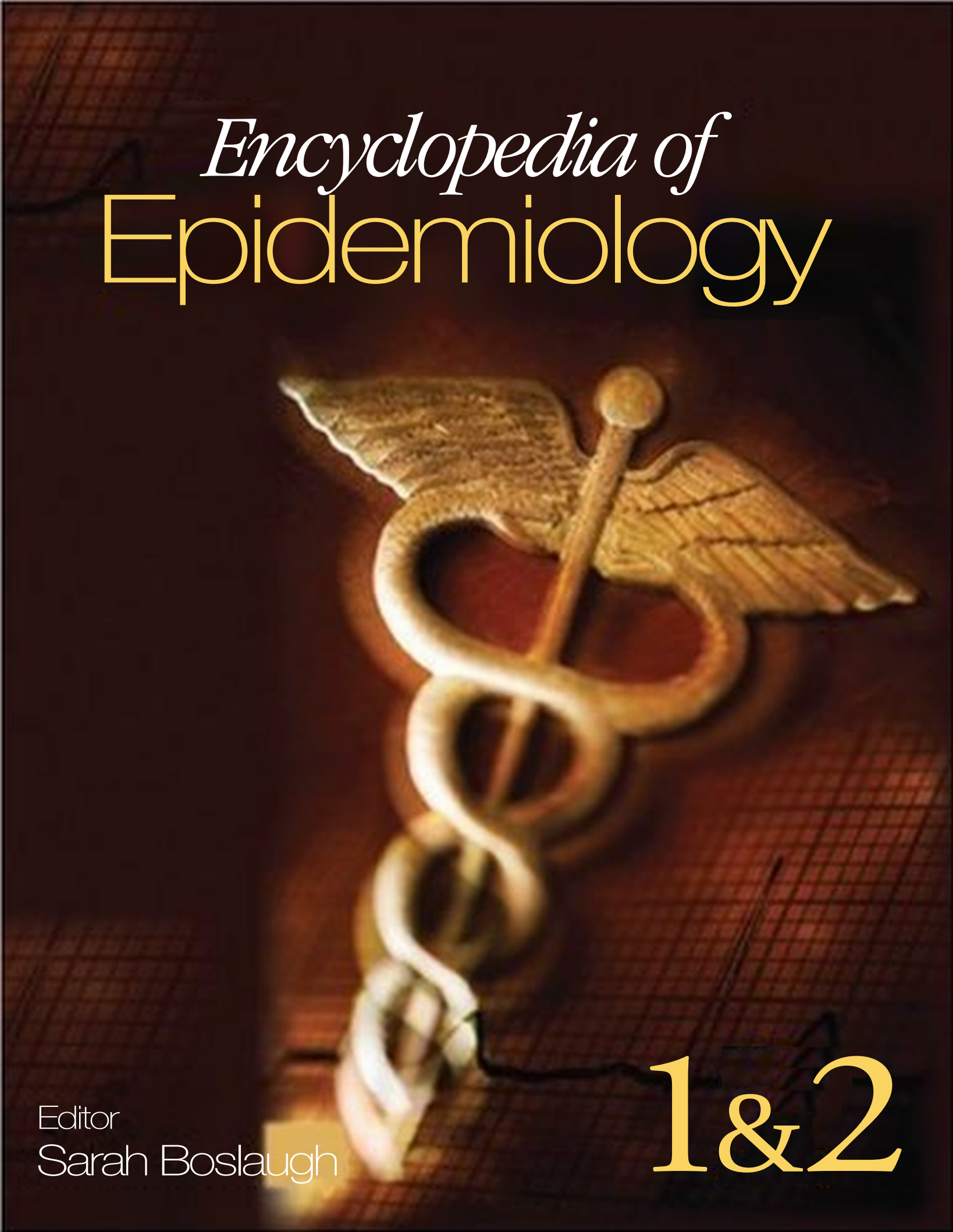
Encyclopedia of Epidemiology



Editor
Sarah Boslaugh

1&2

Encyclopedia of Epidemiology

Encyclopedia of Epidemiology

1&2

Editor

Sarah Boslaugh

Washington University, St. Louis, School of Medicine
BJC HealthCare, St. Louis

Associate Editor

Louise-Anne McNutt

University at Albany, State University of New York

A SAGE Reference Publication

SAGE Publications
Los Angeles • London • New Delhi • Singapore

For information:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
33 Pekin Street #02–01
Far East Square
Singapore 048763

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Encyclopedia of epidemiology/[edited by] Sarah Boslaugh.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-4129-2816-8 (cloth : alk. paper)
1. Epidemiology—Encyclopedias. I. Boslaugh, Sarah.
[DNLM: 1. Epidemiology—Encyclopedias—English. WA 13 E556 2008]

RA652.E533 2008
614.403—dc22 2007031747

This book is printed on acid-free paper.

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

Publisher:	Rolf A. Janke
Acquisitions Editor:	Lisa Cuevas Shaw
Developmental Editor:	Diana E. Axelsen
Reference Systems Manager:	Leticia Gutierrez
Production Editor:	Melanie Birdsall
Copy Editor:	QuADS Prepress (P) Ltd.
Typesetter:	C&M Digitals (P) Ltd.
Proofreaders:	Cheryl Rivard and Dennis W. Webb
Indexer:	Kathy Paparchontis
Cover Designer:	Candice Harman
Marketing Manager:	Amberlyn Erzinger

Contents

Editorial Board

Editor

Associate Editor

Advisory Board Members

List of Entries

Reader's Guide

The Reader's Guide provides an overview to all the entries in the *Encyclopedia* as well as a convenient way to locate related entries within an area of interest. Articles are arranged under headings, which represent broad categories of subjects. For instance, the heading Branches of Epidemiology lists the titles for the 20 main entries on fields of study and practice within epidemiology included in this *Encyclopedia*, from Applied Epidemiology to Veterinary Epidemiology. Similarly, under the heading Epidemiologic Data are listed the main entries on that topic, including types of data (e.g., Administrative Data), specific sources of data (e.g., the Behavioral Risk Factor Surveillance System), and issues related to data management (e.g., Biomedical Informatics). The guide is also useful for finding articles related to a particular topic. For instance, if you are interested in Genetic Epidemiology, you will find that topic under the heading Genetics, which also includes articles on topics such as Epigenetics, the Human Genome Project, and Linkage Analysis. Some topics appear under more than one heading (e.g., Genetic Epidemiology appears under both Branches of Epidemiology and Genetics), reflecting the interrelationships among the broad categories represented by the headings.

Behavioral and Social Science

Acculturation
Bioterrorism
Community-Based Participatory Research
Community Health
Community Trial
Cultural Sensitivity
Demography
Determinants of Health Model
Ecological Fallacy
Epidemiology in Developing Countries
EuroQoL EQ-5D Questionnaire
Functional Status
Genocide
Geographical and Social Influences on Health
Health, Definitions of
Health Behavior
Health Belief Model
Health Communication
Health Communication in Developing Countries
Health Disparities
Health Literacy
Life Course Approach
Locus of Control
Medical Anthropology
Network Analysis
Participatory Action Research
Poverty and Health
Quality of Life, Quantification of
Quality of Well-Being Scale (QWB)
Race and Ethnicity, Measurement Issues With
Race Bridging
Rural Health Issues
Self-Efficacy
SF-36® Health Survey
Social Capital and Health
Social-Cognitive Theory
Social Epidemiology
Social Hierarchy and Health
Social Marketing
Socioeconomic Classification
Spirituality and Health
Targeting and Tailoring
Theory of Planned Behavior
Transtheoretical Model

Branches of Epidemiology

Diseases and Conditions

Epidemiological Concepts

Epidemiologic Data

Ethics

Genetics

Health Care Economics and Management

Health Risks and Health Behaviors

History and Biography

Budd, William
Doll, Richard
Ehrlich, Paul
Epidemiology, History of
Eugenics
Farr, William
Frost, Wade Hampton
Genocide
Goldberger, Joseph
Graunt, John
Hamilton, Alice
Hill, Austin Bradford
Jenner, Edward
Keys, Ancel
Koch, Robert
Lind, James
Lister, Joseph
Nightingale, Florence
Pasteur, Louis
Public Health, History of
Reed, Walter
Ricketts, Howard
Rush, Benjamin
Snow, John
Tukey, John
Tuskegee Study

Infrastructure of Epidemiology and Public Health

American College of Epidemiology
American Public Health Association
Association of Schools of Public Health
Centers for Disease Control and Prevention
Council of State and Territorial Epidemiologists
European Public Health Alliance
European Union Public Health Programs
Food and Drug Administration
Governmental Role in Public Health
Healthy People 2010
Institutional Review Board
Journals, Epidemiological
Journals, Public Health
National Center for Health Statistics
National Institutes of Health
Pan American Health Organization
Peer Review Process
Publication Bias
Public Health Agency of Canada
Society for Epidemiologic Research
Surgeon General, U.S.
United Nations Children's Fund
U.S. Public Health Service
World Health Organization

Medical Care and Research

Allergen
Apgar Score
Barker Hypothesis
Birth Defects
Body Mass Index (BMI)
Carcinogen
Case Reports and Case Series
Clinical Epidemiology
Clinical Trials
Community Health
Community Trial
Comorbidity
Complementary and Alternative Medicine
Effectiveness
Efficacy
Emerging Infections
Escherichia coli
Etiology of Disease
Evidence-Based Medicine
Gestational Age
Intent-to-Treat Analysis
International Classification of Diseases
International Classification of Functioning, Disability, and Health
Latency and Incubation Periods
Life Course Approach
Malnutrition, Measurement of
Medical Anthropology
Organ Donation
Pain
Placebo Effect
Preclinical Phase of Disease
Preterm Birth

Specific Populations

Statistics and Research Methods

About the Editor

Sarah Boslaugh, Ph.D., MPH, has more than 20 years of experience working in data management and statistical analysis. She is currently a performance review analyst in the Center for Healthcare Quality and Effectiveness of BJC HealthCare and an adjunct professor in the Washington University School of Medicine, both in St. Louis, Missouri. She previously worked as a biostatistician and methodologist at Montefiore Medical Center in New York City and for the School of Public Health of Saint Louis University. She is the author of two books, *Secondary Data Sources for Public Health* (2007) and *SPSS Programming: Using Syntax for Data Management* (Sage, 2004), and of numerous book chapters and professional articles. Her research interests include multilevel modeling, quality improvement in healthcare settings, and measurement theory. She received her Ph.D. in research methods and evaluation from the City University of New York and her MPH from St. Louis University.

About the Associate Editor

Louise-Anne McNutt, Ph.D., is an associate professor in the Department of Epidemiology and Biostatistics, University at Albany, State University of New York. Her areas of expertise include infectious disease epidemiology, violence intervention, epidemiologic methods, and quantitative methods. She primarily focuses her international research in the post-Soviet East studying transmission of bloodborne pathogens in health care settings and disease transmission affecting women of reproductive age. In the United States, she conducts research to inform policies on intimate partner violence screening and intervention in primary care settings and sexually transmitted disease interventions. After completing her graduate education, she joined the Epidemic Intelligence Service, Centers for Disease Control and Prevention (CDC). She has held many positions in national organizations, is currently cochair of the Associated Schools of Public Health's Biostatistics and Epidemiology Council, and will join the Editorial Board of the American Journal of Public Health in Fall 2007. Her funding sources for research activities include the National Institutes of Health, CDC, New York State Department of Health, and multiple foundations. She has an MS in biostatistics from the Medical College of Virginia and a Ph.D. in preventive medicine from the University of Iowa.

About the Advisory Board

Elena M. Andresen, Ph.D., is Professor and Chair, Department of Epidemiology and Biostatistics, at the University of Florida College of Public Health and Health Professions. Her training and research interests include chronic disease epidemiology among older adults, disability epidemiology, and outcomes research methods and measures. She has special expertise in research methods and measures regarding older adults and groups with disability and currently receives support for research from the Centers for Disease Control and Prevention, National Institute on Disability and Rehabilitation Research, National Institute on Aging, Department of Veterans Affairs, and the Komen Foundation. She has served as a member of the Institute of Medicine committees on issues of aging and disability and has numerous service activities for a number of federal agencies, including the Centers for Disease Control and Prevention, National Institute on Disability and Rehabilitation Research, and the Agency for Healthcare Research and Quality. She received her Ph.D. in epidemiology from the University of Washington in 1991.

Peter S. Arno, Ph.D., is a health economist and professor in the Department of Health Policy and Management at New York Medical College. He was a Pew Postdoctoral Research Fellow at the Institute for Health Policy Studies and the Institute for Health and Aging at the University of California, San Francisco (1984–1986); a scholar of the American Foundation for AIDS Research (1989–1992), and recipient of an Investigator Award in Health Policy from the Robert Wood Johnson Foundation (1998–2003); and coauthor of *Against the Odds: The Story of AIDS Drug Development, Politics & Profits*, nominated for a Pulitzer Prize. His recent work includes studies of regulation and pricing practices of the pharmaceutical industry; the economics of informal caregiving and long-term care; public health and legal implications of regulating tobacco as a drug; cost, access, quality, and outcome measures related to HIV disease; and the impact of income support policies on health. Based on this work, he has testified before numerous U.S. House and Senate committees.

Ross C. Brownson, Ph.D., is a professor of epidemiology and Director of the Prevention Research Center at Saint Louis University School of Public Health (SLU-SPH) in Missouri. As a chronic disease epidemiologist, he is involved in numerous community-based interventions designed to reduce modifiable risk factors such as physical inactivity and tobacco use. In particular, he is interested in the impacts of environmental and policy interventions on health behaviors. He also conducts research on dissemination of evidence-based interventions. His research is funded by the Centers for Disease Control and Prevention, the National Institutes of Health, and the Robert Wood Johnson Foundation. He is the author of 6 books, more than 200 peer-reviewed articles, and is on numerous editorial boards. His books include *Chronic Disease Epidemiology and Control, Applied Epidemiology,* and *Evidence-Based Public Health*. He has led in the development of the internationally recognized training course in evidence-based public health. He is the recipient of the Award for Excellence in Prevention Research and Research Translation in Chronic Disease (2000, from CDC) and the Abraham Lilienfeld Award for outstanding contributions in teaching and mentoring from the American Public Health Association. Prior to joining SLU-SPH, he was a division director (a cabinet-approved appointment) with the Missouri Department of Health. In this capacity, he wrote or cowrote bills on tobacco access to minors, state clean indoor air, and private insurance coverage of screening mammography and Pap smear testing.

Diane Sperling Lauderdale, Ph.D., is an associate professor in the Department of Health Studies and the College of the University of Chicago. She has been a member of the faculty at the University of Chicago

since 1996 and teaches courses in epidemiologic research methods and the history of public health in the United States. As a chronic disease epidemiologist, she carries out population-based studies of how chronic disease and health-related behaviors, including physical activity and obesity, receipt of preventive medical care, and sleep, vary in the population. She is the author of more than 60 papers in peer-reviewed journals. Her research is funded by the National Institutes of Health, the American Cancer Society, and the Robert Wood Johnson Foundation. Much of her research has focused on the health and health care of immigrant populations in the United States, including studies of birth outcomes, osteoporosis, mortality, and discrimination experiences in medical care. She received an AB in the comparative study of religion from Harvard University, MA degrees in divinity and library science from the University of Chicago, and a Ph.D. in public health (epidemiology) from the University of Illinois at Chicago School of Public Health.

William D. Shannon, Ph.D., is Associate Professor of Biostatistics in Medicine in the Department of Medicine, Washington University School of Medicine. As an applied biostatistician and statistical methodologist, he is involved with a broad range of biomedical data analysis problems, including the genetics of leukemia, field trials for the eradication of lymphatic filariasis in Egypt, developing new methods for sleep medicine in cardiovascular and HIV/AIDS research, designing clinical trials on the impact of automated computer alerting systems to improve physician compliance with medical guidelines, and developing prediction rules for sepsis in intensive care settings. As Director of the Washington University Department of Medicine's Biostatistical Consulting Center, he is continually presented with the challenges of analyzing biomedical data. Of particular interest to him are the statistical methods of cluster analysis and classification analysis. He is on the Board of Directors of the Classification Society of North America and Interface of Computing Science and Statistics and has organized and hosted three of their annual meetings at Washington University. He has received his Ph.D. from the University of Pittsburgh, Graduate School of Public Health, in 1995.

Contributors

Annette L. Adams
Oregon Health & Science University

H. Douglas Adams
Saint Louis University

Saifuddin Ahmed
Johns Hopkins University

Icek Ajzen
University of Massachusetts

Joerg Albrecht
John H. Stroger, Jr., Hospital of Cook County, Chicago, Illinois

Dominik D. Alexander
Exponent, Inc.

Matthew Allison
University of California, San Diego

Paul D. Allison
University of Pennsylvania

Emily E. Anderson
American Medical Association

Jennie E. Anderson
Harvard University

Jennifer Andrews
University of Arizona

Mary K. Anglin
University of Kentucky

Alyson Anthony
University of Kentucky

Marie T. Ascher
New York Medical College

Najood Ghazi Azar
Azusa Pacific University

Emilia Bagiella
Columbia University

Kate Bassil
University of Toronto

Roberto Augusto Becker
Pan American Health Organization/WHO

Robert Bednarczyk
University at Albany, State University of New York

Michael Begleiter
Children's Mercy Hospital

Judith Marie Bezy
University of Washington

Jerome E. Bickenbach
Queen's University

Michael Bigby
Harvard Medical School

Scott M. Bilder
Rutgers, State University of New Jersey

Gretchen L. Birbeck
Michigan State University

Marjorie E. Bond
Monmouth College

Luisa N. Borrell
Columbia University

Rickie Brawer
Thomas Jefferson University Hospital

Joel Breman
National Institutes of Health

Naomi Breslau
Michigan State University

Nathaniel J. Brown
Saint Louis University

Emily K. Brunson
University of Washington

Carol Anne Bryant
University of South Florida

Keneshia Bryant
University of Phoenix

Jane K. Burke-Miller
University of Illinois at Chicago

Merlin G. Butler
Children's Mercy Hospital and University of Missouri

Maia Butsashvili
Georgian Rehabilitation Center

Alexia Campbell
University of Toronto, Canada

Carlos Campillo
Balearic Health Service, Palma de Mallorca, Spain

Belinda Campos
University of California, Los Angeles

Ana W. Capuano
University of Iowa

Martha Ann Carey
Azusa Pacific University

Cecilia Casanueva
University of North Carolina at Chapel Hill

Alissa R. Caton
University at Albany, State University of New York

Keely Cheslack-Postava
Johns Hopkins University

Howard D. Chilcoat
GlaxoSmithKline, Worldwide Epidemiology, North Carolina

Margaret Chorazy
University of Iowa College of Public Health

Hwan Chung
Michigan State University

Janci L. Chunn
Exponent, Inc.

Eddie M. Clark
Saint Louis University

Margaret H. Clark
Southern Illinois University-Carbondale

Lynn Clemow
Columbia University, College of Physicians & Surgeons

Norman A. Constantine
Public Health Institute and University of California, Berkeley

Chris M. Coombe
University of Michigan

Hannah L. F. Cooper
National Development and Research Institutes, Inc.

Christopher Cunniff
University of Arizona

Jennifer Currie
University of Arizona

Amy B. Dailey
University of Florida

Thomas M. Daniel
Case Western Reserve University

Troy Day
Queen's University, Kingston, Ontario, Canada

Martha Decker
Harder + Company Community Research

Deborah L. Dee
University of North Carolina at Chapel Hill

Erin L. DeFries
University of Florida

Michael A. DeRosa
QualityMetric, Inc.

Diane Mary Dewar
University at Albany

Scott R. Diehl
New Jersey Dental School

James M. DuBois
Saint Louis University

William W. Eaton
Johns Hopkins University

Charlotte H. Edinboro
Exponent, Inc.

Millicent Eidson
New York State Department of Health

Felicity Boyd Enders
Mayo Clinic

Molly Engle
Oregon State University

Daniel Epstein
Pan American Health Organization

Robert P. Erickson
University of Arizona

R. Gregory Evans
Saint Louis University

James A. Fain
University of Massachusetts

Margaret Daniele Fallin
Johns Hopkins University

Changyong Feng
University of Rochester

Megan Dann Fesinmeyer
University of Washington

Amanda Bush Flynn
Anderson Area Medical Center, South Carolina

Betsy Foxman
University of Michigan

Lawrence Frank
University of British Columbia

Rongwei (Rochelle) Fu
Oregon Health & Science University

Sandro Galea
University of Michigan

Jorge Bacallao Gallestey
Universidad Médica de La Habana

Joseph C. Gardiner
Michigan State University

Chandler Gatenbee
University of Utah, Salt Lake City

David F. Gillespie
Washington University in St. Louis

Karen Glanz
Emory University

Mary Earick Godby
Tailored Statistical Solutions, LLC

Binh Y. Goldstein
University of California, Los Angeles

Karen J. Goodman
University of Alberta, Edmonton, Canada

Renee Goodwin
Columbia University

José A. Tapia Granados
University of Michigan, Ann Arbor

Thomas K. Greenfield
Alcohol Research Group

Sander Greenland
University of California, Los Angeles

Erik J. Groessl
VA San Diego/University of California, San Diego

Stefano Guerra
University of Arizona

Paul Gustafson
University of British Columbia

V. Anna Gyarmathy
Johns Hopkins University

Perry N. Halkitis
New York University

Rebecca Harrington
Johns Hopkins University

Mark Gerard Haug
University of Kansas

Randall A. Heidenreich
University of Arizona

Susan J. Henly
University of Minnesota

Cheryl L. Holt
University of Alabama at Birmingham

Steven P. Hooker
University of South Carolina

David C. Howell
University of Vermont

John J. Hsieh
University of Nevada

Shazia Hussain
University of Albany School of Public Health

Floyd Hutchison
University of Rochester, New York

Ana S. Iltis
Saint Louis University

Deborah D. Ingram
National Center for Health Statistics

Lydia Isaac
Johns Hopkins University

Barbara A. Israel
University of Michigan, Ann Arbor

Gudmund R. Iversen
Swarthmore College

Susan Jamieson
University of Glasgow

Michael Jernigan
Massachusetts General Hospital, Boston

Roy Jerome
New York University

Carey V. Johnson
Fenway Institute at Fenway Community Health

Adam Jones
Yale University

Namratha Kandula
Northwestern University

Robert M. Kaplan
University of California, Los Angeles

Abhijay P. Karandikar
University of Illinois at Chicago

Sangeeta Karandikar
Independent scholar

David L. Katz
Yale University

Gwen Kay
State University of New York at Oswego

Michael A. Kelsh
Exponent, Inc., and University of California, Los Angeles

Daniel Kim
Harvard School of Public Health

Leo Kinlen
Oxford University

Michelle Kirian
San Francisco Department of Health

Denise Koo
Centers for Disease Control and Prevention

Alan M. Kraut
American University

Matthew W. Kreuter
Saint Louis University

Allison Krug
Artemis Biomedical Communications, LLC

Andrzej Kulczycki
University of Alabama at Birmingham

Naa Oyo A. Kwate
Columbia University

Helen L. Kwon
Columbia University

Jodi Lapidus
Oregon Health & Science University

Thomas A. LaVeist
Johns Hopkins University

William F. Lawrence
Agency for Healthcare Research and Quality

Sarah A. Leavitt
National Building Museum, Washington, D.C.

Brian K. Lee
Johns Hopkins University

Li-Ching Lee
Johns Hopkins University

Robert C. Lee
University of New Mexico

Justin Lessler
Johns Hopkins University

Barry S. Levy
Tufts University School of Medicine

Hilary James Liberty
National Development and Research Institutes, Inc.

Donald A. Lloyd
Florida State University

Mary M. Lopez
Azusa Pacific University

Alexandra M. Lord
Office of the Public Health Service Historian

Molly Lund
Children's Mercy Hospital

Zhehui Luo
Michigan State University

Anu Manchikanti
University of North Carolina at Chapel Hill

JoAnn Manson
Brigham and Women's Hospital

William N. Marshall
University of Arizona, Tuscon

Hilarie Martin
Male Circumcision International Research Collaborative

Dan Mayer
Albany Medical College

David L. McArthur
University of California, Los Angeles

Leslie McIntosh
Washington University in St. Louis

James F. McKenzie
Ball State University

Colleen McLaughlin
University at Albany, State University of New York

Darlene McPherson
Azusa Pacific University

F. John Meaney
University of Arizona

Lee Joseph Melton III
Mayo Clinic

Scott Menard
University of Colorado

Lynne C. Messer
EPA/NHEERL

Eva A. Meyers
Azusa Pacific University

Briana Mezuk
Johns Hopkins University

Yvonne L. Michael
Oregon Health Sciences University

Timothy Miller
University of Arizona

Matthew J. Mimiaga
Harvard University School of Public Health

David M. Morens
National Institute of Allergy and Infectious Diseases/National Institutes of Health

Hal Morgenstern
University of Michigan School of Public Health

Libby M. Morimoto
Exponent, Inc.

Martina Morris
University of Washington

Margaret P. Moss
University of Minnesota, Minneapolis

Fionna S. Mowat
Exponent, Inc.

Oscar J. Mujica
Pan American Health Organization/World Health Organization

Sean Nagle
Columbia University

Philip C. Nasca
University of Massachusetts

Britta Neugaard
James A. Haley VA Medical Center

Craig Newschaffer
Drexel University

Ellen Aagaard Nohr
Aarhus Universitet

Liam M. O'Brien
Colby College

Anne P. Odell
Azusa Pacific University

Babatunde Olowokure
HPA Regional Surveillance

Jorn Olsen
University of California, Los Angeles

Todd M. Olszewski
Yale University

Danielle C. Ompad
New York Academy of Medicine

Margret O'Neall
Saint Louis University

Joe Palau
VA San Diego Healthcare System

Nigel Paneth
Michigan State University

Andrew Park
IRD, Montpellier, France

Edith A. Parker
University of Michigan, Ann Arbor

Jerry C. Parker
University of Missouri, Columbia

D. Maxwell Parkin
Cancer Research UK Centre for Epidemiology

Judea Pearl
University of California, Los Angeles

Daniel Peck
Milesquare Associates, Ruckersville, Virginia

Brian Perron
University of Michigan, Ann Arbor

Marjorie A. Pett
University of Utah

Sydney Pettygrove
University of Arizona

Ruth Pfeiffer
National Cancer Institute

Carl V. Phillips
University of Alberta, Edmonton, Canada

Robert R. Pinger
Ball State University

James Plumb
Thomas Jefferson University and Hospital

Anna E. Price
University of South Carolina, Columbia

Kathy Prue-Owens
University of Arizona

Dee Quinn
University of Arizona

Kathleen Ragsdale
Mississippi State University

Shoba Ramanadhan
Harvard School of Public Health

Colleen A. Redding
University of Rhode Island

Rachel Relosa
Johns Hopkins University

Angela Reyes
Detroit Hispanic Development Corporation

Sydney Rice
University of Arizona

Stefan Riedel
University of Iowa Hospitals & Clinics

Naomi B. Robbins
NBR

Kevin Robinson
University of Michigan

Brad Rodu
University of Louisville

Kara E. Rogers
University of Arizona

Anthony Roman
University of Massachusetts at Boston

Lindsay Rosenfeld
Harvard University

Joseph S. Rossi
University of Rhode Island

Ellen Rosskam
University of Massachusetts, Lowell, and University of Surrey

Rima E. Rudd
Harvard School of Public Health

Marcia Russell
Pacific Institute for Research and Evaluation

Vetta L. Sanders-Thompson
Saint Louis University

Carolyn E. Sartor
Washington University School of Medicine

Donna Scemons
Health Care Systems, Inc.

Darcell P. Scharff
Saint Louis University

Marc Schenker
University of California, Davis

Mario Schootman
Washington University

Amy J. Schulz
University of Michigan, Ann Arbor

Rachel D. Schwartz
Saint Louis University

Mary Alice Scott
University of Kentucky

Denise M. Seabert
Ball State University

Elizabeth Serraillier
Boston University School of Public Health

William R. Shadish
University of California, Merced

Patti Shakhshir
Azusa Pacific University

Jerne Shapiro
Florida Department of Health

Karen Shaw
Health Protection

Mona Shum
Exponent, Inc.

Victor W. Sidel
Montefiore Medical Center and Albert Einstein College of Medicine, New York

Laura A. Siminoff
Case Western Reserve University

Elyse Singer
Bard College

Merrill Singer
Hispanic Health Council

Margie Skeer
Harvard School of Public Health

Tara C. Smith
University of Iowa

Thomas Songer
University of Pittsburgh

Carl Stepnowsky
VA San Diego Healthcare System

Carla L. Storr
Johns Hopkins University

Marc Strassburg
University of California, Los Angeles

Donna F. Stroup
Centers for Disease Control and Prevention

Parminder S. Suchdev
Centers for Disease Control and Prevention

Wan Tang
University of Rochester

Michael E. Tarter
University of California, Berkeley

Angela Taylor
National Development Research Institutes, Inc., New York

Cynthia Taylor
University of Arizona

Stacie Ezelle Taylor
Tailored Statistical Solutions, LLC

Steven Teutsch
Merck & Co., Inc.

Stephen B. Thacker
Centers for Disease Control and Prevention

Christopher Tillquist
University of Louisville

David Todem
Michigan State University

Ghazwan G. Toma
University at Albany, State University of New York

Amy Tsui
Johns Hopkins University

Renjin Tu
Columbus State University

Xin Tu
University of Rochester

Katherine L. Tucker
Tufts University

Mary Tully
University of North Carolina at Chapel Hill

Diane M. Turner-Bowker
QualityMetric, Inc.

Rita Velikina
University of California, Los Angeles

David Vlahov
New York Academy of Medicine

Meghan E. Wagner
Exponent, Inc.

James Walkup
Rutgers University

Robert B. Wallace
University of Iowa

Julia Walsh
University of California, Berkeley

Eve Waltermaurer
State University of New York at New Paltz

Hongyue Wang
University of Rochester

John E. Ware, Jr.
QualityMetric, Inc.

Christian Warren
New York Academy of Medicine

Brian M. Waterman
BJC HealthCare

Vanessa Watts
Harvard University

Nancy L. Weaver
Saint Louis University

Lara Carson Weinstein
Thomas Jefferson University

June M. Weintraub
San Francisco Department of Public Health

Sheila Weiss Smith
University of Maryland

April L. Winningham
University of Arkansas at Little Rock

Ashby Wolfe
University of California, Berkeley

Lisa S. Wolff
Harvard School of Public Health

Nathan D. Wong
University of California, Irvine

Diana Lynn Woods
University of California, Los Angeles

J. Michael Wright
U.S. Environmental Protection Agency

Alvin L. Young
University of Oklahoma

Qin Yu
University of Rochester

Zuo-Feng Zhang
University of California, Los Angeles

Introduction

The Field of Epidemiology

Epidemiology is the study of the frequency and determinants of morbidity and mortality in populations. Besides being a discipline in its own right, the science of epidemiology is one of the foundational sciences of public health and of evidence-based medicine. The Centers for Disease Control and Prevention (CDC) estimate that of the more than 30 years increase in life span experienced by Americans between the years 1900 and 1998, approximately 25 of those years can be attributed to improvement in public health. Epidemiology played a key role in these improvements by identifying risk factors and causal agents for disease and by aiding in the development and evaluation of public health and educational programs designed to reduce morbidity and mortality. Epidemiology has proved equally useful in the drive to prevent and control infectious disease through measures such as provision of clean water, the institution of widespread vaccination programs, and in the reduction of morbidity and mortality from chronic disease through identification of risk factors such as tobacco smoking, consumption of a high-fat diet, and lack of physical activity.

Epidemiologic studies often begin by describing variation in disease occurrence or other health outcomes according to the variables of person, place, and time. *Person* variables describe who is becoming ill—their gender, ethnic group, usual diet, and so on. *Place* variables may describe where the individual was exposed to the disease-causing agent, the location where they became ill, or where the agent (such as a rodent or mosquito) became infected with the disease. *Time* variables may relate either to measurements such as the age or birth cohort of those infected, the duration of an infection or disease, or to cyclical or secular trends in disease occurrence. Descriptive studies of disease occurrence often suggest hypotheses that may be investigated further in analytic studies that examine many variables to discover causal or risk factors related to disease occurrence.

Epidemiologic studies have led to massive improvements in public health, sometimes even before the mechanisms of disease causation are completely understood. For instance, Ignaz Semmelweis's observations of the differing rates of puerperal fever within different wards of the Vienna General Hospital in the mid-19th century led him to institute antiseptic procedures that drastically lowered both morbidity and mortality rates, years before the germ theory of disease was accepted. Similarly, Joseph Goldberger was able to demonstrate that pellagra, a disease epidemic in the American South in the early 20th century, was not an infectious disease but was related to poor nutrition. Goldberger was also able to prevent outbreaks of pellagra by having individuals consume small amounts of brewer's yeast, a breakthrough achieved a decade before niacin deficiency was definitively identified as the cause of pellagra. And the first recognition that cigarette smoking was a primary risk factor for lung cancer came from observational epidemiologic studies conducted by Sir Richard Doll and Sir Austin Bradford Hill in Great Britain in the 1950s, studies conducted before any specific mechanism relating smoking and lung cancer had been proposed.

The scope of epidemiology is continually widening to accommodate expanded definitions of health and to accommodate our ever-broadening understanding of the influence of diverse factors on health. Although epidemiologic thought may be traced back to ancient times, growth and development of epidemiology as a profession began in the 19th century, when the science of epidemiology was developed partly in response to public health concerns such as the cholera epidemics that regularly occurred in many large cities. So successful were the interventions devised in this early

period that many infectious diseases that still plague the developing world are almost unknown in the industrialized world, thanks to public health measures such as the provision of potable tap water, control of mosquitoes and other pests, and institutions of widespread vaccination programs. Infectious diseases still remain a concern in industrialized countries, however, and the emergence of diseases such as HIV and multidrug-resistant tuberculosis pose new challenges for the science of epidemiology.

The first great expansion in the scope of epidemiology was the inclusion of chronic as well as infectious diseases. This is a logical development since, partly due to greater control of infectious disease, the greatest causes of morbidity and mortality in developed countries today are chronic diseases such as cancer, cardiovascular disease, and stroke. These diseases are also significant health factors in the developing world: In fact, the World Health Organization estimates that 80% of deaths due to chronic disease now occur in developing countries. Increased study of chronic diseases has brought with it a massive increase in complexity, since most such diseases do not have a single causal factor or solution. While cholera is caused by ingestion of the bacterium *Vibrio cholerae* and effective prevention requires the public health measure of preventing the bacterium from polluting drinking water or removing or killing it through filtration, chemical treatment, or boiling, heart disease may be influenced by many things, including genetics, obesity, tobacco smoking, and comorbid conditions, and prevention and treatment are likely to require an individualized combination of behavioral and lifestyle changes as well as medical intervention. In fact, many of the risk factors for chronic diseases are individual behaviors, and epidemiology's growing awareness of the importance of individual health behaviors on morbidity and mortality led to the establishment, in the 1980s, of the Behavioral Risk Factor Surveillance System. It has also led to numerous public health campaigns aimed at lowering chronic disease morbidity and mortality by inducing individuals to make alterations in their lifestyle, such as increasing physical activity, eating more fruits and vegetables, and quitting smoking and moderating alcohol consumption.

An even newer focus of concern for epidemiology is the influence of social and geographical factors on health. We now realize that not only individual behavior but also the material facts of a person's life may be major determinants of their health. Geographical influence on health includes not only obvious factors such as living in a polluted region or in a high-crime area but also more elusive qualities such as the type and strength of social relationships that typify a given neighborhood. Social influences also range from the obvious to the remarkable: It's easy to see how poverty could negatively affect health through lack of access to health care or inability to purchase nutritious food, or how the regular experience of discrimination could harm one's mental and physical health, but recent studies indicate that even in affluent societies where health care is available to all, a person's place within the social hierarchy can be a major influence on their health. Sorting out the influence of the many variables involved remains a challenge for epidemiology.

Epidemiologists are concerned not only with disease but also with injury. This includes accidental injuries, such as those caused by motor vehicles or by falling off a ladder, as well as injuries deliberately caused by oneself or other persons. The inclusion of topics such as suicide, child abuse, and intimate partner violence within the scope of epidemiology represents a commonsense broadening of scope, as does the study of risk factors such as the presence of firearms in the home. War and terrorism have also been included within the scope of epidemiology, which is entirely appropriate given the great toll they extract in human life and suffering.

Another expansion of scope for epidemiology, which was originally concerned with the health of human populations, has been the development of the science of veterinary epidemiology that studies the occurrence of health and disease of animal populations. This expansion is entirely logical because not only are many diseases communicable between animals and humans, but animal husbandry also plays an essential role in securing both the food supply and economic welfare of many human populations.

Organization of the *Encyclopedia*

As the study of epidemiology may potentially incorporate information about anything that affects human health, it incorporates descriptive and analytical techniques borrowed from many areas of study. However, a two-volume encyclopedia cannot include detailed discussions of every topic relevant to epidemiology, so we took the approach of including overview articles on many topics relevant to epidemiology, with their length proportional to the centrality of the topic to epidemiology. Anyone studying epidemiology will need to know something about biostatistics, for instance, but not so much as a full-time biostatistician.

Similarly, a basic knowledge of health economics is relevant to the practice of many epidemiologists, although that knowledge need not be as extensive or detailed as someone working as a health economist. We have organized this material into almost 500 entries, which may be viewed as belonging to a number of broad categories. One way to look at the content of this *Encyclopedia* is by major topics and representative entries included within each:

Behavioral and Social Science. This deals with topics such as acculturation, community health, demography, health behavior, health communication, network analysis, and social epidemiology.

Branches of Epidemiology. This includes 20 fields of study, based on methodological approach or content studied, from applied epidemiology to veterinary epidemiology.

Diseases and Conditions. This deals with the incidence, prevalence, prevention, and control of 40 major infectious, chronic, and psychiatric diseases and conditions, from Alzheimer's disease to zoonotic diseases.

Epidemiological Concepts. This deals with topics that form the basis of the science of epidemiology, including fundamental concepts such as attack rate, attributable fractions, incidence and prevalence, effect modification and interaction, morbidity and mortality rates, herd immunity, and direct and indirect standardization.

Epidemiologic Data. This deals with topics related to important data sets and the acquisition and management of data, including administrative data, cancer registries, notable epidemiologic studies such as the Framingham Heart Study and the Harvard Six Cities Study, publicly available data sets, sampling techniques, and secondary data.

Ethics. This deals with standards of ethics in health care, public health and human subjects research, and when doing epidemiology in developing countries, and related topics such as definitions of health, eugenics, and genocide.

Genetics. This deals with topics related to the burgeoning science of genetic epidemiology, including epigenetics, family studies, genetic counseling, gene-environment interaction, linkage analysis, molecular epidemiology, and twin studies.

Health Care Economics and Management. This deals with topics such as biomedical informatics, economic evaluation, functional status, health economics, managed care, and quality-of-life measurement.

Health Risks and Health Behaviors. This deals with what is known about the influence of numerous agents and conditions on health and disease from Agent Orange to urban sprawl and individual health behaviors from alcohol use to tobacco.

History and Biography. This deals with the history of epidemiology and public health, and biographies of pioneers in the field, from William Budd to John Tukey.

Infrastructure of Epidemiology and Public Health. This deals with epidemiologic and public health organizations, governmental and nongovernmental agencies, major journals, and the publication process.

Medical Care and Research. This deals with topics at the intersection of medicine and epidemiology, including clinical epidemiology; evidence-based medicine; the International Classification of Diseases and the International Classification of Functioning, Disability and Health; medical anthropology; and screening.

Specific Populations. This deals with major threats to health and patterns of disease and disability within specific populations, as defined by race or ethnicity, gender, age category, urban or rural residence, immigrant or refugee status, and sexual identity.

Statistics and Research Methods. This deals with theoretical background and applied techniques of epidemiology and statistics, including categorical data analysis, bias, causation and causal inference, classification and regression trees, cluster analysis, factor analysis, geographical and spatial analysis, multilevel modeling, nonparametric statistics, qualitative methods, regression, sequential analysis, and structural equation modeling.

The Growth of the Field of Epidemiology

The demand for epidemiologists is growing faster than the available supply of trained personnel, although enrollment in both public health programs in general, and epidemiology programs in particular, has been steadily expanding for years. In the United

States, most formal epidemiologic study takes place at the graduate (master's and doctoral) level at one of 37 accredited schools of public health, who enroll students from all parts of the world. University study in epidemiology and public health is offered in many other countries as well, including Canada, Japan, and numerous European countries.

Epidemiology is truly an international science, and epidemiologic research today is conducted in all parts of the world. The international growth of epidemiology as a science is demonstrated by the expansion of the International Epidemiological Association, which began as an "International Corresponding Club" founded by three British and American researchers in 1954, that now has members from more than 100 countries and whose triennial Congress in 2002 was attended by more than 1,200 people. Regional epidemiologic conferences also regularly draw participants from outside their geographic region; for instance, the 2004 European Congress of Epidemiology accepted individuals from 38 countries, representing all six continents.

In 2004 to 2005, U.S. schools of public health produced 6,656 graduates, a 43.6% increase from 1994 to 1995. Epidemiology was the most popular field of study, with 19.5% of students choosing this concentration. Despite this increase in graduates, there is a shortage of epidemiologists in the United States and worldwide. Many people who are currently working in epidemiology without formal training, such as the Council of State and Territorial Epidemiologists (CSTE), estimates that almost half the individuals working as epidemiologists in state and territorial health departments are not academically trained in the subject, with the largest gaps in those working in infectious disease epidemiology and injury epidemiology. Of course, many of these individuals are not able to stop working and enroll in a full-time course of study; even if they were, there are not sufficient places in university epidemiology programs to accommodate them. This situation led to development by the CDC and the CSTE of the *Competencies for Applied Epidemiologists in Governmental Public Health Agencies,* which specify what topics should be included in competency-based, on-the-job training for epidemiologists in government public health agencies.

Rationale for the *Encyclopedia*

The extraordinary growth of the field of epidemiology in the past several decades, coupled with an ever-increasing demand for people with epidemiologic training, has created a void in the literature. There are several highly technical encyclopedias directed to trained epidemiologists and biostatisticians and basic dictionaries that define key terms in the field, but there is no single reference work that describes basic epidemiologic concepts in sufficient depth for practitioners and that is accessible to the nonspecialist also. In addition, there is no reference that combines technical information about the techniques used in epidemiology with substantive information about the epidemiology of specific diseases and conditions or about the health status of particular population groups.

The *Encyclopedia of Epidemiology* is designed to be comprehensive in its coverage of topics but not exhaustive in its treatment of any one topic. Basic information about epidemiology is presented in a manner that is appropriate for both a student beginning university study in the field and the interested general reader. Technical articles such as those on biostatistical techniques are presented in a nontechnical manner so that they may be understood without advanced statistical training. Particular emphasis has been placed on communicating with individuals outside the field. Epidemiology plays too important a role in the modern world to be left to epidemiologists alone, and many people working outside the field need a working knowledge of at least some aspects of epidemiology. Such persons include health care professionals such as nurses and physicians, researchers in other social science fields, social workers, and journalists.

Content and Organization

This *Encyclopedia* was designed to include every topic that would be of major concern to persons studying or working epidemiology; thus, many topics from related fields such as biostatistics, health psychology, and health economics are included. Because many topics are interrelated, and to avoid redundancy, all topics include cross-references to related topics. In some cases, a topic was completely covered within an article on another topic; in that case, the smaller topic is listed in its alphabetical position with a cross-reference to the entry in which it is discussed.

To help the reader navigate the *Encyclopedia*, a Reader's Guide is provided that organizes the content into the 14 large topics previously enumerated. Additionally, a list of the entries is presented in alphabetical order.

The content of each article is intended to provide a concise and nontechnical summary of the major aspects of the topic, providing the reader with a fundamental

A

Acculturation

Acculturation is the process of individual or group adjustment to a new culture. As a result of continuous contact with a culture that is different from his or her culture of origin, an individual typically undergoes psychological, behavioral, and attitudinal changes that can affect psychological and physical health. The relevance of acculturation to public health is an issue of national, and increasingly global, importance.

The process of acculturation can take many paths. Individuals may embrace new cultural beliefs and practices, strive to retain culture of origin, or develop bicultural identities. Acculturation in the United States, however, has repeatedly been associated with numerous adverse health behaviors such as increased smoking and drug use. Understanding acculturation processes and constructs that underlie acculturative change is central to developing effective public health practices and policies that promote healthy populations. This entry examines Latino acculturation to the United States to illustrate the importance of acculturation to epidemiology and to highlight the gains to public health that increased understanding of acculturation may bring. The processes discussed, however, may apply to other immigrant populations acculturating to new environments.

Definitions of Acculturation

Acculturation has traditionally been defined as a linear movement away from the culture of origin and toward a new culture. The process of acclimating to a new culture, however, is increasingly recognized to be a complex and multidimensional one, in which individuals retain the values and practices from their culture of origin and adopt the new culture's values and practices to varying degrees. Recent research on biculturalism, for example, suggests that two cultural identities can be retained simultaneously and with varying degrees of psychological ease. Newer measures of acculturation, such as the Acculturation Rating Scale for Mexican Americans-II (ARSMA-II), take this variation into account and can simultaneously generate measures of linear, multidimensional, and bicultural types of acculturation.

Measuring Acculturation

Acculturation measures ideally are tailored to the cultural adjustment experience that is being examined. For example, a researcher studying the health impact of Latino acculturation to the United States strives to measure the behaviors and beliefs that are most representative of Latino culture and U.S. European American culture. This type of measure, however, is not always practical in applied or epidemiologic settings. Thus, many researchers rely on proxies of acculturative change, such as an individual's country of birth, time spent in the new cultural environment, and the extent to which an individual prefers the language, media, and values of the new culture versus culture of origin. These proxies are often useful estimates of acculturation but lack the fine-grained distinctions of more multidimensional measures. Both direct

and proxy measures, however, may fail to address the underlying causes of change in disease risk.

Health and Acculturation

Acculturation is relevant to many health processes, including (1) reproductive health (e.g., sexual behavior, experience and outcomes of pregnancy), (2) disease risk and management (e.g., diabetes and various forms of cancer), (3) psychosocial processes that affect health (e.g., stress and coping patterns), and (4) the development of effective prevention and intervention procedures.

Cultural beliefs and practices surrounding sex and pregnancy are related to differences in how individuals manage their reproductive health. Relative to U.S.-born European Americans, for example, less acculturated Latina immigrants to the United States are known to delay sex longer, have fewer partners over their life course, and be less likely to engage in sexually risky behaviors. Once pregnant, these women are more likely to have a positive attitude toward pregnancy and motherhood and have comparable birth outcomes to U.S.-born European Americans. With greater acculturation to the United States, Latinas begin to show a more adverse pattern that includes earlier sex, more partners, less positive attitudes toward pregnancy and motherhood, and poorer birth outcomes.

Acculturation is also linked with changes in health behaviors that can influence risk factors for developing diseases such as diabetes and cancer. Less acculturated Latinos, for example, have lower mortality from all causes and lower rates of various cancers relative to the rest of the U.S. population. With greater time in the United States, Latinos acquire less nutritious dietary habits (e.g., greater reliance on fast food) and become more likely to smoke, drink alcohol, and use drugs than their less acculturated counterparts. Research suggests that these changes result from greater access to convenience food and from the acceptance of smoking, alcohol, and drugs in the United States.

Psychosocial processes, such as strategies for coping with stressful life events, are also associated with cultural beliefs and have been found to vary by acculturation and to play a role in health outcomes. For example, more acculturated Latinas in the United States report feeling more everyday stress and pregnancy-related anxiety than less acculturated counterparts. Researchers have suggested that the process of adjusting to a new culture accounts for increased perceptions of stress, but another possibility is that the culture of origin may be protective against stress or may promote more effective coping strategies.

Finally, the development of effective strategies for preserving health, such as obtaining and complying with prevention or intervention information, is linked to acculturation. Turning again to the example of Latino acculturation to the United States, understanding changes in norms for leisure-time activity and communication styles is important for developing effective prevention and intervention programs directed at the Latino community. For example, more acculturated Latinos appear to exercise more regularly than less acculturated counterparts and might be persuaded to use this behavior to compensate for a less nutritious diet. In terms of intervention, many cultures, including Latino culture, favor indirect communication and strive to give socially desirable responses. Intervention programs aimed at relatively unacculturated Latinos or members of similarly oriented cultures stand to benefit from taking these norms into account.

—Marc Schenker and Belinda Campos

See also Asian American/Pacific Islander Health Issues; Health Communication; Immigrant and Refugee Health Issues; Latino Health Issues

Further Readings

Berry, J. W. (1980). Acculturation as varieties of adaptation. In A. M. Padilla (Ed.), *Acculturation: Theories, models, and some new findings* (pp. 9–25). Boulder, CO: Westview.

Bethel, J. W., & Schenker, M. B. (2005). Acculturation and smoking patterns among Hispanics. *American Journal of Preventive Medicine*, *29*, 143–148.

Chun, K. M., Balls, O. P., & Marin, G. (Eds.). (2003). *Acculturation: Advances in theory, measurement and applied research.* Washington, DC: American Psychological Association.

Cuellar, I., Arnold, B., & Maldonado, R. (1995). Acculturation Rating Scale for Mexican Americans II: A revision of the original ARSMA Scale. *Hispanic Journal of Behavioral Science, 17*, 275–304.

Addiction

See Drug Abuse and Dependence, Epidemiology of

Additive and Multiplicative Models

Additive and multiplicative models are two alternative approaches to modeling effect of risk factors on disease. In additive models, risk—or disease incidence—changes by some fixed amount when a risk factor is present. The statement "If you take up *X*, you will increase your risk of *Y* by 10%" is an example of an additive model of risk. In contrast, multiplicative models represent the changes in risk as a proportion of the baseline risk. The statement "If you take up *X*, you will double your risk of *Y*" is an example of a multiplicative model of disease risk.

The distinction between additive and multiplicative models becomes especially important when considering the effect of multiple risk factors. For instance, consider the situation described in Table 1.

Table 1 Disease Incidence per 100,000

	No Exposure A	*Has Exposure A*
No Exposure B	10	20
Has Exposure B	30	?

The incidence of disease we expect to see in those with exposure to both A and B differs depending on whether we take a multiplicative or additive approach to disease risk. Under the additive model, we would expect to see 40 cases per 100,000 in those with both exposures, since exposure A increases incidence by 10 per 100,000, exposure B increases incidence by 20 per 100,000, and the baseline rate is 10 per 100,000 $(10 + 10 + 20 = 40)$. Under the multiplicative model, we expect to see 60 cases per 100,000 in the group with both exposures, since exposure A doubles the incidence, exposure B triples the incidence, and the baseline incidence is 10 per 100,000 $(10 \times 2 \times 3 = 60)$. When the observed incidence or risk of disease in people with multiple exposures differs from what is expected based on the model being used (whether additive or multiplicative), there is said to be *interaction* or *effect modification* between the exposures on that scale.

In the analysis of epidemiologic data, the choice of an additive or multiplicative model determines the type of regression analysis performed and the risk measures that will be reported. If an investigator is modeling risk as additive, he or she will generally use linear regression and report risk differences. An investigator who is modeling risk on a multiplicative scale will generally perform a logistic regression and report a relative risk or odds ratio. Epidemiologic investigations that are concerned with disease etiology usually use multiplicative models, while those focused on public health impact are more likely to use an additive risk model.

—Justin Lessler

See also Effect Modification and Interaction; Linear Regression; Logistic Regression

Further Readings

de Gonzalez, A. B., & Cox, D. R. (2005). Additive and multiplicative models for the joint effect of two risk factors. *Biostatistcs, 6*(1), 1–9.

Wayne, W. D. (1999). *Biostatistics: A foundation for analysis in the health sciences* (7th ed.). New York: Wiley.

Administrative Data

Administrative data are collected by organizations and agencies expressly for the purpose of conducting administrative tasks and meeting administrative responsibilities for that organization or agency (e.g., evaluate program performance, agency accountability). These tasks and responsibilities may include contacting individuals within the system, tracking resource utilization, reporting counts to an oversight agency, and projecting trends for resource allocation. Examples of administrative data sources include health maintenance organizations, Medicare and Medicaid programs, vital records administrations, school health systems, hospitals, and

health insurance providers. Although administrative data are collected through a system designed for nonresearch purposes, they can be very useful in epidemiologic research and have been used extensively for this purpose. There are, however, limitations to using administrative data in epidemiologic studies.

The most apparent advantage of administrative data is their availability. Abstracting medical records or conducting surveys to collect the same data in a comparable sample size may not be possible due to time and financial constraints. Administrative data sets often contain records spanning decades. Such data are gathered prospectively by the organization or agency and, therefore, can then become a source for nonconcurrent prospective investigations by the epidemiologist. Access to these data is also relatively easy, since most administrative systems are now computerized and data can be transferred onto electronic media for sharing. Administrative data generally are not for public use and require a research proposal submitted to the agency or organization that collects and oversees use of the data, and sometimes a fee is associated with requests. Administrative databases also tend to be relatively large, allowing for subgroup analysis by various factors such as geographic subdivision, gender, ethnicity, and age. The size of a data set is due mainly to the inclusiveness of the administrative records—everyone in the system is included in an administrative data set. This inclusiveness is another advantage of administrative data as it minimizes the prospect of selection bias.

Although there are no data collection or data entry costs associated with administrative data, data cleaning can still be a time-consuming task on a very large administrative data set. In addition, an administrative data set will seldom have all the data elements needed for a specific epidemiologic study, raising concerns over residual confounding, and variables often need to be recoded. Another disadvantage can be the presence of missing data on key items, although this varies by data source. Finally, as most administrative data sets are not population based, study results from analyses of administrative data may not be generalizable to the larger population or community.

An archetypical example of a population-based administrative data resource that is commonly used in perinatal and pediatric epidemiology is state or provincial birth records. In addition to being able to identify and select a cohort from birth records, birth certificates have a variety of data elements that can be useful in studies of adverse neonatal and childhood outcomes. The content of birth records varies by state or province but usually includes the essential elements of name, date and time of birth, sex, race, parents' names, and place of birth. Some will also include birthweight, gestational age, and age, occupation, and race of the parents. Most epidemiologic studies using birth records for secondary data analysis are cross-sectional or case-control in design. Some have linked birth records of siblings to prospectively examine birth outcomes that may be associated with a previous adverse birth outcome, such as low birthweight. The quality and completeness of data items included on birth records are known to vary widely.

Hospital discharge data are another commonly used administrative data source in epidemiology. Hospital discharge data can be obtained on multiple levels of geographic area, from local hospitals to nationwide data sets. Researchers can work directly with specific hospitals to ascertain discharge records, and discharge data are available from state government agencies as well as the U.S. Agency for Healthcare Research and Quality. For example, the California Office of Statewide Health Planning and Development offers hospital discharge data for purchase by qualified researchers. Nearly every state in the United States has a system of collecting hospital discharge data, although there is a wide range in the data elements that are included and the quality of the data sets (as with birth records), and differences among states in whether reporting is voluntary or mandatory and their policies for making the data available to researchers.

Administrative data sets are sometimes combined with other data sets to allow more specific research questions to be examined. An example of a combined data set that is available to researchers is the SEER-Medicare Linked Database, created by linking the Surveillance, Epidemiology, and End Results (SEER) cancer surveillance data set with Medicare records. The SEER network covers approximately 26% of the U.S. population, while the federally funded Medicare program provides health insurance for the elderly (65 and above) and individuals with end-stage renal disease and certain disabilities. The linkage of SEER and Medicare records started in 1986 and has created a database that includes more than 2.4 million people with cancer for whom Medicare health care claims records are available.

—Craig Newschaffer

See also Birth Certificate; Death Certificate; Health Plan Employer Data and Information Set; Medicaid; Medicare

Further Readings

De Coster, C., Quan, H., Finlayson, A., Gao, M., Halfon, P., Humphries, K. H., et al. (2006). Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: Report from an international consortium. *BMC Health Services Research*, *6*, 77.

Kirby, R. S., & Salihu, H. M. (2006). Back to the future? A critical commentary on the 2003 U.S. national standard certificate of live birth. *Birth, 33*, 238–244.

Admissibility of Scientific Evidence in Law

See Evidence, Legal Admissibility of Scientific

African American Health Issues

Measures of health status highlight disparities in the health of African Americans compared with other racial and ethnic groups in the United States. African Americans are more likely to self-report that their health is fair or poor, and a higher occurrence of many health problems is reported among both lower- and higher-income African Americans. Statistics suggest that the situation is more critical for African American men. This entry provides an overview of the health status of African Americans, describes the nature of the health disparities observed, and discusses the factors most frequently associated with disparities in health outcomes.

African or Black Americans are the descendants of any of the black racial groups of Africa. In the United States, this group includes the descendants of African slaves who have lived here for generations, as well as recent immigrants from Africa, the Caribbean, and other countries. African Americans are currently the third largest ethnic group in the United States, with a population of 36.4 million or 12.9% of the total population. Of the total population, on the 2000 U.S. Census, 12.3% reported Black or African American as their only race. The remaining population, or 0.6%, reported this racial classification in combination with one or more races.

Overall Health Status

As is true of health in the United States in general, African American health has improved. In 2004, consistent with the trend for all subgroups, African American life expectancy reached a record high. Life expectancy at birth was 73.1 years compared with 78.3 for whites and 77.8 for all races combined. African American men have a lower life expectancy than men of other racial and ethnic groups and African American women (see Table 1). However, it is important to note that these statistics represent a narrowing of the black-white life expectancy gap.

Continuing the trend toward improved health, between 2003 and 2004 African Americans showed significant decreases in age-adjusted death rate. The relative magnitudes of these decreases were 3.7% for non-Hispanic African American men and 3.4% for women. The age-adjusted death rate for African Americans was higher than that of the general population, with African American men experiencing the highest rate among all racial and gender groups.

On one of the most sensitive indices of health status, infant mortality rate, African Americans have a rate that is approximately twice that of the general population. Maternal death also occurs at higher rates. African American elders also experience higher morbidity and mortality rates compared with the general population. Also, while unintended pregnancy rates in the United States have been declining, low-income African American women continue to have one of the highest rates.

In 2003, the 10 leading causes of death for African Americans were (in order) heart disease, cancer, stroke, diabetes, accidents and unintentional injuries, assault/homicide, kidney disease, respiratory disease, human immunodeficiency virus (HIV), and septicemia (bacteria in the blood). The top 3 causes of death are similar to those found in the general population, although the mortality rates observed among African Americans are higher. The most significant departures from the general population are deaths due to homicide and HIV. Despite a downward trend, homicide is still the leading cause of death among African American males 15 to 24 years of age. In a change from 2001 and 2002 data, kidney disease accounted for more deaths than HIV.

Table 1 African American Comparative Health Statistics

	All Races		*White*		*African American*	
Measure	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
Life expectancy at birth, 2004	75.2	80.4	75.7	80.8	69.5	76.3
Age-adjusted death rate per 100,000, 2004	955.7	679.2	936.9	666.9	1269.4	855.3
Adult smoking, 2002–2004	23.8	19.4	23.8	20.2	25.1	17.7
% Adult overweight, 1999–2002	68.8	61.6	69.4	57.2	62.6	77.1
% Adult obese, 1999–2002	27.5	33.2	28.0	30.7	27.8	48.8

Measure	*All Races*	*White*	*African American*
Adult age-adjusted percentage distribution: never exercised, 2005	61.3	57.2 61.3	65.7 76.6
10–14 years: birth rates per 1,000, 2004	0.7	0.2	1.6
15–19 years: birth rates per 1,000, 2004	41.1	26.7	63.1
Infant mortality rate, 2001–2003	6.79	5.66	13.6
Maternal mortality rate, 2002	7.6	4.8	22.9

Sources: National Center for Health Statistics, (2006) *Health, United States, 2006, With Chartbook on Trends in the Health of Americans*), Tables 19, 27, 35, and 43; Martin, J. A., Hamilton, B. E., Sutton, P. D., Ventura, S. J., Menacker, F., & Kirmeyer, S., (2004), *Births: Final Data for 2004 National Vital Statistics Reports*, *55*(1), Table 7; National Center for Health Statistics (2002), *Maternal Mortality*; Pleis, J. R., LethBridge-Cejku, M., (2006), Summary of Health Statistics for U.S. adults: National Health Interview Survey, *2005 National Center for Health Statistics. Vital Health Statistics*, *10*(232).

Although African Americans are approximately 13% of the population, they currently account for the majority of new HIV infections in America. In addition, diseases such as asthma are higher among African American children and adults.

As is true in the general population, most causes of death have been declining among African American elders. However, lung cancer and deaths from other lung diseases have increased among older African American men and women. There has also been an increase in mortality associated with hypertension. The leading causes of death among African Americans aged 65 years and above are heart disease, cancer, stroke, diabetes, and pneumonia/influenza. The first four causes of death among African American elders are the same as those generally found in the African American population, and the top three causes of death are the same for elders of all ethnic groups. However, diabetes is a more common cause of death among African American elders than in other ethnic groups with the exception of American Indians.

Health behaviors also play a role in overall health. African Americans are more likely to have lower rates on positive and higher rates on negative health indicators. Among African Americans, mean fruit and vegetable intake was 3.7 servings per day, with only 23.7% of African American adults reporting that they ate three or more vegetables and 35.1% reporting eating two or more fruits per day. African Americans often report that they lack a usual source of care and are less likely to have a yearly dental exam. The trend toward increasing obesity is greater among African American adult females, with a 77.3% prevalence of overweight and obesity. African Americans tend to live in environments with poorer housing quality and disproportionately live in areas where the air contains one or more commonly found air pollutants, including ozone. Children from low-income groups and African American children are also at increased risk of lead poisoning.

Health Disparities

Definitions differ, but in general, the term *health disparities* refers to differences in a population's health status that are not based on biology and that are assumed to be related to differences in social status, including education, income, and access to care. Despite improvements in health care and access, African Americans have mortality rates that are higher than those of whites, and they have the highest rates of morbidity and mortality of any U.S. racial and ethnic group. According to the National Center for Health Statistics, the rate of diabetes among African Americans is greater than that found among whites, and heart disease is more than 30% higher. African Americans tend to have the highest rate of high blood pressure of all groups and tend to develop it younger than others. African Americans have higher rates of diabetes-related complications, such as kidney disease and amputations. Furthermore, the incidence of stroke is disproportionately high among African Americans and the mortality rate is higher than among whites. Age-adjusted death rates from asthma are higher among African Americans. African Americans and Hispanics have higher rates of sexually transmitted diseases than whites.

The poorer health status of African American men is well documented. Watkins indicated that African American men have a lower life expectancy than men in Bangladesh, Iran, Colombia, and Sri Lanka. African American men have a higher incidence of type 2 diabetes, are more likely to die from heart disease, have eight times the AIDS rate, and are nine times more likely to die of AIDS than white males. According to Ward et al., among all racial and ethnic groups, African American men had the highest cancer incidence and mortality rates for all sites combined. African American men had higher incidence and mortality rates for prostate, lung, colorectal, and other specific cancers.

Among African Americans, disparities also exist in the rates on several of the leading health indicators. For example, diets high in fat and calories and low in fruits, vegetables, and fiber, and physical inactivity increase the risk of diabetes, hypertension, heart disease, and cancer. Compared with whites, African Americans consume higher amounts of dietary fat and lower amounts of fruits, vegetables, and dietary fiber and are more likely to report no leisure-time physical activity. According to data from the National Health and Nutrition Examination Survey, overweight is more prevalent among African Americans than non-Hispanic whites. More African American girls aged 6 to 19 were considered overweight compared with white children of the same age. In contrast, the rate of overweight among non-Hispanic white and African American males between ages 6 and 19 is similar.

In addition, African American women are less likely to receive prenatal care in the first trimester of a pregnancy. The gap in vaccination rates among African American and white children has widened and fewer African American children are fully vaccinated.

Factors Associated With African American Health Disparity

While we cannot pinpoint the specific causes of health disparity for specific diseases among members of particular social categories, we can say that race, class, and gender do not in and of themselves produce health disparities. Disparities in African American health status are explained by a number of factors, including income, lack of education, unemployment, differences in lifestyles and health behaviors, differences in environmental and occupational risks and hazards, nutrition, and cultural beliefs about health, as well as discrimination and access to health care. There is evidence that each of these factors plays a role in health disparities.

Studies by the Commonwealth Fund indicate that Hispanic and African American working-age adults have lower access to health care and a higher probability of facing medical debt than white working-age adults. The rate of uninsured African Americans is higher than that of whites but lower than that of Hispanics. Individuals without insurance are more likely to delay care and screenings, less likely to obtain needed medications, and more likely to be diagnosed at later stages of illness. These patterns of access and use of health care are associated with increased morbidity and mortality from disease.

African Americans are less likely to report that they have seen a physician in the previous year. Data indicate that African American adults are more likely to visit an emergency room for a condition that could have been treated by a doctor in an office setting if routine health care were available. In addition, African Americans are less likely to receive preventive health screenings and immunizations. For example, African American and Hispanics are less likely than

whites to receive influenza or pneumonia vaccines and rates of cholesterol and blood pressure screenings are also lower. Studies suggest that African American youth are more likely to eat fast food, which may include more calorie-laden options, as well as watch more television than white children, which is associated with reduced rates of physical activity.

Targeted health promotion campaigns have resulted in positive trends in cancer-screening behavior. For instance, African American women's use of mammography and cervical-screening procedures and the rates of colorectal cancer screening have increased, and their screening rates are now similar to those of whites. African American youth are less likely to smoke than youth of other ethnic populations, and African American men between the ages of 18 and 24 are less likely than white men to be current smokers.

According to the U.S. Census Bureau, approximately 25% of African Americans lived in poverty in 2005, and unemployment is high in the African American community. Even better-off African Americans are often the first generation in their family to achieve middle-class status and are likely to lack the wealth accumulated by white families of similar economic status, meaning they have fewer economic resources to access during an extended illness or health care crisis. And while rates of high school graduation and college completion have increased, African Americans continue to be less well educated than whites. Because of decreased educational levels and personal resources, the awareness of health problems, knowledge of causes and risk factors, and capacity to access medical care may be greatly decreased. Low income and lack of education are associated with increased morbidity and mortality from disease, increased obesity, decreased physical activity, lack of insurance and health care access, and low rates of physical activity.

Studies indicate that the socioeconomic explanation of disparities is limited. Even when income and education are controlled, African Americans and other minorities are more likely to receive care in the lowest-quality facilities with the least likelihood of appropriate follow-up and have more difficulty than majority group members in locating a usual source of care. Similarly, African Americans are more likely to be hospitalized for asthma regardless of income level, and maternal mortality is higher at both lower and upper educational levels.

Discrimination may affect disparities through the historical existence of social inequity and injustice in American education, justice, and economic structures and the disadvantages that have accrued to African Americans based on these differences. In addition, discrimination may influence disparities through the biases of health care institutions and providers. It has been suggested that providers may offer less intensive and sophisticated treatment options to African American patients due to stereotyped beliefs about ability to pay and willingness or ability to engage in and/or accept these services. These biases have been most clearly observed and documented for cardiac care.

Lack of information may be another important factor in African American health disparities. Among African Americans seeking health information, primary care physicians and other professional medical personnel are often perceived as the most credible information sources. However, many African Americans do not maintain an ongoing relationship with a primary care physician. Thus, there may be limited knowledge of risks. For example, surveys have found that African Americans are often unaware of the risk factors associated with diseases and their own personal risk.

Culture can affect health attitudes and behaviors, and it has been argued that African Americans are a population for whom culture might be relevant. Differences in beliefs about the nature and cause of disease, appropriate treatment, and who should provide treatment may influence the rate at which members of different groups access and adhere to treatment, as well as the extent to which they accept and participate in health preventive behaviors. For example, the acceptance of larger body sizes may affect African American concern over the increasing rate of obesity in the population, and traditional foods may contribute to a diet higher in fat. Male gender-role norms may influence the negotiation of condom use among African American men and women.

Culture also affects the interpretation of communications between patient and provider. Recognition of and respect for the communication preferences and patterns of African Americans may improve provider-patient communication. Culturally sensitive care requires that the health care provider adjust to the special needs and circumstances of the patient. In the case of African Americans, this may mean adjusting assessment to accommodate the varying skin tones noted among

African Americans or particular attention to the need to convey respect in interactions.

—Vetta L. Sanders-Thompson

See also Eco-Epidemiology; Health Behavior; Health Disparities; Social Hierarchy and Health

Further Readings

Brach, C., & Fraser, I. (2000). Can cultural competency reduce racial and ethnic health disparities? A review and conceptual model. *Medical Care Research & Review, 57*(Suppl 1), 181–217.

DeNavas-Walt, C., Proctor, B. D., & Hill Lee, C. (2006). *Income, poverty, and health insurance coverage in the United States: 2005* (U.S. Census Bureau, Current Population Reports, P60-231). Washington, DC: Government Printing Office.

Geiger, J. (2003). Racial and ethnic disparities in diagnosis and treatment: A review of the evidence and a consideration of causes. In B. Smedley, A. Stith, & A. Nelson (Eds.), *Unequal treatment: Confronting racial and ethnic disparities in healthcare.* Washington, DC: National Academies Press.

McKinnon, J. (2001). *The black population: Census 2000 brief.* Washington, DC: U.S. Census Bureau.

National Center for Health Statistics. (2006). *Health U.S. 2004, Chartbook, National Health and Nutrition Examination Survey*. Hyattsville, MD: Author. Retrieved March 10, 2007, from http://www.cdc.gov/nchs/products/pubs.

National Health and Nutrition Examination Survey. *Overweight among U.S. children and adolescents*. Hyattsville, MD: Centers for Disease Control, National Center for Health Statistics. Retrieved April 4, 2007, from http://www.cdc.gov/nchs/data/nhanes/databriefs/overwght.pdf.

Ward, E., Jemal, A., Cokkinides, V., Singh, G. K., Cardinez, C., Ghafoor, A., et al. (2004). Cancer disparities by race/ethnicity and socioeconomic status. *CA: A Cancer Journal for Clinicians, 54*(2), 78–93.

Watkins, K. (2005). *Human development report 2005. International cooperation at a crossroads: Aid, trade and security in an unequal world*. New York: United Nations Development Programme.

Witt, D. (2006). Health disparities in African American males. *Primary Care: Clinics in Office Practice, 33*(1), 35–43.

Agent Orange

Agent Orange is a herbicide used by American forces during the Vietnam conflict to remove leaves and other plant life that provided cover to enemy forces. It is a formulation of the two commercially available and widely used herbicides: 2,4-dichlorophenoxy acetic acid (2,4-D) and 2,4,5-trichlorophenoxy acetic acid (2,4,5-T). Before the Vietnam conflict, commercial mixtures of the two herbicides were routinely used worldwide in rangeland, rights-of-way, and forest management programs. During the Vietnam War, the 50:50 mixture of esters of the herbicides was applied in jungle areas to clear vegetation and expose enemy infiltration routes, base camps, and weapons placements, and to clear vegetation from the perimeters of friendly military bases and along lines of communication. The objective of the herbicide spraying program was to defoliate thick jungles that provided cover and concealment for the enemy, who would engage in ambushes and other disruptive tactics. During the period from 1965 to 1970, the U.S. Air Force applied more than 44 million liters of Agent Orange in South Vietnam. In 1969, the contaminant 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD or dioxin) in the 2,4,5-T herbicide was found to be teratogenic (causing birth defects) in laboratory animals, and in April 1970, the U.S. Department of Defense terminated all uses of Agent Orange in Vietnam. Although exposure to Agent Orange was once considered to be the cause of a variety of physical and emotional problems suffered by American veterans and Vietnamese people exposed to Agent Orange, independent studies have not documented any association between Agent Orange exposure and the health conditions claimed to have been caused by it.

Veterans Health Concerns

Hundreds of studies have been conducted of groups exposed to TCDD and/or the phenoxy herbicides 2,4-D and 2,4,5-T, either in the production of the herbicides or as end users in agriculture or forestry. In addition, the U.S. Air Force and the Departments of Veterans Affairs (DVA) in the United States and Australia have conducted studies of Vietnam veterans. These studies include the Air Force Health Study of the men who sprayed the herbicide from fixed-wing aircraft in Vietnam (commonly referred to as "Ranch Hand" personnel after the Air Force Operation Ranch Hand); the DVA study of personnel who served in the U.S. Army Chemical Corps and sprayed herbicide on base perimeters using helicopters and ground equipment; and the Veteran Studies by the Centers for Disease Control and Prevention (CDC). The results of these long-term epidemiologic studies of Vietnam

veterans have consistently shown few, if any, health effects that were not also seen in veterans who did not serve in Vietnam, or in those who were not affiliated with defoliation programs in Vietnam. Moreover, no chloracne, the hallmark of substantial dioxin exposure, was identified in any of the veteran studies.

The Agent Orange Act of 1991

After many years of denials of benefits to Vietnam veterans for lack of evidence that the diseases claimed had been caused by exposure to Agent Orange, Congress passed, and the president signed, the Agent Orange Act of 1991, Public Law 102–4.

There is no test that can show if a veteran's health problems were caused by Agent Orange or other herbicides used in Vietnam, and thus, by law, the Veterans Administration must presume that all Vietnam veterans were exposed to Agent Orange.

Consequently, the Agent Orange Act of 1991 allows benefits to be awarded based on presumptions of exposure and diagnosis with a disease based on an ''association'' with Agent Orange, so that proof of medical causation is not required. Congress provided for a series of reports of the scientific literature by the National Academy of Sciences' Institute of Medicine (IOM) applying the statutory ''association'' standard to assist the Department of Veterans Affairs in establishing by regulation a list of diseases for which Agent Orange benefits would be paid. The IOM was directed by Congress to shift the burden of proof away from veterans, and the IOM reports clearly identify their findings as associations that do not amount to a causal association. In fact, most of the evidence considered by the IOM as providing a basis for an association is not present in most veterans' cases. In accordance with the IOM findings, the following health conditions are presumptively recognized by DVA for service connection, meaning that their occurrence in a Vietnam veteran is presumed to have been caused by Agent Orange:

- Chloracne (must have occurred within 1 year of exposure to Agent Orange)
- Non-Hodgkin's lymphoma
- Soft tissue sarcoma (other than osteosarcoma, chondrosarcoma, Kaposi's sarcoma, or mesothelioma)
- Hodgkin's disease
- Respiratory cancers, including cancers of the lung, larynx, trachea, and bronchus
- Prostate cancer
- Acute and subacute transient peripheral neuropathy (must have occurred with 1 year of exposure and resolved within 2 years of date of onset)
- Type 2 diabetes, *Diabetes mellitus*
- Chronic lymphocytic leukemia

In addition, spina bifida (except spina bifida occulta) is recognized for presumptive compensation in children of Vietnam veterans.

The Scientific Evidence: TCDD Levels in Vietnam Veterans

Notwithstanding the major epidemiologic studies of Vietnam veterans and Agent Orange over the years, the highly relevant data for the important questions of exposure and actual absorbed dose have not been sufficiently appreciated in the literature. The development of increasingly sophisticated methods for biomonitoring low levels of chemicals in human tissues is regarded as one of the premier achievements of environmental science. Beginning in 1986, the CDC collected serum TCDD in a group of 646 U.S. Army veterans who served as ground troops in the regions of Vietnam most heavily sprayed with Agent Orange and also from 97 veterans who had not served in Vietnam. The distributions of TCDD levels were ''nearly identical'' in the two groups, both having means and medians that were well within the range of background levels at that time; that is, the levels were those that everyone had as a result of small amounts of TCDD principally ingested from food.

The CDC study had a statistical power of 99% to detect differences, even if only a moderate proportion of Vietnam veterans had elevated TCDD levels while they were in Vietnam. Neither military and spraying records nor self-reported history of exposure could reliably identify either high- or low-exposure groups. The investigators concluded that most U.S. Army ground troops were likely not exposed to any detectable levels of TCDD.

About 1,200 U.S. Air Force personnel directly handled Agent Orange in support of Operation Ranch Hand, the aerial spraying of herbicides in Vietnam from Air Force planes during the period from 1962 to 1971, and another 1,000 men in the Army Chemical Corps were responsible for spraying the perimeter of bases in Vietnam. Serum TCDD levels confirmed that these individuals had measurable exposure and

that their exposure was often heavy. Yet the long-term epidemiologic studies of these men, when compared with their matched comparisons, do not indicate any major differences in health status.

The Air Force Health Study's results concerning Ranch Hand are particularly informative as to possible health effects of Agent Orange, because the study design incorporated a comparison group of military aircrews with similar duties operating similar types of aircraft in Vietnam and surrounding countries, whose planes were engaged in transporting cargo rather than spraying Agent Orange. Thus, researchers were able to control for other possible confounding causes of disease in Vietnam veterans more effectively than the other major studies.

—Alvin L. Young

See also Admissibility of Scientific Evidence in Law; Carcinogen; Environmental and Occupational Epidemiology; Pollution

Further Readings

Centers for Disease Control. (1988). Serum 2,3,7, 8-tetrachlorodibenzo-*p*-dioxin levels in U.S. Army Vietnam era veterans. *Journal of the American Medical Association, 260*, 1249–1254.

Institute of Medicine. (2005). *Veterans and Agent Orange: Update 2004.* Washington, DC: National Academies Press.

SAIC for the Air Force Research Laboratory. (2006). *The Air Force health study: Final report*. Brooks City-Base, TX: Air Force Research Laboratory. Retrieved June 12, 2006, from http://www.ntis.gov.

Veterans Health Initiative. (2002). *Vietnam veterans and Agent Orange exposure.* Washington, DC: Employee Education System, Department of Veterans Affairs.

Young, A. L., Cecil, P. F., Sr., & Guilmartin, J. F., Jr. (2004). Assessing possible exposures of ground troops to Agent Orange during the Vietnam War: The use of contemporary military records. *Environmental Science & Pollution Research, 11*(6), 349–358.

Young, A. L., & Reggiani, G. M. (Eds.). (1988). *Agent Orange and its associated dioxin: Assessment of a controversy.* Amsterdam: Elsevier Science.

Aging, Epidemiology of

The epidemiologic study of older people in the community has been pursued for many years. Although there is probably nothing unique about the application of epidemiologic methods to older populations, there are many important differences and special issues that merit separate consideration. Vital records, census counts, and demographic studies have documented the increased longevity of populations in nearly all developed countries over the past half century and the growth of the numbers of older persons. The increase in older populations has led to new opportunities for more detailed exploration of health and disease occurrence, risk factors for morbidity and mortality, and health outcomes. The growth in numbers of older persons has also caused a public health mandate for improved surveillance and control of important conditions in those populations, both in the community and in institutional settings. This entry examines differences between epidemiologic studies of older adults compared with other age groups. It also discusses issues in conducting population surveys among older persons.

There are a number of general differences in the epidemiologic study of older populations from young and middle-aged groups, including differences in the clinical manifestations of disease and conditions among older persons, alterations in medical practice, and variations of the "natural history" trajectories and outcomes. As a generalization, diseases that occur in young and middle-aged persons are single entities and have few complicating secondary conditions, at least early in their history. The natural conditions occurring in youth and middle age tend to have more genetic influence. Noninvolved bodily systems tend to be generally intact, and most treatments relate to the primary condition. Social support systems for coping with these conditions are more often well developed. In contrast, health and disease states among older people differ in many ways.

1. *Multiple, Simultaneous Health Conditions (Comorbidity).* The presence of multiple health conditions and physiological dysfunctions is the rule in older people, whether an acute or severe disease is present or not. This necessitates that epidemiologic assessment of older people contain detailed information on a wide range of organ systems. Otherwise, many will not be detected. Having one condition may lead to increased medical surveillance for other conditions ("detection bias"), thus altering the natural history of the latter condition. Also, the presence of comorbidity requires that studies of disease causation consider how the conditions other than the one of primary interest affect the risk factors. For example, if

one is exploring the role of tobacco use in Condition X, it is important to know if comorbid Condition Y, which may have clinically emerged earlier, had an effect on smoking habits.

2. *Increased Prevalence of Common Clinical Signs and Symptoms.* Surveys of older persons reveal that prevalence rates for common clinical signs and symptoms are higher than in younger populations, having several implications for understanding health status. First, even if there is no obvious major acute or chronic illness present, these clinical phenomena may negatively affect personal comfort and function. Pain syndromes are notoriously present at high rates and have important consequences. Chronic leg cramps is an important, if less well-studied, example. Another implication of common signs and symptoms is that there may be a loss of diagnostic specificity. For example, if dry and itching skin is a potential indicator of clinical or subclinical hypothyroidism, possibly leading to a diagnostic inquiry, a prevalence rate of 40% among older persons may preclude those skin symptoms as potential diagnostic indicators, because relevant thyroid conditions are much less common. This then becomes a challenge for the clinician and the epidemiologist, because there may be less intense diagnostic activity for thyroid disease.

3. *Atypical or Delayed Clinical Presentations of Many Conditions.* In older persons, for reasons that are only partly understood, certain diseases may present clinically in different ways among elders than in other age groups. An important example is the presentation of myocardial infarction (MI), which may be more clinically "silent" among older persons—that is, it may occur without classical signs of chest pain and related symptoms. This is known from the coincidental detection of MIs in clinical settings and from serial electrocardiographic studies of older community populations. Thus, determining the frequency of MI in population surveys, based on a physician's report to a survey respondent ("Has a doctor ever told you that you have ...?"), may disproportionately underestimate the burden of this condition among older persons, as well as fail to appreciate alterations in the manifestations of coronary disease in general.

Another important alteration in the way a disease may present is the case of acute infections. It has been reported classically that among the oldest old, pneumonia may not present in the usual manner of typical symptoms such as cough, expectoration, fever, and malaise. Rather, elders may present later than usual and with atypical problems such as confusion or falling. This is important to the accurate surveillance of such infections, particularly nosocomial pneumonia, among institutionalized patients. The altered clinical presentation of pneumonia among this age group has been attributed in part to comorbid dementia (see below), but the principal problem remains.

4. *An Altered Ability to Acquire Clinical Information.* In population epidemiology, the designation of a case of a particular disease or condition is rarely determined by extensive or elaborate testing procedures in the field. Rather, such designations usually depend on report and documentation of standard diagnostic practices within the health care system. Thus, ostensible disease rates will vary according to access to and utilization of health care and the diagnostic practices of individual health professionals. However, as suggested above, usual diagnostic and therapeutic processes may be especially challenging for older persons. Increased rates of cognitive impairment, as well as sensory impairments (particularly decreased hearing and vision), may impede the ability to take a full medical history, explain illnesses, and invoke optimal self-care. Some of this may be appropriately alleviated by caregivers, but such persons may not always be available. Such impediments to the clinical process may alter diagnostic and therapeutic outcomes, and thus spuriously alter estimates of diseases and conditions in population studies.

5. *Altered Variation in Physiological and Biochemical Measures.* Part of the diagnostic process, in addition to medical histories and physical examinations, depends on laboratory determinations, such as blood cholesterol or hemoglobin levels, and physiological measures, such as blood pressure and electrocardiograms. However, an important issue for such applications among older persons is that the "normal" ranges of these measures may be different from those in middle age. This highlights the question of what is "normal" and whether changes in the distributions of these measures represent "normal" aging or subclinical diseases, or even differences in environmental exposures, such as diet or sunlight. An important example is the change in diastolic blood pressure (DBP) with age; population DBP increases until around age 70 and then decreases back toward the normal range. But this may be due to arterial stiffening, which may have its own adverse effects and is not a return to decreased risk. The conundrum of what is normal, biologically

obligate aging versus disease makes designation of normal ranges of laboratory determinations for older persons a difficult and philosophical issue of great import, and demands careful consideration of what clinical questions are being addressed by the test. As in other parts of the basic diagnostic processes, this problem may lead to variation in disease designation and alter the epidemiology of the conditions at hand for definitional reasons.

6. *Increased Vulnerability to Environmental Challenges and Threats.* Whatever the conceptual challenges of defining normal aging versus disease, older persons clearly have a lesser capacity for homeostasis, the ability to return to the normal or previous physiological state after an important environmental challenge. For example, death rates are increased among elders when extremes of heat and cold temperatures occur, as well as in some acute infections or even adverse social occurrences. These challenges will change morbidity and mortality patterns and the relative contributions of the risk factors responsible for them.

7. *Increased Emphasis on Characterization of Functional Status.* One of the great contributions of geriatrics and clinical gerontology has been to incorporate and refine measures of human function, including social, physical, cognitive, and mental, into clinical and epidemiologic lore. This better helps describe and characterize how individuals fare in their social and physical environments. Summaries of these measures across individuals allow further characterization of the functional status of older persons, and hence levels of disability and dependence, in variously defined groups, including those that are geographically referent. Measures such as activities of daily living, instrumental activities of daily living, and mobility have added greatly to the understanding of population health states.

From a clinical perspective, functional characterization of individuals allows increased accuracy of clinical trajectories and outcomes, as well as more precise determination of the type and extent of intervention regimens and programs. From a community and public health perspective, incorporating measures and interventions of social function can improve both health status and quality of life. Epidemiologists are particularly interested in causative factors for ill health, and despite the many successful applications of function status measures in general, they present a problem for etiologic research. This is because functional impairments or decrements are complex syndromes that are multicausal and not single biological entities. Thus, finding the "causes" of dysfunction can be challenging without the specificity that comes with studying individual illnesses.

8. *Altered Medical Care.* Many of the categories noted above have discussed forces that alter medical practice and thus the designation of disease status among older persons. There are additional factors that frequently occur among older persons that alter diagnosis and treatment. One is that many drugs are metabolized differently in older persons, and thus, the therapeutic process on balance may be altered in ways not fully characterized. Another dimension of many comorbid conditions is that health practitioners may understandably prioritize the conditions to be addressed at any one confrontation and devote the most time to those that are most severe or immediate. This too will change the natural history of some medical conditions. Because of substantial comorbidity, older persons may receive care from many different practitioners and health care venues, and thus, coordination and thorough documentation of all relevant clinical activities is a challenge to both medical care and epidemiologic characterization. Elders also spend more time residing in institutions, and with this may come greater exposure to nosocomially generated conditions. This can add to elder morbidity and mortality. Finally, there are many ethical and moral issues tied to geriatric care, each potentially altering the nature history of conditions and their outcomes.

9. *Altered Clinical Outcomes.* While not universal, older persons may have worse clinical outcomes than middle-aged persons presenting with various medical conditions at a similar stage. This has been particularly studied for many cancers and various heart diseases. There may be many reasons for this, including those cited above. Comorbidity, including age-related degenerative conditions and accompanying frailty, poly-pharmacy, and altered medical care approaches may be among the important reasons. These altered outcomes, however mutable, lead, ceteris paribus, to altered prevalence rates of diseases and possibly to different approaches to community disease control. Age-related alterations in disease outcomes may also necessitate age-specific benchmarking when comparing health care outcomes among institutions and when creating quality-of-care standards.

10. *A General Absence of Scientific Studies With Which to Guide Clinical and Public Health Interventions.* Older persons are less likely to participate in clinical and population research studies, and for a variety of historical reasons, many studies in prevention and treatment did not include many persons above the age of 70 years. Thus, the scientific basis for clinical and public policy among older persons is much less secure, and it is one of the frontiers of epidemiologic research.

Population Survey Research on Older Populations

Conducting surveys in communities to ascertain health states and their causes, as well as community needs and policy directions, requires special consideration and preparation. The following is a brief review of selected issues. Challenges in sampling older populations begin with deciding whether those in institutional or quasi-institutional settings are part of the target. In the United States, about 5% of persons live in long-term care settings, and many are disabled. Access to these settings for epidemiologic study may be difficult. Similarly, access to retirement or assisted living communities may require special techniques and resources beyond what is normally available.

Even when conducting interviews and related data collection activities in usual community settings, there are special problems. In general, community participation rates for epidemiologic studies decrease with increasing age; enhancing completeness of data collection often requires the use of surrogate respondents. Increased levels of disease and disability with increasing age lead to reluctance or inability to participate, only partly offset by the greater likelihood that these individuals may be more likely to be at home. Elders may be more embarrassed by their appearance or that of their households because of disabilities, further deterring participation. Elders with substantial illnesses are more likely to have caregivers and gatekeepers who may (appropriately) deny access, leading to the higher rates of proxy interviews and data collection with increasing age. Increasing rates of cognitive impairment among older respondents may decrease the quality of the information collected from such respondents and, equally important, raise issues on the ability of such persons to provide informed consent. This latter issue needs to be clarified with institutional review boards before embarking on data collection.

Consideration of the mode of data collection for older persons is important. Having increased levels of sensory or cognitive impairment may deter various types of data collection and increase the possibility that the collection activity will be assisted by others; this would be true of mailed questionnaires as well as telephone and personal interviews. Older people are increasingly likely to have access to computers and the Internet, and data collection by this means is increasingly feasible. However, such access rates are still lower than those for young and middle-aged adults, necessitating consideration of mixed-mode approaches. In telephone and personal interviews, attention to ability to hear and see collection materials, the possibility of fatigue, and disruptions due to illnesses or caregiving chores by the respondent are important. In the general interview setting, item nonresponse is generally not different from that in surveys of other age groups. However, the increased utilization of brief cognitive testing in surveys may cause some increased resistance compared with other health-related items.

—Robert B. Wallace

See also Alzheimer's Disease; Arthritis; Functional Status; Informed Consent; Public Health Surveillance

Further Readings

Anton, B., Vitetta, L., Cortizo, F., & Sali, A. (2005). Can we delay aging? The biology and science of aging. *Annals of the New York Academy of Sciences, 1057*, 525–535.

Colsher, P. L., & Wallace, R. B. (1989). Data quality and age: Health and psychobehavioral correlates of item nonresponse and inconsistent responses. *Journal of Gerontology, 44*, 45–52.

del Carmen, M. G., & Joffe, G. (2005). Informed consent for medical treatment and research: A review. *Oncologist, 10*, 636–641.

Klabunde, C. N., Harlan, L. C., & Warren, J. L. (2006). Data sources for measuring comorbidity: A comparison of hospital records and Medicare claims for cancer patients. *Medical Care, 44*, 921–928.

O'Neill, P. D. (1997). Aging homeostasis. *Reviews in Clinical Gerontology, 7*, 199–211.

Ostir, G. V., Carlson, J. E., Black, S. A., Rudkin, L., Goodwin, J. S., & Markides, K. S. (1999). Disability in older adults. 1. Prevalence, causes and consequences. *Behavioral Medicine, 24*, 147–156.

Rosin, A. J., & van Dijk, Y. (2005). Subtle ethical dilemmas in geriatric management and clinical research. *Journal of Medical Ethics, 31*, 335–339.

Unal, B., Capewell, S., & Critchley, J. A. (2006). Coronary heart disease policy models: A systematic review. *BMC Public Health, 6*, 213.

U.S. Census Bureau. (in press). *An aging world.*

Alcohol Use

Beverage alcohol results from fermentation of substances containing simple sugars and is catalyzed by yeast. Given that substances containing sugar and yeast occur widely in nature, alcohol was "discovered" early and often in the development of civilization. The intoxicating properties of alcohol have contributed to its widespread use, as have its medicinal properties. However, alcohol is one of few substances voluntarily consumed in amounts approaching those causing coma and death. Acute overconsumption can lead to unintentional alcohol poisoning and increase the likelihood of accidental injuries. Chronic overconsumption can cause social, legal, and medical problems, including psychiatric disorders. Thus, misuse of alcohol represents substantial suffering by individuals and a major societal burden. This entry reviews the epidemiology of alcohol use, alcohol problems, alcohol abuse and dependence, and alcohol-related health consequences.

Physiological Properties of Alcohol

Alcohol is rapidly absorbed from the gastrointestinal tract, enters the bloodstream, and is distributed in the body water compartment. Most alcohol is metabolized to acetaldehyde by alcohol dehydrogenase in the liver. However, chronic heavy drinking may induce added metabolic capacity by the microsomal ethanol-oxidizing system, an enzyme sited predominantly within hepatocyte microsomes that degrades a number of drugs. Acetaldehyde, 100 times more toxic than alcohol, is rapidly metabolized by acetaldehyde dehydrogenase to water and carbon dioxide in most people, but it may accumulate in those who have inherited a less efficient form of acetaldehyde dehydrogenase, giving rise to a dysphoric "flushing syndrome" most prevalent in Asians and often associated with abstention and light drinking patterns.

In general, alcohol has a biphasic effect on mood. As it is absorbed from the gastrointestinal tract, rising blood alcohol levels (BALs) are usually associated with elevated mood. In contrast, as alcohol is metabolized and BALs fall, people are more likely to feel dysphoric and, in some cases, drink more, seeking to regain initial feelings of well-being. Within these broad parameters, there is substantial individual variability in response to alcohol. Studies of subjective response to alcohol indicate that nonalcoholic sons of alcoholics are less affected by a given alcohol dose than nonalcoholic sons of nonalcoholics. Paradoxically, innate tolerance (experiencing fewer effects during initial drinking) appears to be a risk factor for developing alcohol dependence.

Monitoring Alcohol Use

Per capita alcohol consumption, based on alcohol sales and census data, is a standard ecological measure of alcohol use. It provides both cross-jurisdictional (across states and nations) and cross-time-series data for epidemiologic studies of alcohol's impact on rates of morbidity and mortality. Efforts to improve state-specific estimates in the United States include adjusting for abstention rates and taking into consideration the ethanol content of specific beverage types and their market shares in each state.

National surveys dating from the early 1940s provide individual-level data on alcohol use to complement ecological data, and the number, size, and sophistication of alcohol surveys have increased dramatically in recent decades. The National Institute on Alcohol Abuse and Alcoholism (NIAAA) was established in 1973 to lead national efforts to reduce alcohol-related problems. The Alcohol Research Group, an NIAAA Alcohol Research Center, has conducted periodic National Alcohol Surveys since 1964 at approximately 5-year intervals with standardized measures since 1979. Special Alcohol Supplements have been administered in conjunction with several National Health Information Surveys, and the National Health and Nutrition Examination Surveys also include questions on alcohol use. Similarly, questions on alcohol use are asked in the annual National Household Survey on Drug Abuse, sponsored by the National Institute on Drug Abuse. The largest U.S. alcohol surveys today are those conducted by NIAAA's intramural program—the National Longitudinal Alcohol Epidemiologic Survey and the National Epidemiologic Survey of Alcohol-Related Conditions.

Self-report measurement of consumption, though known to underreport alcohol sales, has been

improving with increased methodological rigor. The addition of questions on patterns of drinking providing information on heavy drinking episodes and assessing ethanol content of beverages consumed improves both coverage and estimates of alcohol exposure. Comparability of survey findings is hampered by variability in the measures employed, and relatively little information is available on lifetime patterns of alcohol use.

Factors Influencing Alcohol Use in the United States

Long-Term Trends

The use of alcohol over the past century has been characterized by oscillating secular trends characterized by increases in consumption until this produces negative consequences severe enough to precipitate a backlash in policy enactments and preventive measures that, it has been argued, lead to the subsequent periods of declining drinking. In the United States, an extreme example is the passage and ratification of the Prohibition Amendment, taking effect in January 1920. Although this greatly reduced consumption and alcohol-related problems such as cirrhosis mortality, alcoholic psychosis, and "drunk and disorderly" arrests, increases in criminal activity associated with the production and sale of illegal liquor and social discontent led to repeal in 1933. Per capita consumption increased in subsequent decades, particularly after the 1960s as postrepeal laws were eroded, reaching a peak of nearly three gallons of absolute alcohol per person aged above 14 by 1981.

In the 1980s, concern regarding drunk driving mobilized advocacy groups such as Mothers Against Drunk Driving and Students Against Drunk Driving. Effective campaigns for stronger penalties, more stringent law enforcement, and policies limiting youth access to alcohol followed. A result was federally enacted incentives leading all states to raise the legal drinking age to 21 years, and more recently to reduce the threshold defining driving while intoxicated to a uniform .08 BAC (blood alcohol content). By the late 1990s, per capita consumption reached a new nadir close to that seen in the 1950s, just above two gallons of alcohol. Since then it has been inching up again, starting an apparent new wave. Lifetime drinking patterns of birth cohorts are influenced by these secular fluctuations in the "wetness" or "dryness" of prevailing alcohol attitudes and policies.

Alcohol Use Over the Life Course

Most U.S. individuals who drink alcohol begin in adolescence or early adulthood. Among those who drink regularly, frequency of consumption tends to increase and quantity (i.e., drinks per drinking day) tends to decrease with age. Thus, drinking by older teens and young adults is often characterized by heavy episodic consumption. The overall amount of alcohol consumed, sometimes referred to as volume and expressed as average drinks per day, is often used to summarize measures of alcohol intake. However, it has the disadvantage of obscuring drinking patterns; for example, frequent heavy drinkers who have seven drinks per drinking day once a week have the same volume as moderate drinkers who have one drink per drinking day seven times a week. The importance of taking drinking patterns into account when assessing the consequences of alcohol use is being increasingly recognized. Such patterns can interfere with important developmental achievements and are associated with automobile crashes and other injuries. Although the majority "mature out" of risky drinking patterns as they take on adult social roles involving work and family responsibilities, others maintain or increase heavy intake, thus increasing their risk of social, legal, and work-related alcohol problems, as well as acute and chronic health problems. In addition to the typical moderation seen with aging (i.e., more frequent consumption of smaller quantities), abstention rates increase. Factors associated with these reductions in alcohol intake include age-related changes in body composition (i.e., greater percentages of body fat relative to body water, resulting in higher BALs associated with a given alcohol intake), increased use of medications that contraindicate alcohol use, chronic conditions that impair the ability to cope with alcohol effects, and fewer social opportunities for drinking. In addition, some people attempt to stop abusive use of alcohol as they get older.

In the United States, as in virtually all countries, men drink more than women and suffer more alcohol-related problems. However, women tend to be more vulnerable to alcohol effects, both socially,

because of gender-related stigmas regarding intoxication, and physiologically. In general, for a given alcohol intake, women achieve higher BALs than men but may offset this by tending to drink more slowly. Some metabolism of ethanol takes place in the gastrointestinal lining, preventing it from reaching the bloodstream, a process more efficient in men than women. More important, women tend to have smaller cellular water compartments in which to distribute the alcohol they absorb, both because they tend to be smaller and because of a higher percentage of body fat. Finally, there is some evidence to suggest that even when BALs are held constant, women may respond more strongly than men to alcohol.

Race and ethnicity play an important role in overall consumption and in the time course of drinking. African Americans, Asian Americans, and Hispanics have higher abstention rates (especially for women) than Caucasian counterparts. African Americans and Hispanics who do drink tend to drink more heavily and may less often reduce consumption in middle age, shifting upward the age distribution of alcohol problems. Certain American Indian and Alaska Native groups have extremely heavy and dependent drinking patterns, which seem to be influenced by genetic as well as current and historical environmental factors. As indicated earlier, Asian Americans may be partially protected by the flushing response. Although broad generalities are often made about drinking patterns of minority groups, these patterns vary markedly with sociodemographic characteristics, Indian tribal affiliation, and Hispanic and Asian country of origin, acculturation, and immigrant status.

Alcohol Use Around the World

In countries where wine is traditionally part of the diet (e.g., France and Italy), high cumulative alcohol intakes have been associated with high rates of chronic alcohol-related disease such as cirrhosis but with relatively low rates of social problems related to drinking. In contrast, Nordic drinking is characterized by heavy episodic drinking on weekends and festive occasions, relatively low total consumption, and high rates of acute problems such as injuries and violence. In Russia, lax controls over drinking and a tradition of drinking to excess led to a dramatic increase in alcohol-related mortality in the period after perestroika. Italy has reduced per capita consumption considerably in recent years, and today more young people are consuming beer, leading to some homogenization of European drinking cultures.

With the proliferation of "new" products as part of the global expansion of alcohol multinationals, there is concern about rising rates of alcohol use in developing countries. In general, consumption of traditional beverages, while often continuing, has been overlaid with heavily marketed manufactured products, mostly beer and spirits. Increasing heavy episodic drinking conjoined with economic development and burgeoning vehicular traffic has led to sharp increases in accidents and injuries.

Alcohol Misuse

Alcohol-Related Problems

The burden represented by chronic and acute health effects associated with alcohol abuse is roughly equivalent to the burden represented by social, mental, and emotional consequences. These include violence, vandalism, public disorder, interpersonal difficulties, financial and work-related problems, and reduced educational performance. In U.S. surveys, it has long been noted that higher problem rates are seen among drinkers who consume very large quantities fairly often (e.g., 12+ drinks on a single day, at least monthly) and those who engage in frequent heavy drinking (e.g., 5+ drinks on a single day, on a weekly basis).

According to the *USDA Dietary Guidelines for Americans 2005*, moderate drinking is defined as the consumption of up to one drink per day for women and up to two drinks per day for men. Twelve fluid ounces of regular beer, 5 fluid ounces of wine, or 1.5 fluid ounces of 80 proof distilled spirits count as one drink for purposes of explaining moderation. This definition of moderation is not intended as an average over several days but rather as the amount consumed on any single day. Although heavy drinkers have higher problem rates, the so-called moderate drinkers contribute most to the total number of alcohol problems because there are so many more of them. This has been termed *the prevention paradox* because prevention efforts targeting heavy drinkers may not have as great an impact on the overall

number of alcohol-related problems as interventions targeting moderate drinkers.

Alcohol Abuse and Dependence

Alcoholism was defined as a disease in the 1950s, having previously been thought of as a bad habit by many. Alcohol abuse and dependence are psychiatric disorders described in the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition, published by the American Psychiatric Association in 2000. Impairment of self-control is still regarded as a key element of alcohol use disorders. In the psychiatric epidemiologic tradition, survey methods assess alcohol abuse and dependence using structured interviews. A classic example is the Diagnostic Interview Schedule, developed in the 1980s for use in the NIMH-sponsored Epidemiologic Catchments Area studies. It allowed lay interviewers to obtain measures of alcohol abuse and dependence. These surveys revealed that the general adult household population prevalence of meeting criteria for an alcohol disorder was vastly greater than the number in alcohol treatment. People so diagnosed were much younger than clinical cases, giving rise to the term *the two worlds of alcoholism*—the world of untreated alcoholics and the world of alcoholics who enter treatment. These discrepancies directed attention to the phenomenon of ''natural recovery'' from alcoholism (i.e., recovery without formal treatment) and stimulated the development of programs to increase screening and referral for alcohol problems, for example, through emergency departments, employee assistance, and drunk driving programs. More recently, the U.S. Preventive Services Task Force has recommended screening and brief counseling interventions addressing alcohol abuse in primary care settings, and NIAAA has published guidelines to aid such providers.

Numerous risk factors for alcohol disorders have been investigated (e.g., genetic, neurological, psychological, social, and environmental availability). Several alcohol typologies have been proposed based on findings from these studies, and the life course of alcohol disorders is highly variable. Many individuals who have alcohol experiences consistent with a diagnosable disorder appear able to respond by abstaining or moderating their drinking, usually with no formal treatment. In contrast, others struggle for years, need a great deal of support to abstain or cut down, and continue to be at risk of relapse when stressed. Early onset of alcohol use and early intoxication are associated with a greater likelihood of later alcohol dependence; however, prior conditions such as family disruption, conduct disorder, multiple-risk-taking dispositions, and genetic factors probably contribute to this relationship.

Alcohol-Related Health Consequences

Heavy drinking for years is a necessary and sufficient cause of conditions such as alcoholic liver disease and fetal alcohol syndrome, and it increases the likelihood of numerous other health problems. Causal relationships have been demonstrated between alcohol consumption and more than 60 disease conditions and disabilities, including traffic fatalities and other kinds of injuries, homicide, and suicide. In one method of estimating the burden of disease caused by alcohol in a given population, epidemiologic studies identify the fraction of a given disease that can be attributed to alcohol, the alcohol-attributable fraction (AAF). Thus, 100% of alcoholic liver disease is attributable to alcohol, but a relatively small percentage increase in breast cancer is associated with alcohol use. To estimate burden, AAFs may be applied to subsets of the population having the relevant diseases and summed. Another measure of alcohol's impact on society sums years of potential life lost (YPLL)—age at death from an alcohol-related cause subtracted from 65 years. Because many alcohol-related deaths, such as those resulting from automobile crashes and even cirrhosis, occur in relatively younger individuals compared with victims of heart disease and cancer, alcohol-related YPLLs are higher than would be expected given their mortality rates. Contemporary work sponsored by the World Health Organization has quantified the burden of disease attributable to alcohol and other factors, finding that the total global burdens of disease for alcohol (4.0%) and tobacco (4.1%) are on a par.

Alcoholic Liver Disease

Mortality data on liver cirrhosis serve as a surrogate for monitoring alcoholic liver disease because alcohol abuse is often omitted from death certificates to spare the family. Age-adjusted cirrhosis mortality rates dropped during Prohibition, peaked during the 1970s, and have fallen since then. That mortality rates began falling before per capita alcohol consumption decreased has been attributed to improved

access to and effectiveness of alcoholism treatment. Cirrhosis mortality was substantially higher among blacks compared with whites in 1970, but age-adjusted rates have fallen more rapidly among blacks, such that rates are now similar. In contrast, cirrhosis mortality rates remain higher among white Hispanics than blacks or whites. Not only is heavy drinking a primary cause in alcoholic liver disease, it also exacerbates liver damage from other factors such as chronic hepatitis C virus (HCV) infections and obesity, often co-occurring with these conditions. Increases in liver cirrhosis are anticipated as the cohort most affected by the HCV epidemic between the 1960s and 1980s reaches ages at which liver disease usually manifests, exacerbated by the epidemic increases in early onset obesity.

Cardiovascular Disease

The relation of alcohol use to cardiovascular disease (CVD) is complex. Numerous studies have reported J-shaped or U-shaped alcohol dose-response curves with CVD endpoints, suggesting that drinking moderately may be protective. Others have argued that such curves may be artifacts related to unadjusted confounding factors (e.g., a generally healthier lifestyle among moderate drinkers compared with abstainers, a tendency of sicker persons to abstain or reduce intake, and failure to take drinking patterns and lifetime drinking trajectories into consideration).

—Marcia Russell and Thomas K. Greenfield

See also Cancer; Cardiovascular Disease; Injury Epidemiology; Intimate Partner Violence; Suicide; Vehicle-Related Injuries; Violence as a Public Health Issue

Further Readings

Babor, T., Caetano, R., Casswell, S., Edwards, G., Giesbrecht, N., Graham, K., et al. (2003). *Alcohol: No ordinary commodity. Research and public policy.* New York: Oxford University Press.

Clark, W., & Hilton, M. (Eds.). (1991). *Alcohol in America: Drinking practices and problems.* Albany: State University of New York Press.

Greenfield, T. K. (2001). Individual risk of alcohol-related disease and problems. In N. Heather, T. J. Peters, & T. Stockwell (Eds.), *International handbook of alcohol dependence and problems* (pp. 413–438). New York: Wiley.

Room, R., Babor, T., & Rehm, J. (2005). Alcohol and public health. *Lancet*, *365*, 519–530.

Russell, M. (2001). *The epidemiology of alcohol use, alcohol abuse and dependence, and alcohol-related morbidity and mortality*. Retrieved September 26, 2006, from http://rsoa.org/lectures/01/index.html.

Web Site

National Institute on Alcohol Abuse and Alcoholism Publications: http://www.niaaa.nih.gov/Publications. See especially *Alcohol Research and Health* and *Surveillance Reports.*

Allergen

An allergen is a molecule that stimulates an allergic response in a sensitive individual. Allergens are almost always proteins; however, not all proteins are allergens. A protein that acts as an allergen displays two fundamental properties: induction of an IgE immune response on first exposure and induction of a clinical response to the same or similar protein on subsequent exposures. Some examples of allergens include dust mites, peanuts, pollen, and pet dander; however, a comprehensive list of allergens is impossible to assemble, as the list of possible substances is extensive (i.e., food, air particles, drugs, animal products, and insect stings). Sensitivities vary greatly from one person to another; and the response to allergens also varies in severity from allergic rhinitis and hives to asthma and anaphylaxis. Allergen sensitivity testing is used in epidemiology studies to characterize atopy and its association with allergic diseases in populations.

An allergen provokes an allergic response in the immune system by acting as a substance that stimulates an IgE-mediated antibody response. On first exposure to the allergen, IgE antibodies are developed that specifically recognize the allergen. These antibodies are then bound to mast cells, a cell type that is very common in our gastrointestinal and respiratory systems. If the body is exposed to the allergen again, the mast cell-bound IgE antibody recognizes the antigen and causes the mast cells to release substances (e.g., histamine, leukotrienes, and interleukins) that cause cell damage and inflammation, resulting in the symptoms we commonly associate with allergies such as a runny nose, congestion, hives, and swelling. Atopy refers to the condition of

raised IgE antibody levels following exposure to common allergens.

Exposure to an allergen can occur through inhalation (e.g., pollen), ingestion (e.g., peanuts), or direct contact (e.g., latex gloves). If the allergen is in the air, the allergic reaction will likely occur in the eyes, nose, and lungs. If the allergen is ingested, the allergic reaction will likely occur in the mouth, stomach, and intestines. An allergic reaction can occur throughout the body, such as hives or analphlaxis, if the reaction is severe enough. Many protein allergens have been sequenced, but there are only a handful of common characteristics associated with these allergenic proteins. Several allergen types (food and plant) belong to only a few of the thousands of protein families, suggesting common structural features. Other common features include a high dose of the protein in the ingested or inhaled substance and resistance to digestion, two features that make immunologic resistance more difficult.

The goal of diagnosing an allergy is identifying the causative allergen; IgE antibodies specific to a particular allergen are measured in the blood in RAST (radioallergosorbent) testing, or skin-prick testing (SPT) is performed. SPT involves the controlled application of a variety of allergens and positive and negative controls to the skin. Itchy, red skin indicates a positive response to the applied allergen—that is, an IgE-mediated reaction. SPT has been used in a variety of ways in epidemiologic studies to characterize atopy in populations, including to identify the prevalence or incidence of reactivity to specific allergens in predisposed populations such as asthmatics; to identify a sensitive population for use in a case-series or case-control study; to examine the trend of skin test reactivity over a particular time frame in a population; to calculate the association of specific skin reactivities with particular allergy types (e.g., asthma or chronic obstructive pulmonary disorder) or the severity of those allergies; and to calculate the association between reactivity to an allergen and the levels of that allergen in an environment (e.g., pollen reactivity and pollen counts).

Allergic diseases are common, with an estimated 40 million persons in the United States suffering from some type of allergic disease. Correspondingly, the reported prevalence of positive skin test responses to common allergens is relatively high in epidemiologic studies. For example, the National Health and Nutrition Examination Survey (NHANES), a survey designed to be representative of the general U.S. population, conducted SPT testing using 10 common allergens among participants interviewed from 1988 through 1994. The majority of the participants (i.e., 54.3%) had at least one positive skin test response, with the most common positive responses occurring with exposure to dust mite, cockroach, perennial rye, and short ragweed allergens. Males tend to be more likely to have a positive skin test response and higher IgE levels, compared with females. Also, skin test reactivity and IgE levels tend to peak in young adulthood and decline in later years.

The vast majority of asthmatics are sensitized to at least one common allergen, including pet, rodent, and dust mite allergens. Allergen sensitization is a condition that is highly associated with prevalent asthma, as well as a risk factor for the future development of asthma. Allergen sensitization is thought to be an intermediate factor in the causal pathway leading to asthma. Also, allergen sensitization appears to modify the effects of other exposures, meaning that atopic persons are more likely to develop asthma symptoms than nonatopic persons, given the same level of allergen exposure.

The treatment for allergies involves avoidance or controlled exposure to the offending allergen, the administration of drugs that counteract the effects of the substances released by mast cells (e.g., antihistamines), and allergen immunotherapy (AIT). AIT is the controlled, systemic administration of an ''allergy vaccine'' composed of a specific mixture of allergens that is designed to modify the allergic mechanisms and ultimately desensitize the person to the allergen.

—Meghan E. Wagner and Janci L. Chunn

See also Asthma; National Health and Nutrition Examination Survey

Further Readings

Akdis, C. A. (2006). Allergy and hypersensitivity. Mechanisms of allergic disease. *Current Opinion in Immunology, 18*, 718–726.

Arbes, S. J., Gergen, P. J., Elliott, L., & Zeldin, D. C. (2005). Prevalences of positive skin test responses to 10 common allergens in the U.S. population: Results from the Third National Health and Nutrition Examination Survey. *Journal of Allergy and Clinical Immunology, 116*(2), 377–383.

Pearce, N., Beasley, R., Burgess, C., & Crane, J. (1998). *Asthma epidemiology: Principles and methods*. New York: Oxford University Press.

Alzheimer's Disease

Alzheimer's disease (AD) is a progressive, degenerative brain disease that impairs memory, thinking, and behavior. AD, named after Alois Alzheimer, a German neurologist who published his observations of a patient, Augusta, D., in 1906, is the most common form of dementia, comprising 60% of all dementias. AD has been previously known as dementia of the Alzheimer type, senile dementia of the Alzheimer type, and Alzheimer's dementia.

AD currently affects about 4.5 million men and women in the United States, with an annual cost of $100 billion, and the number of persons with AD is expected to rise to 16 million by the year 2050. The incidence of AD increases with age, affecting up to 50% of people above the age of 85, although rare in those below the age of 60. There are other forms of dementia not related to AD, such as dementia with Lewy bodies (20%), vascular dementia (15%), and rare forms such as frontal-temporal dementia (5%) with women at greater risk for AD than vascular dementia. There is no known cause of AD, although it appears likely that a combination of factors, including age, genetic inheritance, environmental factors, diet, and overall general health, is responsible.

Although the first symptoms of AD are often confused with the changes that take place in normal aging, AD is not a normal part of aging and is caused by brain pathology. The brain of an affected individual may begin to deteriorate as much as 10 to 20 years before any visible signs or symptoms of AD appear. Typical symptoms of the early stages of AD may be attributed to many causes, making initial diagnosis difficult; they include memory loss, behavioral symptoms, emotional changes, and changes in judgment and decision making.

Over time, AD progresses through three main stages: mild, moderate, and severe. Memory impairment is a necessary feature for the diagnosis of AD or any type of dementia. Change in one of the following areas must also be present: language, decision-making ability, judgment, attention, and/or other areas of mental function and personality. The rate of progression is different for each person. If AD develops rapidly in an individual, it is likely to continue to progress rapidly. If it has been slow to progress, it will likely continue on a slow course.

Generally, the onset of AD is insidious, with failure of memory of recent events, emotional changes, depression, anxiety, and unpredictable behaviors being the earliest symptoms, appearing up to 3 years prior to diagnosis. Functional and behavioral problems may be evident and may increase anytime within 1.5 to 6 years after diagnosis, followed by progressive apathy, space perception disorders, a shuffling gait, slow and awkward movements, jerky muscle contractions (myoclonus), and irreversible loss of speech and memory. AD eventually progresses to a late vegetative phase consisting of complete inability to think, move, or speak. The patient usually dies of pneumonia, heart attack, or stroke.

While clinical symptoms and a thorough examination resulting in a diagnosis of probable AD are highly correlated with postmortem examination, a definitive diagnosis is made only after a postmortem examination. Postmortem examination of an AD patient will reveal a loss of cells in all cortical layers except the motor cortex and a degeneration of neurofibrils, the filaments found in and around nerve cells. Neurofibrillary degeneration and plaques, composed of amyloid protein, are distinctive histopathological features of the cerebral cortex in AD. The amyloid deposition is thought to be due to an abnormality in the amyloid precursor protein. This abnormality is influenced by several factors, including higher levels of stress hormones and the four alleles of the apolipoprotein E (APOE) genotype.

Risk Factors

Several factors may protect against AD, while others increase susceptibility to AD. Genetic differences appear to play a role in susceptibility to AD. The fact that a high percentage of Down syndrome patients develop AD suggests a possibility of involvement of chromosome 21 in the development of AD. Familial clustering and apparent generation-to-generation transmission are suggestive of autosomal dominant inheritance in rare cases such as familial Alzheimer's disease. Other genetic traits that are associated with higher rates of AD are the presenilin 1(PS1) gene, APOE polymorphisms, and alpha-1-ntichymotrypsin (ACT), an inflammatory protein. APOE polymorphisms occur in either homozygous (3/3) or heterozygous

(2/3) form. The polymorphism containing the four alleles in either heterozygous or homozygous form tends to increase susceptibility to AD and can interact with ACT and amyloid precursor protein to increase the risk and clinical progression of AD.

In addition to genetic factors, interaction between genes and environment can lead to increased risk. One such factor is chronic stress leading to glutamate excitotoxicity and increased amyloid precursor protein. Other factors associated with increased susceptibility to AD include depression, cerebrovascular disease, history of head injury, high cholesterol levels, a prolonged inflammatory response, and oxidative free radicals.

Several factors have been identified that may be protective for AD. These include consumption of antioxidant foods such as curcumin, found in the spice turmeric; exercise; use of nonsteroidal anti-inflammatory medication (NSAIDS) such as aspirin; and the APOE 2 allele.

Diagnosis

To obtain a diagnosis of AD, a person must meet several criteria. These include dementia confirmed by medical and psychological exams; problems in at least two areas of mental functioning; progressive loss of memory and other mental functions; and no other disorders that might account for the dementia, such as hypothyroidism, depression, overmedication, drug-drug interactions, and vitamin B_{12} deficiency.

Tools used to diagnose AD include the Mini-Mental State Examination (MMSE), Clock Test, Functional Assessment Staging (FAST), Alzheimer's Disease Assessment Scale, Cognitive Subscale (ADAS-Cog), Severe Impairment Battery (SIB), Modified Alzheimer's Disease Cooperative Study (ADCS), Neuropsychiatric Inventory (NPI), and the Clinical Rating Scale.

One of the tools frequently used to screen for overall cognitive decline is the MMSE (scored 0 to 30; mild decline is indicated by a score of 21 to 25, moderate 9 to 20, and severe 0 to 8) consistent with the Clinical Dementia Rating (CDR) scoring of 1.0 for mild and 2.0 for moderate dementia. The MMSE consists of 11 questions that cover five cognitive areas: orientation, registration (ability to recognize and name specific items), attention, recall, and language.

The Clock Test, an easy-to-administer indicator of cognitive decline, is used to differentiate between depression and stages of AD. Patients are asked to draw a clock, including all the numbers and a specific time. Scoring for this test includes the numbers drawn, location of the numbers, and location and size of clock hands.

The FAST is used to determine the stage (mild, moderate, or severe) of AD rather than for diagnosis. It assesses a range of activities, including dressing, continence, and ability to speak, sit up, and smile.

Several tests are used more commonly in clinical research regarding AD. The ADAS-Cog, used to gauge change in cognition with a focus on memory and language, is a highly accurate scale in diagnosing and staging mild to moderate AD. One of the limitations of this scale is a "floor effect" (the inability of the test to identify worsening conditions after patients reach the bottom of the scale), which makes it ineffective in measuring severe AD.

The SIB is used to assess the cognitive functioning of severely impaired persons who are unable to take other standardized cognitive scales. It consists of 40 questions (some with multiple parts) that measure cognitive range in orientation, language, memory, and attention. This test is used to assess patients in the moderate to severe stages of AD.

Developed through the modified ADCS, the Activities of Daily Living Inventory (ADCS-ADL) is used to measure functional capacity over a broad range of dementia severities. Patients are evaluated on their response to questions designed to determine their ability to perform specific ADLs, such as bathing, dressing, eating, and walking.

The Behavioral Rating Scale for Geriatric Patients (BGP) assesses both functional and behavioral disturbances in geriatric patients. Assessments include physical disabilities, abilities to perform ADLs, and level of activity versus inactivity.

The NPI evaluates behavioral and psychiatric disturbances with a 12-item questionnaire. Items include delusions or paranoia, hallucinations, agitation or aggression, depressed mood, anxiety, elation or euphoria, apathy or indifference, disinhibition, irritability, motor disturbance, nighttime behaviors, and appetite problems.

The Clinicians' Interview-Based Impression of Change Plus Caregiver Input (CIBIC-Plus), used to interview both patient and caregiver, measures the overall improvement or decline of a patient's cognitive function through a series of interview questions.

Treatment

Treatment of AD is symptomatic and is directed toward preventing further deterioration, given the current pathophysiologic understanding of the disease. Pharmacologic agents for cognitive symptoms include cholinesterase inhibitors and NMDA receptor antagonists. Cholinesterase inhibitors are directed toward increasing the concentration of the chemical messenger acetylcholine and keeping it high so that neural transmission will proceed in the face of the loss of cholinergic neurons found in mild to moderate AD.

NMDA receptor antagonists regulate levels of glutamate, an excitatory messenger chemical that the brain uses to process, store, and retrieve information. However, an excess amount of glutamate excitotoxicity has been implicated in increased susceptibility to AD. This class of medications is used most extensively to treat those with moderate to severe AD, slowing the progression and maintaining functional ability.

Most herbal therapies are directed toward the antioxidant effect that these substances provide in an attempt to reduce oxidative free radicals. While they may be effective in several individuals with AD, these substances are not regulated, and their efficacy has not been established. Two commonly used treatments are coenzyme Q10 and *Ginkgo biloba*. Neither of these can be recommended at this time to decrease or prevent further decline in AD. In addition, *Ginkgo biloba* may be dangerous for some individuals as it has anticoagulant properties and may compromise blood clotting. AD patients often display behavioral symptoms such as irritability, anxiety, sleep disturbances, agitation, restlessness, and pacing, all of which are distressing to the patient and their caregivers, frequently resulting in institutionalization. Events that may trigger these behaviors include a change in caregiver or living arrangement, travel, and bathing.

Given the side effects and limited efficacy of pharmacologic interventions such as antipsychotics used to treat these symptoms, the American Neurological Association currently recommends the use of nonpharmacologic interventions. Several of these interventions have met with varying success and are directed toward increasing relaxation and decreasing stressful environments. These include music therapy, therapeutic touch, aromatherapy such as lemon balm, pet therapy, increased pleasant activities, and simulated presence. Simplifying the environment and providing predictable structure and routine can decrease these behaviors. In addition, improving the therapeutic communication and decreasing the stress of formal and informal caregivers can help to decrease these symptoms.

—Diana Lynn Woods

See also Aging, Epidemiology of; Complementary and Alternative Medicine; Etiology of Disease; Gene-Environment Interaction; Intervention Studies; Psychiatric Epidemiology; Stress

Further Readings

Blennow, K., deLeon, M. J., & Zetterberg, H. (2006). Alzheimer's disease. *Lancet, 368*, 387–403.

Cummings, J. (2005). Clinical evaluation as a biomarker for Alzheimer's disease. *Journal of Alzheimer's Disease, 8*, 327–337.

Doody, R. S., Stevens, J. C., Beck, C., Dubinsky, R. M., Kaye, J. A., Gwyther, L., et al. (2001). Practice parameter: Management of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology, 56*(9), 1154–1166.

Khachaturian, Z. (2000). Epilogue: Toward a comprehensive theory of Alzheimer's disease: Challenges, caveats, and parameters. *Annals of the New York Academy of Sciences, 924*, 184–193.

Woods, D. L., Craven, R., & Whitney, J. (2005). The effect of therapeutic touch on behavioral symptoms of persons with dementia. *Alternative Therapies in Health and Medicine*, *11*(1), 66–74.

American Cancer Society Cohort Studies

The American Cancer Society has conducted numerous cohort studies that provide a wealth of data to help in the fight against cancer. The first of these, the Hammond-Horn Study, was initiated in 1952. It was a prospective cohort study designed to research the effects of tobacco smoking on cancer death rates, as well as death rates from other diseases. With the help of volunteers, 188,000 adult men were recruited for this study and followed yearly through 1955. The Hammond-Horn Study paved the way for the cancer prevention studies (CPS-I, CPS-II, and CPS-3), which began in 1959 and are continuing today. The CPS-I, CPS-II Baseline, and CPS-II Nutrition cohorts studies have produced at least 117, 163, and 34 publications, respectively.

Cancer Prevention Study I

For the CPS-I, which focuses on mortality as an endpoint, approximately 1 million male and female adults in 25 states were recruited between October 1, 1959, and February 1, 1960, and completed extensive baseline questionnaires covering demographic characteristics, height, weight, diet, occupation, alcohol and tobacco use, menstrual and reproductive history (females), and personal and family disease history, including that of cancer. Supplemental questionnaires on changes in tobacco use and cancer were mailed to participants in 1961, 1963, 1965, and 1972 to confirm vital status. Vital status of participants was determined annually from 1960 to 1965, 1971, and 1972 through personal inquiry. Vital status, as well as date and place of all deaths, was recorded in the study, and death certificates were obtained from State Health departments.

Cancer Prevention Study II

For the CPS-II, approximately 1.2 million male and female adults in 50 states, the District of Columbia, and Puerto Rico were recruited in 1982 for the Baseline Cohort and also completed extensive baseline questionnaires covering topics similar to those of CPS-I. Vital status of participants is updated in conjunction with the National Death Index through computerized linkage. Follow-up on vital statistics of the cohort is completed through 2002, and cause of death for more than 98% of all deaths for the cohort has been documented. As of 2002, more than 385,000 deaths have occurred in the CPS-II Baseline Cohort.

In 1992, a subgroup of 184,194 males and females from the CPS-II Baseline Cohort was mailed a new questionnaire covering additional detailed information on diet, updates on aforementioned lifestyle factors, and self-reported cancer diagnoses, which constitutes the CPS-II Nutrition Cohort. The Nutrition Cohort allows cancer diagnosis to be studied as an endpoint in addition to mortality. Participants of the Nutrition Cohort are aged 50 to 74 and reside in 21 states that have population-based state cancer registries, including California, Connecticut, Florida, Georgia, Illinois, Iowa, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Mexico, New Jersey, New York, North Carolina, Pennsylvania, Utah, Virginia, Washington, and Wisconsin. The Nutrition Cohort is mailed this additional questionnaire biennially starting in 1997. Self-reported cancer diagnoses are verified with medical records when consent is given, and self-reported cancer diagnoses are supplemented using computerized linkage with the state cancer registries.

Beginning in 1998, blood samples were obtained from a subgroup of 39,371 surviving males and females from the CPS-II Nutrition Cohort, comprising the CPS-II Lifelink Cohort. The Lifelink Cohort allows for future epidemiologic research on potential nutritional, hormonal, and genetic risk factors for cancer and other diseases. In June 2001, collection of blood samples was completed. Additionally, in January 2001, collection of buccal, or cheek, cells from 70,000 males and females of the CPS-II Nutrition Cohort using a mailed collection kit began as an alternative to donation of blood. All biospecimens are being stored in liquid nitrogen for future research.

Cancer Prevention Study 3

Recruitment for the Cancer Prevention Study-3 is currently under way. The goal is to enroll 500,000 males and females aged 30 to 65 years from various races/ethnicities. Nearly 25 years have elapsed since recruitment of the CPS-II Baseline Cohort in 1982 so that this new CPS-3 Cohort likely has new lifestyles and behaviors compared with the previous cohorts that need to be researched and captured.

—Binh Y. Goldstein and Zuo-Feng Zhang

See also Cancer Registries; Longitudinal Research Design; Observational Studies; Study Design

Further Readings

Thun, M. J., Calle, E. E., Rodriguez, C., & Wingo, P. A. (2000). Epidemiological research at the American Cancer Society. *Cancer Epidemiology Biomarkers & Prevention, 9*, 861–868.

Web Sites

American Cancer Society: http://www.cancer.org.

American College of Epidemiology

The American College of Epidemiology (ACE) was incorporated in 1979 with the purpose of addressing the professional concerns of epidemiologists and developing criteria for their professional recognition.

Goals of the ACE include advocating for policies and actions that enhance the science and practice of epidemiology, promoting the professional development of epidemiologists, recognizing excellence in epidemiology, issuing timely policy statements for the profession, and developing and maintaining an active membership base representing all aspects of epidemiology. The first president of the ACE was Abraham M. Lilienfeld, who was also one of the signatories of the ACE Articles of Incorporation, along with Cedric F. Garagliano and Curtis L. Meinert.

The ACE is a professional membership organization for people working in epidemiology or closely related fields: Individuals must apply for membership and have their application reviewed by the ACE Board of Directors. A recommendation for admission is based on credentials, including training, education, experience, and contributions to the profession. There are three main categories of membership: associate member, member, and fellow. Associate members are individuals who are enrolled in training, which, when completed, would qualify them for admission into the ACE. Members are individuals who have a doctoral degree in epidemiology, a doctoral degree in a related field and a master's degree in epidemiology, or a doctoral or master's degree plus related experience in epidemiology. Fellows are individuals who meet the requirements to be a member and have also demonstrated significant and sustained contributions to epidemiology through research or leadership in the field.

The ACE holds an annual scientific meeting and every 5 years holds the American Congress of Epidemiology in conjunction with the Epidemiology Section of the American Public Health Association, the Canadian Society for Epidemiology and Biostatistics, and the Society for Epidemiological Research. The First Congress of Epidemiology was held in 2001 in Toronto, Ontario, and the second in 2006 in Seattle, Washington.

Annals of Epidemiology is the official journal of the ACE. It is a peer-reviewed, international journal published 12 times per year by Elsevier. *Annals of Epidemiology* focuses on epidemiologic research and methodological development and encourages the use of epidemiology in a multidisciplinary approach to studying disease etiology. The ACE has issued official statements on several major topics within epidemiology, including minority representation in epidemiology and within the ACE; ethics; health data control, access, and confidentiality; data sharing from epidemiologic studies; and a statement of opposition to Proposition 54 in California, which would have restricted the collection and use of racial, ethnic, and national origin data within California.

—Sarah Boslaugh

See also American Public Health Association; Journals, Epidemiological; Society for Epidemiologic Research

Further Readings

Camargo, C. A., & Clark, S. (2006). Increasing diversity among the American College of Epidemiology membership. *Annals of Epidemiology, 16*(7), 529–532 (Epub March 3, 2006).

Web Sites

American College of Epidemiology: http://acepidemiology2.org.
Annals of Epidemiology: http://www.annalsofepidemiology.org.

American Indian Health Issues

North Americans indigenous to the lands of the 48 contiguous states refer to themselves variously as Indians, American Indians, Native Americans, or as members of their tribes (e.g., Menominee, Cherokee, Navajo) or cultural groups (e.g., Ojibwe), and at times simply as Americans. The term *American Indian* is used in this entry. Alaska Natives will be included as North American continental indigenous peoples with distinct political ties to the United States. Similarities and differences in tribal origins, cultures, and traditions are a backdrop to understanding health and health beliefs. European Columbian contact in 1492 precipitated abrupt changes in life ways and initiated a political history felt as historical trauma. American Indians today experience health disparities that reflect the cultural disruption and political disadvantage of the past. Culturally acceptable, culturally supportive health services, American Indian–initiated and –conducted research, and strengths inherent in traditional ways of life are identified as pathways to optimal health in the future.

Tradition and Culture

American Indians and Alaska Natives are the descendants of the indigenous peoples of the lands now

occupied by the 48 contiguous states and Alaska in the United States. Anthropological evidence suggests an Asian origin for native North American people, with migration waves across Bering Strait land bridges beginning well over 15,000 years ago. Hundreds of distinct tribes each with their own unique and overlapping traditions have been located on North American lands from time immemorial. Well over 200 distinct languages have been distinguished.

Over the centuries, American Indian cultures have reflected local geographic and climatic conditions. Migrating societies thrived through subsistence hunting, fishing, and gathering of plant materials, while sedentary groups relied largely on cultivation of crops and hunting and lived in villages both large and small. Groups appeared, shifted, and regrouped over the centuries. For example, around 1250 AD to 1540 AD, there were significant shifts during what is termed the Southwest Pueblo IV period. Through migration, the entire Pueblo world's demography shifted east from the Four Corners region, causing dramatic demographic, social, economic, and political reconfigurations.

Complex tribal social and religious structures included clans, clan families, birth families, father's clan, kiva groups, and other medicine and priest societies, and other connections such as a spiritual father. Social and religious roles and duties were carried out through dance, music, and storytelling. In some groups, music was reserved for sacred activities and rites. A common thread was a holistic integration of practical day-to-day life with nature-based spirituality.

In the past and continuing today, most tribal groups recognized physical, emotional, mental, and spiritual balance in harmony with nature as the core of health. Sacred medicine wheels, hoops, or other symbols portray tribal beliefs about health across the life cycle. Illness reflected imbalance with self, relationships, or nature. Healers were recognized as medicine people, midwives, bone pressers, stomach massagers, or members of medicine societies. Each had special knowledge of practical and spiritual methods for treating injury, illness, and emotional disturbances. Ceremonies, botanicals, including herbs and tobacco, visions, stories, and songs were used to restore balance and health.

European Contact

The American Indian population in 1492 is estimated broadly by historians and demographers to have been from 1 to 18 million. It declined precipitously following European Columbian contact. The decline was due to deaths from epidemic disease and warfare. Entire villages and tribes were annihilated as smallpox, measles, and other infectious diseases for which American Indians had no immunity and no effective treatment swept across the continent. Epidemics weakened the ability of American Indian communities to respond politically or militarily to early European settlement pressures and subsequent U.S. actions for removal to designated Indian Territory west of the Mississippi River.

The Cherokee Trail of Tears (1838–1839), Navajo Long Walk (1864), and other operations were brutal death marches where disease and starvation were commonplace. Those who survived faced harsh living conditions, sociocultural disruption, and poverty that continue to affect Indian life today. That impact has been termed *historical trauma* and is regarded as a factor in the emergence of health disparities in life span and the modern plagues of chronic illness. It is thought that loss of identity, grief, guilt, and other posttraumatic symptoms lead directly or indirectly to diabetes, stroke, heart disease, liver disease, cancer, and maladies impairing social functions, such as depression, alcoholism, substance abuse, and injury.

Political Status

The U.S. Constitution (Article II, Section 8), treaties, federal statutes, and seminal Supreme Court decisions serve as the basis for federal recognition of the 569 Indian tribes and Native Alaska corporations and for policy guiding government-to-government interaction with them, termed the *federal trust responsibility*. Political recognition of tribes as sovereign nations sets the American Indian experience uniquely apart from that of other U.S. minority groups. The U.S. government provides health services to members of recognized tribes in fulfillment of the federal trust responsibility. The Indian Health Service, Tribal Health Services, and Urban Health Services (ITU) are components of the federal system of care. The Indian Health Service (IHS) is a national system of hospitals, clinics, and satellite offices providing a widely varying range of health services for enrolled members of federally recognized tribes.

In the 20th century, an increased birth rate and slow but steady improvement in life expectancy

contributed to a rebound in the American Indian/Alaska Native population from its 1900 nadir of about 250,000 individuals. According to the 2000 census, approximately 2.5 million individuals (0.9%) of the U.S. population self-identified as only American Indian/Alaska Native and 4.4 million (1.5%) reported mixed ancestry. The number of U.S. citizens reporting some Indian ancestry or cultural identification is increasing. American Indians are U.S. citizens and citizens of the states in which they reside, but in addition can be citizens of their tribe if they become enrolled members. Tribal governing bodies set criteria for citizenship, which vary among Indian Nations. For example, Navajo Nation requires a blood quantum level of 1/4, whereas Cherokee Nation requires direct descent from individuals on the Dawes Rolls (closed in 1907) rather than a minimum blood quantum. Enrolling as a tribal member is not automatic, and one must apply for the privilege.

Most of today's American Indian/Alaska Native population is located west of the Mississippi River, with the greatest number in California (mostly urban Indians) and high concentrations in Alaska, Oklahoma, Arizona, and New Mexico. Navajo Nation has the largest reservation-based population at about a quarter of a million, while Cherokee Nation has the largest nonreservation-based population. Approximately 40% of the American Indian/Alaska Native population continues to reside on or near reservation lands. Following federal policy encouraging nonreservation relocation during the World War II era, most of the American Indian/Alaska Native population now lives in urban areas. Among Indian elders, there is a trend toward remigration from cities to reservation residence.

Modern Diseases

Today, life expectancy overall equals that of whites, but the change is distributed unequally across the American Indian population. Males and females residing near reservations in 2001 had a life expectancy of 5.9 and 4.3 years less than whites with slightly above-average income and education living in areas other than the upper Midwest, Appalachia, and the Mississippi Valley and some Asian Americans.

Quality data are an essential foundation for describing the burden of disease in a population, planning effective health services, and setting research agendas. Obtaining data about American Indian health is challenging due to inconsistencies and difficulties in defining, identifying, and accessing the American Indian/Alaska Native population. Nevertheless, it is clear that the American Indian population is younger than the U.S. overall population, with 8.6% below 5 years as compared with 6.8% for the U.S. overall in 2000. There are fewer elders above 54 years (11.9% compared with 21%). Life expectancy at birth increased from 63.5 in 1972 to 73.2 in 1992. Yet American Indian/Alaska Native teens and young adults residing on or near reservations have among the highest mortality rates of groups in the United States.

Despite lower rates of low birthweight, American Indian/Alaska Native infant mortality rates are higher than U.S. overall (9.7 vs. 6.8 per 1,000 in 2001 data). Elevated postneonatal death rates account for the discrepancy. Sudden infant death syndrome (SIDS; the sudden death of an infant younger than 1 year of age that remains unexplained after a thorough case investigation that includes autopsy, death scene investigation, and review of clinical history) is the leading cause of postneonatal death. The IHS Aberdeen Area (North Dakota, South Dakota, Nebraska, Iowa) reported rates ranging from 3.46 to 3.66/1,000 live births from 1992 to 1998 (compared with 0.7/1,000 live births for all U.S. races in 1999). The disparity persisted in 2003, when American Indian/Alaska Native SIDS deaths occurred at 2.3 times the rate for infants of non-Hispanic white mothers, even as SIDS rates for all U.S. groups were falling. Reasons for the elevated incidence of SIDS in American Indian/Alaska Native infants are uncertain. Risk factors are the same as those in other populations (smoking, delayed prenatal care, nonsupine sleep position, low socioeconomic status, layered clothing), although smoking occurs at relatively high rates in the American Indian/Alaska Native population. Serotonergic brainstem abnormalities found in non-Indian and American Indian/Alaska Native victims of SIDS may play a role. Until a more complete understanding of SIDS pathogenesis is available to guide effective prevention, treatment is directed toward reduction of risk factors and compassionate support of families whose babies died of SIDS.

Alcoholism appeared in American Indian populations during early days of European contact when ammunition, alcohol, and tobacco were primary trade goods. Now, a constellation of alcoholism, injuries (homicide, suicide, motor vehicle accidents,

traumatic brain injury), substance abuse, domestic and community violence, and depression disproportionately affects American Indian/Alaska Native youth and young adults. Accidents and homicide are among the top three causes of death for American Indian/Alaska Natives from 1 to 24 years of age. Suicide among American Indian/Alaska Native youth occurs over 2.5 times more frequently than among the U.S. all races group. The alcoholism death rate for Indians from 15 to 24 years of age is 5.5 per 100,000 compared with 0.3 for all U.S. races. Many Indian youth rate their health status as low. The social toll is enormous, affecting all aspects of life for affected individuals, their families, and tribal communities. Poverty, lack of educational opportunity, isolation, and cultural disruption are all thought to be etiologic factors. Among adults, perceived discrimination is associated with higher levels of depression, and practice of traditional life ways appears protective. A sense of social competence, positive school achievement, and avoidance of substance use are related to well-being among youth. Many American Indian communities view elimination of substance abuse as a key to reducing the spectrum of social pathologies and draw on their cultural strengths to design interventions ranging from programs for individual treatment to community-wide prevention.

Epidemic type 2 diabetes emerged in the American Indian/Alaska Native population in the mid-20th century. It is recognized as a ''white man's disease'' resulting from disruption of traditional life ways, including diet and activity patterns. Prevalence is markedly higher, onset occurs at younger ages, and renal, cardiovascular, and retinal complications develop at higher rates than in the general population. Depression comorbidity is high as well. Genetic and behavioral factors have been implicated in the increasing incidence and prevalence of type 2 diabetes in the American Indian/Alaska Native population. The highest prevalence of diabetes in the world occurs among the Pima Indians of Arizona. The diabetes burden of disease is expected to grow as today's youthful American Indian/Alaska Native population with diabetes ages with longer exposure to risk factors for complications. Increased physical activity, improved diet, and weight management are key aspects of both primary and secondary prevention because they may prevent or delay onset of diabetes and mitigate its debilitating complications.

Access to effective care is critical to the monitoring required for early identification and treatment of diabetes complications. A challenge in diabetes prevention and control is altering beliefs that the disease is inevitable. Awakening the Spirit, an American Indian/Alaska Native program sponsored by the American Diabetes Association, emphasizes that diabetes did not occur in the past and that people with diabetes can manage the disease, live full lives, and be well enough to watch their grandchildren grow up. Within ITU programs, community diabetes interventions reflect local priorities. In an expression of hope and responsibility for the well-being of future generations, many programs focus on diabetes prevention efforts for children.

There is also a disparity between health problems of elderly American Indians and Alaska Natives compared with the United States at large. Chronic diseases and resulting functional declines appear earlier than in the general population, are of greater severity, and usually result in a shorter life span. For these reasons, American Indian elders are often eligible for IHS, tribal health care services, or urban Indian health services at age 50. The service base is spare, however. The elderly on reservations have almost no formal services because long-term care is not a principal function of IHS or tribal health care. There are only 12 reservation-based nursing homes in the country. Due to moratoriums on building new nursing home beds in 32 states, many needed beds will never be built. Clashing health beliefs and traditions and laws are barriers that keep American Indian elders from obtaining off-reservation long-term care, which perpetuates a cyclical wheel of disparity.

Pathways to Optimal Health

Beginning in the early 1800s, military physicians provided public and personal health services aimed at control of infectious disease among Indian people. After 1849, the Bureau of Indian Affairs in the Department of the Interior administered medical care. Through the Snyder Act of 1921, Congress authorized funds for health services to federally recognized tribes. Responsibility was moved to the U.S. Public Health Service in 1954, with the IHS serving 12 areas. The IHS, headquartered in Rockville, Maryland, comprises a national health system operating 33 hospitals, 59 centers, 34 urban health clinics, and 50 stations. Under the Indian Self-Determination Act

(PL 93–638), tribes receive direct IHS services or operate health services using funds allocated through IHS.

The mission of the IHS, in partnership with the American Indian/Alaska Native people, is to raise physical, mental, social, and spiritual health to the highest level. Improvements in Indian health across the 20th century have been attributed, in part, to efforts of the IHS. However, the IHS faces continuing challenges in providing services.

Many American Indian/Alaska Natives eligible for services lack access to IHS or tribal facilities. Of the estimated 3.3 million members of federally recognized tribes and their descendents who are eligible for care, only about 1.8 million receive IHS services. Intertribal urban health clinics decrease geographic barriers by providing services in some major metropolitan areas with high Indian populations.

The IHS employs more than 15,000 health care professionals. Yet the system is strapped by ongoing shortages of personnel, estimated at 12% in 2003. Remote locations are an employment disincentive, as is unfamiliarity of many health care providers with Indian culture. American Indian/Alaska Natives are underrepresented among health professionals. Health professions awareness programs beginning in elementary schools and scholarship/mentoring programs supporting American Indian students in higher education are designed to ease the shortage.

Traditional Indian medicine and Western care are used to varying degrees by many American Indian/Alaska Native people. American Indian/Alaska Native health professionals are leaders in encouraging use of tribal healers. The American Association of Indian Physicians includes as part of its mission the honoring of traditional healing practices. The National Alaska Native American Indian Nurses Association logo incorporates bear claws to signify strength and medicine for healing. IHS policy directs local service units to collaborate with tribes in provision of traditional healing to their patients.

American Indians are sometimes characterized as among the most overstudied and underserved of U.S. populations. Yet the need for research is recognized even though there are daunting challenges in implementing research and converting findings to action. While the American Indian/Alaska Native population shares some commonalities, tribes are also culturally distinct. Research procedures acceptable with one group may be anathema to another, and findings may not generalize. Research tools such as survey questionnaires developed in the majority population may appear peculiar to American Indian/Alaska Native respondents. For example, depression-screening tools may include language and symptoms inconsistent with interpretation among American Indians.

Indian-initiated and -conducted research is one pathway to the elimination of American Indian health disparities. The Native Research Network links American Indian/Alaska Native investigators to promote excellence and integrity in research. The National Institutes of Health, Minority Access to Research Careers, and Bridges to the Future programs (including the University of Minnesota M.S. to Ph.D. Nursing Science Bridge Program) reflect a federal commitment to developing a cadre of researchers from underrepresented minority groups including American Indians. The Resource Centers for Minority Aging Research (including the Native Elder Research Center at the University of Colorado) emphasize postdoctoral training. The Native American Research Centers for Health program is a mechanism for linking tribes with resources, including technical expertise from collaborating research universities, to carry out their research agendas. American Indian health scientists have firsthand appreciation of the complexity of health and illness experiences among American Indians and Alaska Natives, and they know the importance of combining tribal self-determination, traditional spiritual values, and the highest-quality science for the resolution of disparities.

—Susan J. Henly and Margaret P. Moss

See also Aging, Epidemiology of; Alcohol Use; Complementary and Alternative Medicine; Diabetes; Health Disparities; Maternal and Child Health Epidemiology

Further Readings

Burhansstipanov, L., & Satter, D. E. (2000). Office of Management and Budget racial categories and implications for American Indians and Alaska Natives. *American Journal of Public Health*, *90*, 1720–1723.

Burrows, N. R., Engelgau, M. R., Geiss, L. S., & Acton, K. J. (2000). Prevalence of diabetes among Native Americans and Alaska Natives, 1990–1997. *Diabetes Care*, *23*, 1786–1790.

Henly, S. J., Struthers, R., Dahlen, B. K., Ide, B., Patchell, B., & Holtzclaw, B. J. (2006). Research careers for American Indian/Alaska Native nurses: Pathway to

elimination of health disparities. *American Journal of Public Health*, *96*, 606–611.
Indian Self-Determination and Education Assistance Act, Pub. L. 93–638, 25 U.S.C. 450 et seq. (1975).
Kunitz, S. J. (1996). The history and politics of U.S. health care policy for American Indians and Alaska Natives. *American Journal of Public Health*, *86*, 1464–1473.
Moss, M. P., Roubideaux, Y. D., Jacobsen, C., Buchwald, D., & Manson, S. (2004). Functional disability and associated factors among older Zuni Indians. *Journal of Cross-Cultural Gerontology*, *19*, 1–12.
Murray, C. J. L., Kulkarni, S. C., Michaud, C., Tomijima, N., Bulzacchelli, M. T., Iandiorio, T. J., et al. (2006). Eight Americas: Investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Medicine*, *9*, 1513–1524. Retrieved October 10, 2006, from http://medicine.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pmed.0030260, and available at www.plosmedicine.org.
Rhoades, E. R. (Ed.). (2000). *American Indian health: Innovations in health care, promotion, and policy.* Baltimore: Johns Hopkins University Press.
Struthers, R., Lauderdale, J., Nichols, L. A., Tom-Orme, L., & Strickland, C. J. (2005). Respecting tribal traditions in research and publications: Voices of five Native American nurse scholars. *Journal of Transcultural Nursing*, *16*, 193–201.
Thornton, R. (1987). *American Indian holocaust and survival.* Norman: University of Oklahoma Press.

American Public Health Association

The American Public Health Association (APHA), founded in 1872, is the oldest public health association in the world. It is also the largest, with more than 50,000 members in 2006, representing many different occupations related to public health. The APHA engages in a variety of activities to promote public health, including disseminating information through press releases, books and journals, holding an annual meeting, providing opportunities for professional development and continuing education, and bestowing a number of awards on individuals who have aided in public health efforts.

The APHA was founded by Stephen Smith, then Commissioner of the Metropolitan Health Board in New York City. The first annual meeting was held in 1873 and, as is still the rule today, combined presentation of scholarly papers with inclusion of more popular speakers. This format reflects APHA's interests in garnering support for public health among politicians and the general public as well as providing a forum for the exchange of ideas among people working in public health. APHA has held annual meetings every year from 1872 to the present, except for 1945; most have been held in the United States, but a few have taken place in Mexico, Canada, or Cuba.

As membership grew, members with common interests began forming sections to discuss topics of mutual interest. Among the first founded were the sections devoted to laboratory investigation (1899), health administration (1908), statistics (1911), and sanitary engineering (1911). Currently, there are 24 sections representing interests from alcohol, tobacco, and other drugs through Vision Care. APHA currently has 17 caucuses, which are formed around either members with common backgrounds, such as ethnicity or occupation, or common concerns and outlooks, such as Socialism or Peace.

The official journal of APHA is *The American Journal of Public Health*, which began publication in 1911 and publishes primarily scholarly research articles. APHA also publishes the monthly newspaper *The Nation's Health*, which carries articles written in a more popular style, often summarizing current research and reporting on public policy issues. Books relating to a number of public health concerns are also published through APHA.

The APHA fosters awareness of public health through presentation of a number of awards honoring individuals and organizations. The William Thompson Sedgwick medal has been awarded annually since 1929, to honor distinguished service by those working in public health. The Presidential Citation of the American Public Health Association is awarded on an irregular basis for service to public health by someone not working specifically in public health: Recipients have often been politicians or journalists. Other awards bestowed by the APHA include the David P. Rall Award for Advocacy in Public Health, the Milton and Ruth Roemer Prize for Creative Public Health Work (awarded to a health officer working at the city, county, or other local governmental level), and the APHA Distinguished Public Health Legislator of the Year (awarded to a local, state, or federal lawmaker).

—Sarah Boslaugh

See also Governmental Role in Public Health; Health Communication; Journals, Public Health; Public Health, History of

Further Readings

Bernstein, N. R. (1972). *The first one hundred years: Essays on the history of the American public health association*. Washington, DC: American Public Health Association.

Web Sites

American Public Health Association: http://www.apha.org.

Analysis of Covariance

Analysis of covariance (ANCOVA) is a combination of analysis of variance (ANOVA) and regression analysis because the model contains both quantitative and qualitative independent variables. The idea is to enhance the ANOVA model by adding one or more quantitative independent variables that are related to the dependent variable. These variables are called concomitant variables or covariates. Increasing the precision of the model results in reducing the error terms. Without a covariate, the error mean square may be so high that a simple ANOVA may not detect differences between treatments. Covariates can also be used to remove the effect of an extraneous variable from the dependent variable. An extraneous variable influences the outcome of an experiment but is not of particular interest to the researcher.

Consider the study of a new weight-loss medication. In a double-blind study, weight is measured on subjects who have been randomly assigned to one of two treatment groups. One group receives the new medication, and the other group receives a placebo. The researcher wants to know whether the new medication produces significant weight-loss results. Since the effect of the medication may be related to the individual's initial weight, initial weight is used as a covariate in the analysis to reduce within-treatment variability. The ability to detect differences between treatments is now strengthened.

ANCOVA is also used to explore the nature of treatment effects rather than for increasing the precision of the model. In a study of the effect of two different cognitive therapy treatments, children with behavioral problems are assessed by a mental health professional using a questionnaire. Each child is given a total problem score. Parents are asked to fill out a questionnaire to establish a socioeconomic status (SES) score for each child as well. The SES score is used as a covariate in the analysis. In this case, the relationship between total problems and SES score for each treatment is of primary concern rather than the effect of the treatments on total problem score.

ANCOVA is often used as a means of correcting for bias when treatment groups are noticeably different from each other. In the double-blind study of a new medication to reduce blood pressure, subjects are randomly assigned to one of two treatments: those treated with a placebo or those treated with the medication. Suppose it is found that by chance the initial blood pressure for subjects in one group is found to be substantially higher than that of the other group. Adding initial blood pressure as a covariate in the model helps remove that bias. Using ANCOVA for this purpose must be done with caution, however. If the covariate is related to the treatment variable, any conclusions are questionable at minimum. For instance, in a study of attitudes toward two different blood glucose monitors for diabetics, it is found that older patients tend to like one monitor while younger patients tend to like the other. With little regard, age is a covariate in an ANCOVA in an attempt to remove the bias. As it turns out, however, age is related to monitor preference. Therefore, using age as a covariate could actually lead to the wrong conclusion.

Selecting the right covariate requires careful thought and consideration. If the covariate is not related to the dependent variable, nothing is gained by adding it to the model. An ANOVA is more appropriate and less complicated in this situation. Covariates are usually observed before the study begins and should be independent of any effect of the treatments to obtain meaningful results. If the covariate is measured during the study, it is important to ensure that it is not influenced by the treatments. Examples of covariates are prestudy attitudes toward treatments, questionnaire scores, and prestudy health condition measurements, such as weight and blood pressure.

The ANCOVA model starts with an analysis of variance model. One or more terms are added to the model to reflect the relationship between the dependent variable, Y, and the independent variable, X.

The model for one covariate is as follows:

$$Y_{ij} = \mu + \tau_i + \gamma(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij},$$

where

μ is an overall mean;
τ_i are fixed treatment effects subject to the restriction $\Sigma\tau_i = 0$;
γ is a regression coefficient for the relationship between Y and X;
X_{ij} are covariate values;
ε_{ij} are independent and normally distributed with mean 0 and variance σ^2;
$i = 1, \ldots, r$ treatments; and
$j = 1, \ldots, n_i$, n_i is the number of subjects in treatment i.

The basic properties of the ANCOVA and the ANOVA models are the same. However, since there is more to the ANCOVA model, it has additional properties.

In ANOVA, all observations of the ith treatment have the same mean response (μ_i). In ANCOVA, however, the mean response depends on the treatment as well as the value of the covariate, X_{ij}. Therefore, the mean response for the ith treatment is given by

$$\mu_{ij} = \mu_{.} + \tau_i + \gamma(X_{ij} - \bar{X}_{..}),$$

which is the mean response for treatment i at any value of X. When $X_{ij} - \bar{X}_{..} = 0$, the value of Y is $\mu + \tau_i$ and γ is the slope of each regression line.

Since the slope of the regression line is the same for all treatments, the difference between the mean responses is the same at any value of X. Figure 1 shows three regression lines for three treatments using hypothetical data. Treatment 1 has a higher mean response than Treatments 2 and 3, and Treatment 3 has a higher mean response than Treatment 2. The difference between the mean responses for Treatments 1 and 3 is the same regardless of the value of X because the slopes are parallel. This property of the ANCOVA model is referred to as *constancy of slopes.* It allows for comparison of treatment effects at any convenient level of X, such as $X_{ij} = \bar{X}_{..}$. Without constancy of slopes, ANCOVA is inappropriate. A test for parallel slopes should be conducted before proceeding with ANCOVA.

The question of interest in ANCOVA is similar to that in ANOVA: Are there significant treatment effects and, if so, what are they? ANCOVA uses regression procedures to adjust the dependent variable for the

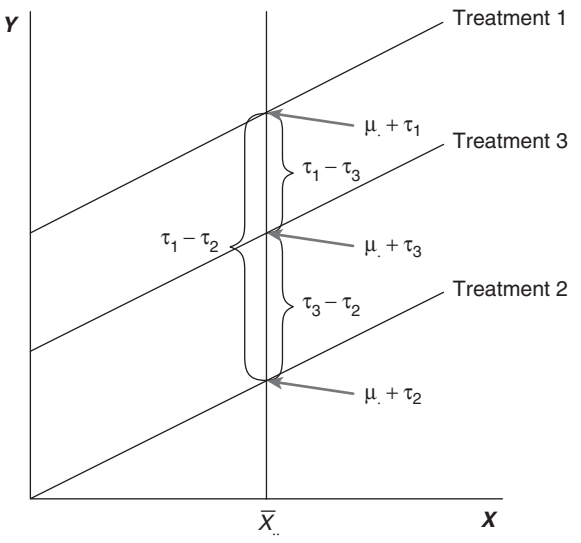


Figure 1 Regression Lines for Three Treatments in ANCOVA

effect of the covariate. In ANCOVA, therefore, the question is modified: Are there significant differences between treatments if the covariate is held constant? If the treatment regression lines are parallel, the hypothesis to test for differences between treatments, as in ANOVA, is that the population means for each treatment are equal:

$$H_0 : \tau_1 = \tau_2 = \cdots \tau_r = 0.$$

The alternative hypothesis is that at least one τ_i is not equal to zero. In other words, are the regression lines at the same level or is at least one of them higher than the others? If there is a significant difference between treatments, pairwise comparisons of treatment effects $\tau_i - \tau_i'$ can be made. This amounts to comparing the vertical distance between two regression lines. If necessary, more general contrasts of the τ_i can be made as well.

When there is no relationship between Y and X, the error mean square is the same as for ANOVA, and one degree of freedom for the error is lost. Therefore, testing whether the regression coefficient γ is zero is generally not done.

The ANCOVA model can also be expressed in terms of a regression model. In this model, $r - 1$ dummy variables are used to represent the treatments:

$$Y_{ij} = \mu + \tau_i I_{ij1} + \cdots + \tau_{r-1} I_{ij,r-1} + \gamma(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij},$$

where I_{ij1} is the value of dummy variable I_1 for the jth observation from treatment i. The treatment effects $\tau_1, \ldots, \tau_{r-1}$ are regression coefficients for the dummy variables. Interaction terms $I_1(X_{ij} - \bar{X}_{..})$, $\ldots, I_{r-1}(X_{ij} - \bar{X}_{..})$ can be added to this model to allow for nonparallel slopes:

$$Y_{ij} = \mu_{.} + \tau_i I_{ij1} + \cdots + \tau_{r-1} I_{ij,r-1} + \gamma(X_{ij} - \bar{X}_{..}) + \beta_1 I_{ij1}(X_{ij} - \bar{X}_{..}) + \cdots + \beta_{r-1} I_{ij,r-1}(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij}.$$

The regression form of the ANCOVA model is useful for testing for parallel slopes:

$$H_0 : \beta_1 = \cdots = \beta_{r-1} = 0.$$
$$H_a : \text{ at least one } \beta \text{ is not equal to zero}$$

An F test is used to compare the error mean squares of the models with and without interaction terms.

There may be situations where more than one covariate is appropriate. Extending the ANCOVA model is fairly straightforward. In the case of two covariates, X_1 and X_2, the model becomes

$$Y_{ij} = \mu_{.} + \tau_i + \gamma_1(X_{ij1} - \bar{X}_{..1}) + \gamma_2(X_{ij2} - \bar{X}_{..2}) + \varepsilon_{ij}.$$

The relationship between Y and X discussed in this article is linear, but linearity is not required. A linear model is preferable, since it is easier to interpret. In some cases, nonlinear data may be transformed so that a linear model can be used. When it is not reasonable to use a linear model, a nonlinear model may provide more meaningful results. The relationship, for instance, could be a quadratic one:

$$Y_{ij} = \mu_{.} + \tau_i + \gamma_1(X_{ij} - \bar{X}_{..}) + \gamma_2(X_{ij} - \bar{X}_{..})^2 + \varepsilon_{ij}.$$

When using a nonlinear model, parallel curves fit to each treatment are of interest.

—Mary Earick Godby

See also Analysis of Variance; Dummy Variable; Regression

Further Readings

Montgomery, D. C. (1984). *Design and analysis of experiments.* New York: Wiley.

Netter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models: Regression, analysis of variance, and experimental designs.* Homewood, IL: Richard D. Irwin.

Analysis of Variance

Most analyses of relationships between variables involve the use of independent and dependent variables. Analysis of variance (ANOVA) is a collection of methods where the independent variable(s) are categorical (nominal) and the dependent variable is quantitative (metric, interval/ratio) with a numerical scale. The analysis compares the means of the dependent variable for the groups defined by the independent variable(s). A more appropriate name might be analysis of means, but variances are used to determine whether means are different.

The simplest case has one independent variable with two categories and one dependent variable. For example, ANOVA can be used to analyze the relationship between the independent variable gender (categorical) and the dependent variable blood pressure (quantitative). The analysis compares the mean blood pressures for females and males. This comparison can also be made using a t test for the comparison of two means, and such a t test is a special case of ANOVA. Furthermore, the comparison can be made using regression analysis with a dummy variable for gender, and such a regression analysis is also a special case of ANOVA.

ANOVA and regression analysis are special cases of the general linear model, and there are mainly historical reasons why the two methods are seen as distinct. ANOVA grew out of analysis of experimental data in agriculture, where yields were compared for different treatments, such as type of fertilizer. Most of this took place in England under the leadership of the great statistician Ronald Fisher (1850–1921). Much of regression analysis has its foundation in economics with its many quantitative variables. Regression analyses used by economists are often called econometrics. The answer to which method of performing ANOVA is the proper approach can itself be analyzed using regression analysis with the dummy variables, but the construction of such variables can be cumbersome. Most statistical software programs still distinguish between ANOVA and regression, and regression analysis typically provides less detailed output than does ANOVA.

With one independent variable, we perform one-way ANOVA, with two independent variables we perform two-way ANOVA, and so on. Introducing location as a second independent variable with values

urban and rural, we can use a two-way ANOVA to study whether there are differences in blood pressure for females and males as well as between an urban and a rural location, using both gender and location as the two variables in one analysis. We could do a one-way analysis for gender and then separately do a one-way analysis for location, but it is more efficient to use both gender and location at the same time in one analysis. With more than one dependent variable, we perform multivariate ANOVA.

Main Results of an Analysis of Variance

ANOVA provides a measure of how strong the relationship is between the dependent and the independent variable(s) on a scale from 0 to 1. The strength of the relationship is measured by the quantity R^2, which takes on the role of a squared correlation coefficient. Such a number is also known as the size of the effect of the independent variable, measured on a scale from 0 to 1 or a percentage from 0 to 100. With several independent variables, we can get such a coefficient for each variable, telling us how strong the relationship is between that particular independent variable and the dependent variable. We can even get an overall coefficient, which sometimes is simply the sum of the coefficients for the individual variables.

Also, the analysis can test whether the relationship is statistically significant—that is, whether the group means are different from each other. A typical null hypothesis is that the group population means are equal. We will reject the null hypothesis if our analysis returns a small p value, meaning it is unlikely that if the population means were equal, results would be as extreme as they were. The p value is found using the theoretical statistical F variable, named for Fisher. If the group means are different, we can also ask whether they are all different from each other or whether only some of them are different from each other. An alternative to a null hypothesis stating that all means are equal is that at least some of them are not equal but not necessarily that they are all different.

One-Way Analysis of Variance

Suppose we think a variety of socioeconomic variables have an impact on obesity. Using census tracts of a city as a substitute for these variables, we ask whether average weight varies for residents living in different city census tracts. Weight as the dependent variable is quantitative, and census tract as the independent variable is categorical. The data set is drawn from random samples of women, one sample from each census tract.

In this analysis, any person's weight may be explained by either of two factors. One is location (census tract of residence) and the other is the combined, net effect of all variables other than location. Such a variable is known as the residual variable. Suppose the residual variable has no effect—weight is determined by location only. In that case, all women in a given census tract will have the same weight. The best estimate of that common value is the mean weight in that census tract. But this is not a realistic premise; we would not expect all the women in any sample to have the same weight. The difference between an individual woman's weight and the mean weight in her census tract is therefore taken to be the effect of the residual variable. Thus, for each woman we can measure the effect of the residual variable. We then need some way to summarize these residual values. It is tempting to take their mean, but that is not feasible as some of the residuals will have negative values and some will have positive values, and by definition their mean will equal zero.

One way around this problem is to square all the residuals to make them positive. The overall effect of the residual variable can then be defined as the sum of all these squared terms. This gives the residual sum of squares (RSS) as

$$\text{Residual variable sum of squares} = \text{Sum (Observation} - \text{Group mean)}^2.$$

The larger the effect of the residual variable is, the larger this sum will be. This sum looks like the numerator in a variance.

Similarly, if no variables influenced the weight of individual women, then all women in the study would have the same weight. The best estimate of this common weight is the overall mean weight of all the women. Thus, the difference between the weight of a woman and the overall mean becomes the effect of all variables. One way to summarize these differences is to take their squares and add all the squares. This gives the total sum of squares (TSS) as

$$\text{Total sum of squares} = \text{Sum}\,(\text{Observation} - \text{Overall mean})^2,$$

as a measure of the effects of all variables.

Now we have the effect of all the variables and the effect of the residual variable, but not the effect of the location variable. If location did not have an effect, then the means in the census tracts would all be equal and equal to the overall mean. Thus, the difference between a census tract mean and the overall mean tells the effect of the location variable. We can find this difference for every woman, square all the differences, and add the squares. This gives the effect of the independent variable (ISS) as

$$\text{Independent variable sum of squares} = \text{Sum}\,(\text{Group mean} - \text{Overall mean})^2.$$

The more the group means are different, the larger this sum will be. This also looks like the numerator in a variance, and we can now begin to see how we use variances to tell if means are different.

The way these sums have been defined gives the identity:

$$\text{Total sum of squares} = \text{Independent sum of squares} + \text{Residual sum of squares}.$$

This equality always holds and is a mathematical consequence of the definitions above.

These sums are often displayed in a table. Sometimes ISS is called the between-group sum of squares, and the RSS is called the within-group sum of squares. The sums of squares are interpreted as the magnitude of effect, and the proportion R^2 tells the effect of the independent variable on a scale from 0 to 1 or as a percentage from 0 to 100.

The effects are based on the variation in the data. If all the data points are equal, then there are no effects and there is no variation. The TSS gives the total variation in the data, and this variation is broken down into two parts: one for the independent variable and one for the residual variable. Thus, we can say the independent variable explains the proportion R^2 of the total variation in the data. Multiplying R^2 by 100 gives the percentage of the variation in the dependent variable that is explained or accounted for by the independent variable.

Finally, what about the populations from which the data came? Most analyses of variance include test(s) of significance. The null hypothesis states that all the population means are equal versus the alternative hypothesis that at least one population mean is different from the others. It may not be surprising that the null hypothesis is not rejected when R^2 is small and most of the effect is accounted for by the residual variable—that is, when the RSS is large in comparison to the ISS. But the RSS, to a large extent, depends on how many observations there are, and the ISS, to a large extent, depends on how many groups there are.

Table 1 Magnitude of Effects

Variable	*Magnitude of Effect*	*Proportion*
Independent variable	ISS	$R^2 = \frac{\text{ISS}}{\text{TSS}}$
Residual variable	RSS	$1 - R^2 = \frac{\text{ISS}}{\text{TSS}}$
Total	TSS	1.00

Notes: ISS, independent sum of squares; RSS, residual sum of squares; TSS, total sum of squares.

We compensate for this by normalizing both sums of squares before comparing them. This is done by dividing RSS by the quantity $n - k$, where n is the total number of observations and k is the number of groups (this is analogous to what we do in the computation of the variance for a set of observations when we divide the sum of squares by $n - 1$). This gives the residual mean square (RMS) as

$$\text{RMS} = \frac{\text{RSS}}{n - k},$$

which then is the variance of the residual terms. Similarly, ISS is divided by $k - 1$ to give the independent mean square (IMS),

$$\text{IMS} = \frac{\text{ISS}}{k - 1}.$$

The two quantities $n - k$ and $k - 1$ are the so-called degrees of freedom for the corresponding sums.

When the null hypothesis is true, the two mean squares are estimates of the same variance and therefore approximately equal. When the null hypothesis is not true, then the IMS is a good deal larger than the RSS. To compare the two mean squares, we take their ratio

$$F = \frac{\text{IMS}}{\text{RMS}}.$$

Table 2 Analysis of Variance

Sources of Variation	*Sums of Squares*	*Proportions*	*Degrees of Freedom*	*Mean Squares*	*F Ratio*
Independent variable	ISS	$R^2 = \frac{\text{ISS}}{\text{TSS}}$	$k - 1$	$\text{IMS} = \frac{\text{ISS}}{k-1}$	$F = \frac{\text{IMS}}{\text{RMS}}$
Residual variable	RSS	$1 - R^2 = \frac{\text{RSS}}{\text{TSS}}$	$n - k$	$\text{RMS} = \frac{\text{RSS}}{n-k}$	
Total	TSS	1.00			

Notes: IMS, independent mean square; ISS, independent sum of squares; RMS, residual mean square; RSS, residual sum of squares; TSS, total sum of squares.

This is the F ratio with $k - 1$ and $n - k$ degrees of freedom. If the observed value is close to 1, then there is very little difference between the means. When F is much larger than 1, there is typically a significant difference between the means. Most statistical software gives p values for the observed values of F. There are also extensive published tables of the F distribution. For example, with 1 (2 groups) and 50 (52 observations) degrees of freedom, the .05 value of $F_{1,50} = 6.30$. Thus, we need a larger value of F than 6.30 to reject the null hypothesis of equal means. This test is valid only when the data in each group are approximately normally distributed.

These computations are often summarized in an ANOVA table, as seen in Table 2. There are times when the column with the proportions is not included in the table.

Finally, when a large F value results in a small p value, such that the null hypothesis of equal group means is rejected, we may also want to determine which of the means are different and which are not different from each other (assuming there are more than 2 groups). The answer is typically found using methods known as multiple comparisons.

Two-Way Analysis of Variance

When there are two independent, categorical variables, it would be possible to do two separate one-way analyses. With an analysis using the first variable, the effect of the second variable would be included in the residual variable, and the same for the second variable. However, if we can take the variance explained by both variables out of the residual variable simultaneously, the effect of the residual variable will be less. Since the effect of the residual variable is used in the denominator for the F tests, a smaller residual effect will result in a larger F and therefore (holding all other facts constant) a better chance to find a significant result for each independent variable.

Two new issues arise. First, the two categorical independent variables themselves define a contingency table. Suppose there are two treatments A and B, and the dependent variable is level of blood sugar. Some subjects would get both A and B, some would get A but not B, some would get B but not A, and some subjects would get neither A nor B. This gives a contingency table with two rows and two columns. How many subjects should we have in each of the four cells? With the same number of observations in each cell, it will be possible to get unique effects for both treatments, using ANOVA. In experiments, we have control over how many subjects there are in each cell. Thus, ANOVA has a close relationship to how experiments are designed.

But if we had a sample survey and categorized people by gender and urban/rural residence, there is no reason to expect there would be the same number of observations in each cell. Thus, if gender and residence are themselves related, we will not be able to get unique measures of the effects of gender and residence. If we did the analysis using multiple regression with two dummy variables, those variables would probably be correlated. With collinearity in the independent variables, we do not get unique sums of squares for each variable.

Second, there may be a so-called interaction effect present. There may be an additional effect of A and B, over and beyond their separate effects. Suppose we are testing two treatments intended to lower blood sugar, A and B, and the subjects that get both treatments show a reduction in blood sugar larger than the combined effects of A and B. This shows that there is an interaction between the two treatments. Fortunately, the strength of this interaction is easily quantified in an ANOVA. The typical output from

a two-way ANOVA will have a row for the first variable, a row for the second variable, a row for the interaction variable, a row for the residual variable, and a total row. There will be sums of squares for each variable that can be used to establish R^2s for each variable, degrees of freedom for each variable, mean squares, and F values for the two independent variables and the interaction variable. If the ANOVA table does not show the interaction variable, it may be that the interaction effect was found to be small and statistically insignificant. It is then common to combine the original RSS and the interaction sum of squares into a new RSS and use this for the basis of the F tests for the two categorical variables.

More Than Two Independent Variables

ANOVA generalizes directly to cases with more than two independent categorical variables. The main difference is that more interaction effects are possible. With three independent variables A, B, and C, there are main effects for each of the variables and two-way interaction effects for A and B, A and C, and B and C. We also get a three-way interaction effect ABC. Higher-order-interaction effects are often difficult to interpret. If they are not statistically significant, their sums of squares and degrees of freedom are often combined with the original RSS and degrees of freedom, and the new RSS and degrees of freedom are used for the F tests for the remaining variables.

Random Versus Fixed Variables

The distinction between random and fixed variables refers to whether we use all the values of the variable in question. Gender is an example of a fixed variable, because we use all the values (female and male) of the variable in the analysis. As an example of a random variable, consider a situation in which we want to take random samples of people in the almost 70 counties in Pennsylvania and compare the counties on some characteristic such as weight, in anticipation of a statewide antiobesity campaign if there are county differences. If there are no differences in weight across counties, the campaign will not take place. Taking a random sample within each county is a large undertaking; it is easier first to take a random sample of counties and then to take a random sample of people within the chosen counties. Since we are not using all the values of the county variable, county becomes a random as opposed to a fixed variable, because the values included represent a random selection of all the values to which we wish to generalize. For more complicated studies, it makes a difference in the analysis whether we have fixed or random variables.

—Gudmund R. Iversen

See also Analysis of Covariance; Degrees of Freedom; F Test; Multiple Comparison Procedures; Multivariate Analysis of Variance; Regression

Further Readings

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07–054). Beverly Hills, CA: Sage.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments.* New York: Springer-Verlag.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (2nd ed.) (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07–001). Newbury Park, CA: Sage.

Antagonism

See Effect Modification and Interaction

Anxiety Disorders

Anxiety disorders are the most common mental disorders among adults with a peak in incidence during young adulthood. Anxiety disorders are associated with high rates of psychiatric comorbidity, physical illness, social and occupational disability, suicidality, and high rates of health service utilization in the United States and worldwide. There are mounting data to suggest that anxiety disorders begin early in life, are common among youth and adults, frequently persist throughout development, and increase the risk of subsequent psychosocial and psychiatric morbidity, including suicidal behavior, from early childhood into adulthood. Moreover, despite rapid increases in availability of efficacious psychotherapeutic and

psychopharmacologic treatments for anxiety disorders in the past three decades, evidence to date suggests that few with anxiety disorders seek and receive treatment.

The prevalence of anxiety disorders has been documented in a number of cross-sectional psychiatric epidemiologic studies that suggest that the lifetime prevalence of anxiety disorders ranges from 24.9% to 44% among adults in the community.

Risk Factors

In recent years, there has been increasing interest into factors that influence an individual's risk of developing an anxiety disorder. In contrast to heavy research investment into a number of other mental disorders, including conduct, mood, and psychotic disorders, there has been comparatively little research into the risk factors for anxiety disorders. Available evidence to date suggests that demographic factors, a history of childhood physical or sexual abuse, family history of anxiety disorders, disruption in parenting, and certain perinatal factors may increase the risk of anxiety disorders. In addition, individual factors associated with risk of anxiety disorders include evidence of mental disorders in childhood and certain genetic and personality factors.

Numerous studies have documented an association between specific demographic factors and increased risk of anxiety disorders. These studies variously show that factors such as female gender, lower socioeconomic status, minority racial status, marital status, and age are associated with increased risk of anxiety disorders in adulthood.

Childhood physical abuse has been associated in numerous studies with increased risk of mental disorders, though there is notably less research investigating these linkages compared with the impact of sexual abuse on later mental health. These studies variously show that exposure to childhood sexual abuse, especially severe abuse, is related to a significantly increased risk of a wide range of mental disorders, including anxiety disorders, in adulthood. Recent studies have shown linkages between childhood abuse and increased rates of panic disorder, panic attacks, social phobia, specific phobia, generalized anxiety disorder (GAD), and post-traumatic stress disorder (PTSD) among youth and adults. In sum, the evidence to date suggests a link between exposure to childhood physical abuse and increased risk of anxiety disorders.

Several studies have documented an association between childhood sexual abuse and increased risk of anxiety disorders. These studies variously show that exposure to childhood sexual abuse, especially contact sexual abuse, is related to a significantly increased risk of anxiety and anxiety disorders in adulthood. In a recent review of this literature, Fergusson and Mullen identified more than 30 studies that found that children known to have been sexually abused are vulnerable to a wide range of behavior problems, mental health disorders, and adjustment difficulties, including anxiety, fear, depression, and other negative attributes.

Family history of anxiety disorders has been shown to be a significant risk factor for the onset of anxiety disorders in offspring in clinical and high-risk samples. Results consistently suggest that having a parent with an anxiety disorder is associated with a significantly increased risk of anxiety disorders in offspring. For instance, Goldstein and colleagues found that panic disorder is more common among offspring of parents with panic and other anxiety disorders, compared with children with parents without anxiety disorders. Similarly, family studies have shown that the individual family member examined with first-degree relatives who have anxiety disorders have significantly higher rates of anxiety disorders, including social phobia, specific phobias, and agoraphobia, compared with those with never mentally ill relatives. Results on familial linkages of obsessive compulsive disorder (OCD) are relatively consistent but remain more mixed, while studies consistently show that panic is familial.

Previous clinical and cross-sectional studies of adults have shown that divorce/early parental separation and loss are associated with increased rates of anxiety disorders, compared with those who have not been exposed to these events. For instance, Davidson and colleagues found that adults with PTSD had significantly higher rates of parental separation before age 10, compared with those without PTSD.

Allen and colleagues examined the relationship between a range of prenatal and perinatal factors in the risk of anxiety and depressive disorders in a prospective, longitudinal study of 579 adolescents in the community. They found that fever and illness during the first year of life, as well as maternal history of miscarriage and stillbirth, were associated with an increased risk of anxiety disorders as were not being

breast-fed and maternal emotional problems during pregnancy.

Numerous studies have documented an association between psychiatric symptoms and mental disorders and increased risk of anxiety disorders, for instance, by Johnson, Cohen, and Brook. These studies variously show that the appearance of symptoms and mental disorders in early childhood, especially anxiety symptoms, is related to a significantly increased risk of anxiety and anxiety disorders in adulthood. Previous studies also show that depression in adolescence is associated with increased risk of anxiety disorders during young adulthood and suggest that anxiety disorders evident at an early age (ages 14 to 16) were associated with increased risk of later anxiety disorders. These associations persist after adjusting for differences in potentially confounding risk factors.

Genetic Factors

A recent meta-analysis by Kendler's group was performed to assess the level of heterogeneity between studies of the genetic epidemiology of anxiety disorders, as well as to combine data from multiple studies for a more powerful estimate of the heritabilities and familial risks of anxiety disorders. The results of this study confirmed that the relative risk of a later family member's developing most anxiety disorders ranges between 4 and 6. Analysis of twin studies suggests that the source of the risk is primarily genetic and found little support for common environment as the source of risk. The heritabilities for anxiety disorders are estimated at 30% to 40%. A number of genomic screens for panic disorder have been performed, with some modest evidence for linkage to several genomic regions. A single genome for OCD suggests linkage to a region on chromosome 9. A large number of candidate gene association studies have been performed for anxiety disorder, with the large majority being carried out in panic disorder data sets. Although generally underpowered, these studies have suggested some association between panic disorder and the genes for catechol-*O*-methyltransferase (COMT), monamine oxidase A (MAOA), the 2A adenosine receptor, and the Type B cholecystokinin receptor. Although these studies have not been adequately replicated, they provide interesting candidate genes for studies of anxiety phenotypes. Fewer such studies in OCD populations suggest the involvement of MAOA and COMT, the serotonin transporter, and the 1B serotonin receptor.

Previous clinical and cross-sectional studies of community samples have also shown linkages between specific personality traits (e.g., neuroticism) and increased rates of anxiety disorders, as well as increased severity, impairment, and poorer response to treatment associated with specific personality factors among patients with anxiety disorders. Moreover, cross-sectional community-based studies of adults have shown that personality traits are associated with increased rates of comorbidity among anxiety disorders, and depressive and substance use disorders, compared with those without these traits.

Outcomes Associated With Anxiety Disorders

Comorbidity is common among patients with anxiety disorders, and there is a relatively large literature on the comorbidity of anxiety disorders among adults in clinical and community samples. Clinical data suggest that comorbidity of mood and substance use disorders occurs among 50% to 80% of adult patients with anxiety disorders with similar rates suspected for pediatric patients; however, fewer data are available as information to investigate this speculation. Clinical data suggest that comorbid cases are more severe in terms of number of lifetime disorders, earlier onset, greater family history of anxiety and other mental disorders, poorer response to treatment, increased number of life events, and longer duration of disorder. Epidemiologic studies to date are largely consistent with these findings and are based on data from cross-sectional studies of population-based samples of adults in the community.

Previous studies have shown cross-sectional associations between anxiety disorders and a range of negative psychosocial outcomes. These include suicidal ideation and suicide attempts, occupational disability, higher rates of medical and psychiatric service utilization, physical morbidity, premature mortality, victimization, crime, poor partner relations, educational achievement, lower socioeconomic status, and others.

A small number of studies have documented an association between early anxiety problems and anxiety disorders and poorer educational achievement. These studies variously show that anxiety problems and anxiety disorders in childhood and adolescence

are related to a significantly lower level of educational achievement in youth and young adulthood.

A number of studies have documented an association between anxiety disorders and impairment in occupational functioning and occupational disability. These studies variously show that anxiety disorders, including panic disorder, social phobia, and generalized anxiety disorder, are related to a significantly increased risk of occupational disability compared with those without.

There have been several studies documenting an association between anxiety disorders and increased risk of suicidal ideation and suicide attempts. These studies variously show that anxiety disorders, especially panic attacks and panic disorder, are related to a significantly increased risk of suicidal ideation and suicide attempts. In addition, a recent literature review showed a significant association between panic and risk of suicide behavior.

—Renee Goodwin

See also Child and Adolescent Health; Intimate Partner Violence; Psychiatric Epidemiology; Suicide; Violence as a Public Health Issue

Further Readings

Allen, N. B., Lewinsohn, P. M., & Seeley, J. R. (1998). Prenatal and perinatal influences on risk for psychopathology in childhood and adolescence. *Development and Psychopathology, 10*(3), 513–529.

Davidson, J. R., Hughes, D., Blazer, D. G., & George, L. K. (1991). Post-traumatic stress disorder in the community: An epidemiological study. *Psychological Medicine, 21*(3), 713–721.

Fergusson, D. M., & Mullen, P. E. (1999). *Childhood sexual abuse: An evidence based perspective.* Thousand Oaks, CA: Sage.

Goldstein, R. B., Wickramaratne, P. J., Horwath, E., & Weissman, M. M. (1997). Familial aggregation and phenomenology of ''early''-onset (at or before age 20 years) panic disorder. *Archives of General Psychiatry, 54*(3), 271–278.

Johnson, J. G., Cohen, P., & Brook, J. S. (2000). Associations between bipolar disorder and other psychiatric disorders during adolescence and early adulthood: A community-based longitudinal investigation. *American Journal of Psychiatry*, *157*, 1679–1681.

Kendler, K. S., Walters, E. E., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1995). The structure of the genetic and environmental risk factors for six major psychiatric disorders in women: Phobia, generalized anxiety disorder, panic disorder, bulimia, major depression, and alcoholism. *Archives of General Psychiatry, 52*, 374–383.

Apgar Score

The Apgar score was devised in 1952 by Dr. Virginia Apgar (1909–1974) as a quick and simple method of assessing the condition of newborn infants. Typically an infant is assessed at 1 and 5 min after birth in five areas of functioning, each of which is assigned a score from 0 to 2. The five scores are summed, and a higher score indicates better health; a score of 7 or higher (out of a possible 10) indicates good to excellent health. The five areas of functioning, with their mnemonic in parentheses, are skin color (**a**ppearance), heart rate (**p**ulse), reflex irritability (**g**rimace), muscle tone (**a**ctivity), and respiration (**r**espiration). These areas were selected from a larger list of objective signs of infant health because delivery room personnel could be easily taught to evaluate them using the 0 to 2 scale.

The first use of the Apgar scale in research was a review in 1953 by Virginia Apgar of 1,025 infants born alive at Columbia Presbyterian Hospital in New York City. She found that Apgar score was related to the type of birth (higher scores were associated with vaginal births with the occiput presenting, and lower scores with breech extraction and version) and the use of anesthesia (which was associated with lower scores). Lower scores were also associated with higher neonatal death rates: Mature infants with scores of 0 to 2 had a 14% death rate, those scoring 3 to 7 a 1.1% death rate, and those scoring 8 to 10 a 0.13% death rate.

The Apgar score was quickly adopted for both infant assessment and research purposes, and the 5-min evaluation became common after a study of more than 54,000 births occurring during 1959 to 1966, in which the Apgar score at 5 min was found to be more predictive of neonatal mortality than the 1-min score. This was confirmed in an analysis of more than 150,000 infants born in Texas in the years 1988 to 1998, which found a strong correlation between the 5-min Apgar score and neonatal mortality. Preterm births with 5-min Apgar scores of 0 to 3 had 315 neonatal deaths per 1,000, those with scores of 4 to 6 had 72 neonatal deaths per 1,000, and those with scores of 7 to 10 had 5 deaths per 1,000.

For full-term births (37 weeks of gestation or later), those with 5-min Apgar scores of 0 to 3 had 244 neonatal deaths per 1,000, those with scores of 4 to 6 had 9 deaths per 1,000, and those with scores of 7 to 10 had 0.2 deaths per 1,000.

The Apgar score has been criticized because it has been used for purposes for which it was never intended—for instance, to predict the neurologic development of the infant or to serve as an indicator of hypoxia. Apgar scores are also influenced by several extraneous factors, including the maturity of the infant (a healthy preterm baby may receive a low score because of its immaturity) and maternal use of certain medications. However, the usefulness of the Apgar score for its original purpose, to quickly evaluate the health of a newborn infant, has been reaffirmed in recent studies and it is used all over the world for this purpose.

—Sarah Boslaugh

See also Maternal and Child Health Epidemiology; Reproductive Epidemiology

Further Readings

American College of Obstetricians and Gynecologists. (1996). Use and abuse of the Apgar score. *International Journal of Gynecology and Obstetrics, 54*, 303–305.

Casey, B. M., McIntire, D. D., & Leveno, K. J. (2001). The continuing value of the Apgar score for the assessment of newborn infants. *New England Journal of Medicine, 344*(7), 467–471.

Finster, M., & Wood, M. (2005). The Apgar score has survived the test of time. *Anesthesiology, 102*(4), 855–857.

Applied Epidemiology

The term *applied epidemiology* has been defined according to five core purposes: (1) the synthesis of the results of etiologic studies as input to practice-oriented policies; (2) the description of disease and risk-factor patterns as information used to set priorities; (3) the evaluation of public health programs, laws, and policies; (4) the measurement of the patterns and outcomes of public health and health care; and (5) the communication of epidemiologic findings effectively to health professionals and the public.

Many in epidemiology and public health view the linkage between etiologic research and public health intervention as implicit. However, it was observed more than two decades ago that the discipline of epidemiology has become increasingly divorced from activities in the real world that result in the improvement of public health. The concept of applied epidemiology has developed in an attempt to address concerns that epidemiology is not responding adequately to the concerns of public health practitioners. A major challenge for applied epidemiology is to improve overall health status by encouraging and measuring the effects of policy change, economic incentives, and behavioral interventions. There is also a shortage of epidemiologists in public health settings. Making epidemiology more relevant to public health practice will require training programs that provide practicing professionals with at least a basic understanding of epidemiologic methods and ways of accurately interpreting the large body of scientific literature.

Background and Historical Evolution

Epidemiology is often considered the key scientific underpinning of public health practice. The pivotal role of epidemiology was emphasized by the Institute of Medicine in its definition of the substance of public health as ‘‘organized community efforts aimed at the prevention of disease and promotion of health. It links many disciplines and rests upon the scientific core of epidemiology’’ (Committee for the Study of the Future of Public Health, 1988, p. 41).

Since 1927, dozens of definitions of epidemiology have been put forth. A widely accepted version is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems. Perhaps the most comprehensive definition, and the one most relevant to public health practice, and applied work was crafted by Terris (1992):

> Epidemiology is the study of the health of human populations. Its functions are:
>
> 1. To discover the agent, host, and environmental factors which affect health, in order to provide the scientific basis for the prevention of disease and injury and the promotion of health.

2. To determine the relative importance of causes of illness, disability, and death, in order to establish priorities for research and action.
3. To identify those sections of the population which have the greatest risk from specific causes of ill health, in order that the indicated action may be directed appropriately.
4. To evaluate the effectiveness of health programs and services in improving the health of the population. (p. 912)

Each of these four functions has a direct application toward improving the overall health of the population. Recognition of epidemiology's role in improving the overall health of the public was not consistently present in early definitions.

Critics of modern epidemiology acknowledge that the science has produced information essential for understanding disease etiology and decreasing the burden of disease, yet many important public health issues are left unaddressed, and the potential role of communities has been inadequately considered. The so-called risk-factor epidemiology approach involves the search for multiple antecedent factors at the individual level (e.g., smokers at high risk of lung cancer), without necessity for determining intervening factors (e.g., the policies that lead to high smoking rates). To some degree, this approach ignores the multilevel, environmental determinants of disease (or health) and may obscure opportunities for intervention.

Using Epidemiology in Applied Settings

Epidemiologists have at their disposal an increasingly large array of tools that can enhance the application of epidemiology within public health practice. These include methodological advances that offer us more sophisticated ways to evaluate the health risks associated with many exposures and environmental contaminants in modern society. New information technologies, including the rapid evolution of microcomputers, software, and the Internet, offer exciting possibilities and quick access to data. In addition, changes in how health care is delivered, particularly the growth of managed care, provide new chances for epidemiologists to become involved in the assessment of health care utilization and quality. Despite these vast possibilities, there are many examples where decision making and health policy making in public health and health care occur in the absence of sound epidemiologic data and scientific reasoning. While epidemiology contributes greatly to the discovery of new knowledge for public health, it falls short in translating existing knowledge to improve practice and the health of the public.

Epidemiology is an important discipline to aid in bridging the gap between science and practice. Various reports and survey data support the importance of epidemiology in the public health setting. A survey of 40 state health agency directors found that among 11 key areas, epidemiology was rated as having the highest importance to respondents (a mean of 9.5 on a 10-point scale). In contrast, the percentage of respondents who believed that research needs in epidemiology were being met by universities was much lower (a mean of 4.4 on a 10-point scale).

Personnel and Training Needs

Although obtaining accurate estimates of personnel needs is difficult and relatively little empirical data exist, it is widely accepted that a shortage of trained epidemiologists has existed in public health agencies for several decades. It is also likely that the continually growing demand for quality measurement in health care will increase the need for epidemiologists in the private and nonprofit sectors. In a recent study of 37 state health agencies, 40.5% identified a shortage of epidemiologists. The shortage of master's- and doctoral-trained epidemiologists may be most acute for noninfectious disease epidemiologists. As of 2003, 43% of responding states did not appear to have a state chronic disease epidemiologist or person recognized as such even if not formally titled.

Training of epidemiologists occurs through a variety of mechanisms. Many epidemiologists at the master's and doctoral levels are trained by schools of public health. Other important sources include schools of medicine and the Epidemic Intelligence Service of the Centers for Disease Control and Prevention Survey data, and expert groups have also shown the need for expanded and perhaps different formal training in epidemiology. Successful educational programs need to maintain close contact with public health practice. Training and job categories in epidemiology are inconsistent. For example, 42% of the current epidemiologic workforce lacks formal academic training in epidemiology. There are promising developments in training across the globe, including Field Epidemiology Training Programs and Public Health

Schools Without Walls. The current epidemiologic and technologic advances provide unprecedented opportunities for practicing health professionals. To take full advantage of these opportunities, continued skill enhancement will be necessary.

—*Ross C. Brownson*

See also Community Trial; Competencies in Applied Epidemiology for Public Health Agencies; Descriptive and Analytic Epidemiology; Governmental Role in Public Health; Outbreak Investigation

Further Readings

Association of State and Territorial Health Officials. (2004). *State Public Health Employee Worker Shortage Report: A civil service recruitment and retention crisis*. Washington, DC: Author.

Brownson, R. C., & Petitti, D. B. (Eds.). (2006). *Applied epidemiology: Theory to practice* (2nd ed.). New York: Oxford University Press.

Committee for the Study of the Future of Public Health, Institute of Medicine. (1988). *The future of public health*. Washington, DC: National Academy Press.

Council of State and Territorial Epidemiologists. (2004, September). *National assessment of epidemiologic capacity in chronic disease: Findings and recommendations*. Atlanta, GA: Author.

Last, J. M. (Ed.). (2001). *A dictionary of epidemiology* (4th ed.). New York: Oxford University Press.

Lilienfeld, A. M., & Lilienfeld, D. E. (1982, June). Epidemiology and the public health movement: A historical perspective. *Journal of Public Health Policy*, 140–149.

Morris, J., Schneider, D., & Greenberg, M. R. (1994). Universities as resources to state health agencies. *Public Health Reports, 109*, 761–766.

Terris, M. (1992). The Society for Epidemiologic Research (SER) and the future of epidemiology. *American Journal of Epidemiology, 136*, 909–915.

Thacker, S. B., Dannenberg, A. L., & Hamilton, D. H. (2001). Epidemic intelligence service of the Centers for Disease Control and Prevention: 50 years of training and service in applied epidemiology. *American Journal of Epidemiology, 154*, 985–992.

Arthritis

Arthritis is a general term that technically means inflammation of the joint(s), but the terminology is somewhat misleading because arthritis also refers broadly to a wide range of joint-related conditions—not all of which involve inflammation per se. Joints have six major components: (1) cartilage, (2) synovial membrane, (3) bursa, (4) muscle, (5) tendon, and (6) ligament. Problems with the functioning of any of these joint components may be described as "arthritis." This entry examines the epidemiology, clinical management, and social and cultural impact of arthritis, focusing in particular on osteoarthritis, rheumatoid arthritis, and juvenile rheumatoid arthritis.

Epidemiology and Clinical Manifestations

The term *arthritis* covers more than 100 different medical conditions, and approximately one third of the population in the United States is affected by arthritis, which is the leading cause of disability. The primary causes for most forms of arthritis are not completely known, but arthritis occurs more commonly in women than in men, and the prevalence of most forms of arthritis increases with advancing age. Nevertheless, arthritis can affect all age groups, including children. Common subtypes of arthritis include osteoarthritis, rheumatoid arthritis, and juvenile rheumatoid arthritis.

Osteoarthritis

Osteoarthritis (OA), sometimes called degenerative joint disease, is the most common form of arthritis and is closely associated with the aging process. OA is associated with worn or frayed cartilage, which subsequently fails to properly cushion the joint. The typical clinical manifestations include painful joints, stiffness, and difficulty with mobility. Prevalence estimates for OA are highly dependent on the criteria used to define it, such as symptomatic pain, radiographic evidence, or self-reported symptoms. However, the incidence of hip arthritis is approximately 88 cases per 100,000 person-years; the incidence of knee arthritis is approximately 240 cases per 100,000 person-years; and the incidence of hand arthritis is approximately 100 per 100,000 person-years. Due to the implications for weight bearing, the presence of radiographic evidence of OA in the knees or hips is often associated with substantial functional impairment. The known risk factors for OA include advancing age, female gender, obesity, major joint trauma, repeated overuse, heredity, prior inflammatory disease, and developmental abnormality. A role for heredity in OA, although

complex, also is suspected because persons whose parents had OA are at higher risk for developing OA themselves. Occupations that involve high physical demands or repetitive movements are associated with a higher incidence of OA.

Rheumatoid Arthritis

Rheumatoid arthritis (RA) affects about 1% of the worldwide population, and approximately 75% of those affected are female. RA is a condition involving chronic inflammation of the synovial membrane that lines the joint. As a result, the joint becomes swollen, tender, and warm; inflammatory activity may eventually cause irreversible damage to cartilage and/or bone. RA typically presents in a symmetrical manner with both sides of the body being similarly affected; the joints of the wrists and knuckles are almost always involved. Persons with RA also commonly experience systemic symptoms such as fatigue, aching muscles, and even a low-grade fever. The major risk factors for RA are advancing age and female gender. Onset is most commonly in the fifth and sixth decades of life, but it can occur at any age. Both genetic and hormonal factors are thought to play a role in the development of RA. Family studies reveal a higher risk for RA among first-degree relatives as compared with persons who are unrelated, although this relationship is not pronounced. With regard to hormonal risk factors, the picture is not entirely clear, but symptoms of RA often remit during pregnancy, and recurrence is likely after birth. Studies have been mixed regarding the role of oral contraceptives and postmenopausal estrogens, but some investigations have suggested that these therapies may lower the risk for development of RA.

Juvenile Rheumatoid Arthritis

As the name implies, juvenile rheumatoid arthritis (JRA) affects children prior to age 16. Population-based studies indicate that the prevalence of JRA is approximately 1 to 2 per 1,000 children; the incidence is 11 to 14 new cases per 100,000 children. The cause of JRA is unknown, but several subsets of the condition exist. Interestingly, onset prior to age 6 typically involves females; onset at an older age typically involves males. As is the case with most other forms of arthritis, the risk factors are unclear, but heredity appears to play a role due to the notably higher incidence in monozygotic versus dizygotic twins. Infectious agents (e.g., rubella virus and Lyme disease) also have been suggested as possible triggering agents for JRA. Life expectancy for adults with JRA is lower than for the general population.

Clinical Management

Due to the wide range of conditions that are subsumed under the term *arthritis*, no single treatment regimen will apply to all. However, for most forms of arthritis, a multimodal approach is employed, which may involve patient education, especially in the form of "self-management" programs. Other common clinical interventions include pharmacologic management, exercise (both strengthening and aerobic), physical modalities (heat and cold), splinting, surgery (including joint replacement), and cognitive-behavioral training, among others. Clinical intervention is also frequently directed at secondary problems such as pain, fatigue, sleep disturbance, mood disorders, deconditioning, treatment adherence, and work disability.

Social and Economic Impact

Arthritis, in all forms, exacts a heavy toll in terms of physical, social, and psychological consequences. The physical consequences include pain, joint stiffness, and limitations of mobility, which can secondarily affect personal care, household activities, and occupational status. Social consequences can include role limitations, increased social isolation, and loss of leisure activities. Psychological consequences can include a higher incidence of depression, anxiety, and perceptions of helplessness, particularly in the context of a frequently unpredictable disease course. These collective physical, social, and psychological impacts all conspire to make arthritis the leading cause of disability in the United States. Lower levels of formal education and lower income levels, possibly due to less adaptive resources, have been identified as risk factors for arthritis-related disability. The economic impact of arthritis is enormous due to the high prevalence of the various conditions, the chronicity of the symptoms, the implications for high health care utilization, and the frequency of losses in gainful employment. The economic impact of arthritis in the United States has been estimated

to be approximately 2.5% of the gross national product.

—*Jerry C. Parker*

See also Aging, Epidemiology of; Disability Epidemiology

Further Readings

Harris, E. D., Budd, R. C., Genovese, M. C., Firestein, G. S., Sargent, J. S., Ruddy, S., et al. (Eds.). (2005). *Kelly's textbook of rheumatology* (Vols. 1 and 2, 7th ed.). Philadelphia: Elsevier Saunders.

Robbins, L., Burckhardt, C. S., Hannan, M. T., & DeHoratius, R. J. (Eds.). (2001). *Clinical care in the rheumatic diseases* (2nd ed.). Atlanta, GA: Association of Rheumatology Health Professionals.

Asbestos

Asbestos, a mineral used as an insulator and fire retardant during much of the first half of the 20th century, is now a nearly ubiquitous environmental contaminant. Found frequently in buildings, asbestos may be present in a broad range of construction materials, from concrete blocks to flooring.

Asbestos is a human health threat because a single submicroscopic fibril, thousands of which make up a single asbestos fiber, can cause severe lung disease. Only when asbestos is pulverized are the dangerous fibrils released into a dust cloud, which may be inhaled or ingested. The safe removal of asbestos is therefore a delicate and time-intensive process and presents a high risk of exposure. Current practice is typically to leave asbestos undisturbed. It is estimated that billions of dollars would be required to eliminate asbestos from all affected buildings.

Asbestos exposure is most commonly associated with occupational exposure, though persons living in areas with naturally occurring asbestos or man-made mines may also be exposed. Historically, high-risk occupations have included mining, milling, working with insulation, shipbuilding/repairing, and working in the textile industry. Men are more commonly exposed to asbestos and, therefore, are more frequently affected by negative sequelae associated with exposure.

The most common diseases caused by asbestos exposure are mesothelioma, asbestosis, and lung cancer. If ingested, asbestos fibrils may also cause stomach cancer. Mesothelioma is a very rare form of cancer of the mesothelium, the membrane lining the various body cavities, which most commonly affects the mesothelium of the lungs or chest cavity, but may also affect the mesothelium of the abdomen or heart. Mesothelioma is nearly always associated with asbestos and is not related to smoking. Asbestosis occurs when asbestos fibrils irritate and scar lung tissue. This damage decreases the ability of the lungs to oxygenate blood, causing shortness of breath. There is no treatment for asbestosis, a progressive disease. The many diseases associated with asbestos exposure have a long latency period, typically many decades. Thus, studies assessing the risk of various health outcomes among asbestos-exposed persons require adequate follow-up time.

Asbestosis is a member of the group of diseases termed *pneumoconioses*. These diseases—coal workers' pneumoconiosis (CWP), silicosis, asbestosis, mixed dust pneumoconiosis, graphitosis, and talcosis—all result from inhaling mineral particles, result in changes in lung tissue, and have no current treatment. From 1968 to 2000, deaths due to pneumoconioses decreased overall and for all specific subtypes except asbestosis. Rather, mortality due to asbestosis increased steadily from 77 to 1,493 during this period. The annual age-adjusted death rate for asbestosis increased from 0.54 per million population in 1968 to 6.88 per million population in 2000. Asbestosis increased throughout the United States during this time, though the coastal states saw the most marked rise, likely due to the shipbuilding industry. Since asbestosis mortality usually occurs 40 to 45 years after exposure, this upward trend in asbestosis mortality likely dates back to the post–World War II era. However, it is not expected that this trend will begin to reverse for at least several years, since asbestos use peaked in the United States in 1973. The peak in asbestosis deaths in the United States will likely be seen around 2013 to 2018.

Beginning in the 1970s, studies indicated that the risk of lung cancer among people exposed to asbestos was increased beyond what would be expected by cigarette smoking (synergism). While is remains clear that lung cancer risk among asbestos-exposed persons differs by smoking status (effect modification), the underlying relationship has been debated recently—namely, whether this model is additive or multiplicative. According to the U.S. Environmental Protection Agency (EPA), cigarette smoking and asbestos exposure interact synergistically in the development of lung cancer. Estimates

of lung and pleural cancer risk following asbestos exposure have ranged from 1.9 to 28.0.

Asbestos is still used worldwide, though it has been limited to primarily commercial uses in the United States and most Western European countries; some European countries have banned its use altogether. In the United States, asbestos use peaked in 1973 at 719,000 metric tons. In 1999, when asbestos use was limited to commercial roofing materials, gaskets, and friction products such as brake shoes and clutches, 15,000 metric tons were used in the United States. Worldwide, 1.93 million metric tons of asbestos were produced in 1999, compared with a 1975 peak global production of 5.09 million metric tons. Many countries rely on asbestos for inexpensive building materials, made by combining cement with asbestos for added strength. Asbestos cement products tend to carry a lower risk of exposure since the cement is relatively effective at securing the asbestos fibers within the solid.

Asbestos may also be found in vermiculite, a mineral used for its fire-resistant and absorbent properties. One of the most recent and most publicized cases of asbestos exposure involved a vermiculite mine in Libby, Montana. In late 1999, following public concern and media reports of asbestos contamination, the EPA and other federal agencies began environmental assessments in Libby, the site of a vermiculite mine and two former vermiculite processing plants. Mining of asbestos-contaminated vermiculite occurred in Libby from the 1920s until the mine's closure in 1990. It is estimated that 80% of the world's vermiculite came from the mine in Libby. Libby was added to the National Priorities List (NPL) in 2002, after which 3,500 properties in the area were inspected. Cleanup work has already begun in Libby, with an expectation that as many as 1,400 properties will require remediation.

In 1989, the United States passed a ban on asbestos-containing materials (ACM). Because the ban was overturned in 1991, however, much of this ban was never implemented. Several categories of ACM were banned after the 1991 decision: flooring felt, rollboard, and corrugated, commercial, or specialty paper. In addition, no new uses of asbestos are permitted in the United States. The EPA primarily regulates asbestos use under the Clean Air Act and the Toxic Substances Control Act.

—Erin L. DeFries

See also Cancer; Environmental and Occupational Epidemiology; Exposure Assessment

Further Readings

Attfield, M. D., Wood, J. M., Antao, V. C., & Pinheiro, G. A. (2004, July 23). Changing patterns of pneumoconiosis mortality: United States, 1968–2000. *Morbidity and Mortality Weekly Report, 53*, 627–632.

Brownson, R. C., Remington, P. L., & Davis, J. R. (Eds.). (1998). *Chronic disease epidemiology and control* (2nd ed.). Washington, DC: American Public Health Association.

Levy, B. S., & Wegman, D. H. (Eds.). (1983). *Occupational health*. Boston: Little, Brown.

Liddell, F. D. K. (2001). The interaction of asbestos and smoking in lung cancer. *Annals of Occupational Hygiene, 45*, 341–356.

National Institute of Occupational Safety and Health. (2002). *The work-related lung disease surveillance report, 2002: Asbestosis and related exposures*. Retrieved June 12, 2007, from http://www.cdc.gov/niosh/docs/2003-111/2003-111.html.

U.S. Environmental Protection Agency. (2006). *Region 8: Libby asbestos*. Retrieved February 7, 2007, from http://www.epa.gov/region8/superfund/libby/background.html.

U.S. Geological Survey. (2001, March). *Asbestos fact sheet*. Retrieved June 12, 2007, from http://pubs.usgs.gov/fs/fs012-01.

Web Sites

U.S. Environmental Protection Agency: http://www.epa.gov/asbestos.

Asian American/Pacific Islander Health Issues

Persons in the United States with ancestral origins in Asia or the Pacific Islands (Polynesia, including Hawaii, Micronesia, Melanesia, and other islands) have often been grouped together for the purpose of gathering health information and calculating statistics. However, this is an extremely heterogeneous group, with national origins that cover one third of the globe. Until recently, there have been relatively few persons in the United States from Asia, and grouping them together was a way to fit these populations into the racial classification system used in data collection in the United States. There is widespread recognition that describing the group in

aggregate might mask important health issues that are particular to one or more subgroups. The Asian American and Pacific Islander population (AAPI) comes from very diverse environments, and their experiences in the migration process and living in the United States have also varied greatly. The economic status of these groups varies greatly: The Japanese and Asian Indians are among the wealthiest ethnic groups in the United States, while persons from Southeast Asia have very high rates of poverty. Pacific Islanders have lower average educational attainment and income than Asian Americans. Ideally, health issues would describe each national origin population separately, such as the Thai, Filipino, or Samoan. Practically, though, data are rarely available to do this, and an important health issue for Asian Americans and Pacific Islanders is the inadequacy of the data.

Data Issues

The aggregate AAPI category is a relatively recent invention, recommended in a 1977 federal document as part of a new effort to collect uniform statistics by race and ethnicity across all federal agencies. The 1980 census was the first census to use the AAPI category. In the 1990s, there were two changes in federal recommendations for race and ethnicity data collection that are important for the AAPI category, both of which are seen in the 2000 census and some other recent data sources. First, Asian Americans and Pacific Islanders were separated into two race categories. Second, persons were allowed to check more than one race code. The multiracial provision has a large impact on AAPI data because about 15% of persons who checked an Asian American race code and 55% of persons who checked a Pacific Islander race code also selected another race code in the 2000 census. Pacific Islanders are the race group most likely to check more than one category. Because it is unclear how to count the multiracial persons, it is difficult to track the size of the Asian American or Pacific Islander populations or to use the census as a denominator when calculating a health indicator when the numerator data source lacks multiple race codes. The inconsistency across data sources in racial or ethnic classification is a serious limitation in comparing data over time or across different study populations; some data are reported with one race category (AAPI), some with two race categories (Asian American and Pacific islander), and some with ethnic subgroup categories (such as Chinese). Sometimes, identification by ethnically distinctive surname is used to infer subgroup in data sources that provide only aggregate racial identification or no racial identification.

There is another important limitation to health data for AAPI. Asian Americans constituted 4.2% of the population in 2000 and Pacific Islanders 0.3%. Nationally representative health surveys, even large ones with several thousand respondents, do not include adequate numbers of participants in the individual AAPI subgroups to permit estimates of health factors, and they may not even have enough participants to calculate stable rates for the aggregate race categories. This is the case for the National Health and Nutrition Examination Surveys. Beginning in the 1980s, many health-related data collections oversampled African Americans and Latinos to ensure adequate numbers for analysis, and there were also a number of epidemiologic studies that exclusively targeted these groups. However, Asian Americans were perceived as a ''model minority'' without health problems requiring special research and public health efforts, and until very recently, there was neither oversampling in large data collections nor targeted studies. A notable exception is the Honolulu Heart Program initiated in 1965 by the National Institutes of Health as a prospective cohort study of environmental and biological causes of cardiovascular disease among Japanese Americans living in Hawaii. More recently, a few large epidemiologic studies have oversampled Asian American populations, including the National Longitudinal Study of Adolescent Health (Add Health), the Multi-Ethnic Study of Atherosclerosis (MESA), and the Study of Women's Health Across the Nation (SWAN). Chinese Americans are best represented in these studies.

Asian American Health Issues

By many standard health indicators, Asian Americans are indeed a very healthy population. Age-specific death rates reported by the National Center for Health Statistics are much lower for Asian Americans than for any of the other race categories and life expectancy is higher, although there are concerns with the completeness and accuracy of Asian American identification on the death certificates used to calculate the death rate numerators. Asian Americans are less likely

to die of heart disease, diabetes, and stroke than non-Hispanic whites. Infant mortality is also lower for Asian Americans than for non-Hispanic whites, although average birthweight is also lower. Because a high proportion of adult Asian Americans are foreign-born, the "healthy immigrant effect" is often cited when discussing the health of this group. The healthy immigrant effect comes from the observation that recent immigrants tend to be in very good health, presumably because of the powerful selection factors associated with their motivation to migrate and their success in the difficult migration process.

Nonetheless, there are several health issues specific to Asian American populations. Since a high proportion of adults are migrants, they may face significant cultural and linguistic barriers accessing health care. Limited English proficiency is common among the adults in some subgroups, particularly persons from Southeast Asia and China. There are also high rates of noninsurance of health among Asian migrants.

Infectious Diseases

The infectious diseases recognized as particular concerns among Asian Americans are tuberculosis and hepatitis B. Many adult Asian Americans are migrants from countries where these infections have a much higher prevalence than in the United States.

Asian Americans have the highest tuberculosis rate of all race groups in the United States, estimated to be as high as 20 times the rate among non-Hispanic whites. Almost one fourth of new tuberculosis cases in the United States occur among Asian Americans, mostly among the foreign-born.

Hepatitis B is more common among Asian Americans than any other race group. Hepatitis B virus can lead to cirrhosis (scarring) of the liver and liver cancer. Hepatitis B is transmitted through bodily fluid, including blood. One mode of transmission is from an infected mother to her infant in childbirth, and this is a common mode of transmission in regions with high prevalence of hepatitis B, such as Southeast Asia, China, and Korea. The vaccine against hepatitis B, introduced in most countries in the 1990s, is effective at blocking mother-child transmission when administered to the infant at birth. However, persons born in some parts of Asia before that have a high likelihood of hepatitis B infection. There have been estimates that chronic infection with hepatitis B may be as high as 15% among some Asian American populations. Asian Americans may also have relatively high rates of hepatitis C, but data are limited.

Cancer

Cancer is the leading cause of death for Asian Americans, although it should be noted that more Asian Americans die of cancer than of heart disease not because the cancer death rate is so high but rather because the heart disease death rate is low. However, there are several cancer sites for which Asian Americans do have high risk compared with most other race groups. These are liver cancer, cervical cancer, and stomach cancer—all cancers of infectious origin. The few studies that allow the disaggregation of Asian Americans, though, suggest that risk is not uniformly high across all Asian American subgroups for these cancers. Liver cancer rates are particularly high for Vietnamese, Koreans, and Chinese. Stomach cancer rates are very high for Japanese and Koreans and somewhat high for Chinese and Vietnamese. Cervical cancer risk is high for Vietnamese women and may be high for Asian Indians. Because cervical cancer is preventable through early detection, there have been concerted efforts to increase rates of screening with a Papanicolaou smear, particularly among women from Southeast Asia.

Coronary Heart Disease

Heart disease rates are very low among all Asian American groups except Asian Indians. Asian Indians, whether living in South Asia or elsewhere, have among the highest rates of coronary heart disease in the world. Although the Asian Indian population is the most rapidly growing subgroup in the United States, increasing by more than 100% from 1990 to 2000, there has been little epidemiologic research in the United States. Many Asian Indians have very high levels of the blood fat triglyceride and low levels of high-density lipoproteins (the "good cholesterol"), both well-recognized risk factors for coronary artery disease. There are also high rates of diabetes among Asian Indians, another risk factor for coronary heart disease. Some recent research suggests that genetic factors may play a role in coronary heart disease risk for Asians Indians.

Diabetes

Some Asian American subgroups appear to have a risk of type 2 diabetes as high or higher than

non-Hispanic whites. These groups are Asian Indians and possibly Japanese and Filipinos. Higher body mass index is associated with increased risk of type 2 diabetes in all populations, but there is some evidence that the relationship is different for Asians, regardless of geographic location. Because of greater percent body fat and visceral adiposity at the same level of body mass index, Asian Americans may have increased diabetic risk at lower levels of body mass index than white populations.

Health Behaviors

Data are limited on health behaviors for Asian American subgroups. The difference between smoking rates for men and women is much larger for Asian Americans than other race groups, reflecting the high prevalence of smoking among men in East and Southeast Asia (but not South Asia) and very low prevalence of smoking among women with the same background. Men from Southeast Asia, Korea, and China have high rates of smoking in the United States. Several studies suggest that more than half of men from Southeast Asia smoke. While smoking is more prevalent among Asian American women born in the United States than among migrants, it is still less common than among non-Hispanic white women.

Asian Americans have lower rates of obesity than other race groups. However, obesity risk is not uniform across subgroups; limited data suggest that Asian Indian women and Japanese men have rates of obesity similar to non-Hispanic whites. Foreign-born Asian Americans are less likely to engage in leisure-time physical activity than persons born in the United States and are no more likely to engage in nondiscretionary physical activity (for transportation or employment-related), which would compensate for lower levels of discretionary activity.

Mental Health Issues

Many small studies have investigated mental health issues among specific populations of Asian Americans, particularly depression and stress related to acculturation. However, cultural differences in the expression, recognition, and description of depression and other mental health problems make it difficult to develop multilingual survey instruments that will function similarly across different populations. Asian Americans appear to have relatively low rates of utilization for mental health services, and there is concern that this represents underutilization. Refugees from Southeast Asia have a high risk of post-traumatic stress syndrome.

Acculturation and Health

Many studies have examined how acculturation affects the health of Asian Americans. The conceptualization and measurement of acculturation are not consistent across studies, with some using a single measure, such as years in the United States or home language use, and others using instruments that include several questions and try to measure an underlying construct of cultural orientation. In general, studies of physical health find declining health and worsening health behaviors with increasing years since migration, or other measures of acculturation. Studies of mental health have conflicting findings.

Pacific Islanders

While data are limited for describing Asian American health, they are much more limited for Pacific Islanders. Pacific Islanders have been nominally included in studies of AAPI, but there is often little information specific to them. Most Pacific Islanders in the United States originate in islands that are part of the United States (Hawaii, Guam, and Samoa) and are therefore not international migrants, even if they have moved to the mainland. They have not experienced the selection pressures that probably account for the good health of migrant Asian Americans.

Pacific Islanders generally experience poorer health than both Asian Americans and non-Hispanic whites. Available data suggest generally high rates of both infectious and noninfectious diseases among Pacific Islanders. They are more likely to develop and die from cancer, heart disease, and diabetes. Obesity rates are very high among native Hawaiians and Samoans. Factors contributing to poor health may include limited access to health care, cultural barriers, and poor nutrition. Rates of hepatitis B and tuberculosis are high for Pacific Islanders, as they are for Asian Americans, and HIV/AIDS may also be higher than among non-Hispanic whites.

—Diane Sperling Lauderdale

See also Honolulu Heart Program; Immigrant and Refugee Health Issues; Race and Ethnicity, Measurement Issues With; Race Bridging

Further Readings

Chen, M. S., Jr. (2005). Cancer health disparities among Asian Americans: What we do and what we need to do. *Cancer, 104*, 2895–2902.

Fadiman, A. (1997). *The spirit catches you and you fall down: A Hmong child, her American doctors, and the collision of two cultures*. New York: Farrar, Straus & Giroux.

Ghosh, C. (2003). Healthy people 2010 and Asian Americans/Pacific Islanders: Defining a baseline of information. *American Journal of Public Health, 93*, 2093–2098.

Kandula, N. R., Kersey, M., & Lurie, N. (2004). Assuring the health of immigrants: What the leading health indicators tell us. *Annual Review of Public Health, 25*, 357–376.

Kumanyika, S. K. (1993). Special issues regarding obesity in minority populations. *Annals of Internal Medicine, 119*, 650–654.

Lauderdale, D. S., & Kestenbaum, B. (2002). Mortality rates of elderly Asian American populations based on Medicare and Social Security data. *Demography, 39*, 529–540.

LaVeist, T. (2005). *Minority populations and health: An introduction to health disparities in the U.S.* San Francisco: Jossey-Bass.

Lew, R., & Tanjasiri, S. P. (2003). Slowing the epidemic of tobacco use among Asian Americans and Pacific Islanders. *American Journal of Public Health, 93*, 764–768.

Salant, T., & Lauderdale, D. S. (2003). Measuring culture: A critical review of acculturation and health in Asian immigrant populations. *Social Science and Medicine, 57*, 71–90.

Srinivasan, S., & Guillermo, T. (2000). Toward improved health: Disaggregating Asian American and Native Hawaiian/Pacific Islander data. *American Journal of Public Health, 90*, 1731–1734.

Uppaluri, C. R. (2002). Heart disease and its related risk factors in Asian Indians. *Ethnicity and Disease, 12*, 45–53.

Association, Genetic

Analytical epidemiologic studies investigate potential associations between exposures (e.g., risk factors) and outcomes (e.g., as identified in quantitative data on traits or diseases) of interest. Genetic association studies are a special type of analytical epidemiologic studies in which the exposure of interest is represented by a genetic factor. Genetic factors are usually represented by common forms of genetic variation, such as single nucleotide polymorphisms (i.e., genetic loci for which different DNA variants can be present among different individuals) or deletions or insertions (i.e., losses or additions of a sequence of DNA in a chromosome).

Identification of genetic associations can have important implications in the prevention, diagnosis, and treatment of diseases, including complex diseases such as cancer and cardiovascular illnesses. For example, identification of specific polymorphisms related to the development of disease may lead to prevention programs targeted to specific subgroups of individuals at high risk. Similarly, genetic variation that affects response to specific medications may be taken into account in selecting the most effective and least toxic pharmacologic treatments for individual patients.

Genetic associations can be investigated within several epidemiologic study designs, including the classic case-control study design. In a case-control study of genetic risk, individuals with and without a specific disease are compared according to their exposure to a specific genetic risk factor, usually whether they have or do not have a particular genetic variant. However, the genetic nature of the exposure influences the methodological approach used in these studies in several respects. One reason for this influence is that even in the simplest case of diallelic loci (i.e., genetic loci at which only two possible alternative alleles are present in the population), double copies of most polymorphisms are present in humans (with the exception of loci on the sex chromosomes), and they are located in proximity to other polymorphisms; therefore, genetic risk factors can be linked to disease in several different ways. For example, the disease of interest can be associated with

1. the frequency of a specific allele among chromosomes from cases (allele analysis);
2. the frequency of a specific combination of alleles at one locus among cases (genotype analysis);
3. the frequency of specific haplotypes (i.e., the sequence of alleles at different loci on the same chromosome) among chromosomes from cases (haplotype analysis); and
4. the frequency of specific combinations of haplotypes among cases (diplotype analysis).

In addition, because at the population level alleles can have different frequencies in different groups (e.g., ethnic groups) and can be assorted at different loci in a nonrandom fashion (i.e., they can be in linkage disequilibrium), the validity of associations between polymorphisms and disease in epidemiologic studies can be confounded by associations related to

population stratification across ethnic groups and/or by the existence of linkage disequilibrium between these polymorphisms and other nearby polymorphic loci. Despite these limitations, genetic association studies are usually less vulnerable to confounding and bias than classic epidemiologic studies. For example, in genetic case-control studies, the risk of recall bias is minimized and the temporal sequence between exposure and outcome is resolved because individual genomic variation is established from conception and, thus, must precede the outcome of interest (e.g., onset of disease).

Yet in the years following the completion of the Human Genome Project, most studies investigating genetic associations for complex diseases have produced variable results. The genetic dissection of complex diseases is proving quite challenging. One possible explanation for these inconsistencies is that the effects of genetic variation on disease risk are influenced by interactions between genes and the environment. Thus, capturing the complexity of these interactions will need to precede the identification of the genetic components of complex diseases.

—Stefano Guerra and F. John Meaney

See also Gene; Gene-Environment Interactions; Genetic Counseling; Genetic Epidemiology

Further Readings

Cordell, H. J., & Clayton, D. G. (2005). Genetic association studies. *Lancet, 366*, 1121–1131.

Khoury, M. J., Little, J., & Burke, W. (2003). *Human genome epidemiology: A scientific foundation for using genetic information to improve health and prevent disease.* New York: Oxford University Press.

Thomas, D. C. (2004). *Statistical methods in genetic epidemiology.* New York: Oxford University Press.

Association of Schools of Public Health

The Association of Schools of Public Health (ASPH) was founded in 1953 and is the only U.S. public health organization representing deans, faculty, and students of ASPH-accredited and associate schools. It is governed by its members and board of directors and currently has 39 member schools. It was created in response to the growth of graduate education programs in public health during the early 20th century, which led to a need to formalize a periodic accreditation process of national programs. The comprehensive accreditation process for schools of public health is now conducted by an independent agency, the Council on Education for Public Health (CEPH), which is recognized by the U.S. Department of Education.

ASPH schools prepare students for entry into careers in public health with a well-rounded approach that includes the major facets of public health. Member schools must be independent academic institutions that have a doctoral degree program and offer degrees in the core areas of public health. These areas include epidemiology, environmental health, biostatistics, health service administration, and health education/behavioral sciences.

At present, 15 types of degrees in specific areas of public health are offered. New areas of study include international and global health, biomedical and laboratory practice, and maternal and child health. Distance learning methods are becoming more popular, with several ASPH member schools offering degrees through distance learning alone.

The ASPH performs a number of functions to further public health and public health education. Its goals, as enumerated on the ASPH Web site, include facilitating communication among member schools, providing a focus and a platform for the enhancement of emerging academic public health programs, cooperating with the federal government to strengthen public health education and the public health profession, assisting in the development and coordination of national health policies, serving as an information center for governmental and private groups and individuals whose concerns overlap those of higher education for public health, and assisting in meeting national goals of disease prevention and health promotion.

Not all educational institutions for public health and health education in the United States are members of ASPH, because they have not gone through the formal review process administered by CEPH. The issues of accreditation, accountability, and cost-effectiveness are controversial, and there is an ongoing dialogue among various public health agencies concerning how to improve the quality of services to the public and how to raise the standard of public health services on a local and national level.

—Sean Nagle

See also Competencies in Applied Epidemiology for Public Health Agencies; Public Health, History of

Web Sites

Association of Schools of Public Health: http://www.asph.org.
Council on Education for Public Health: http://www.ceph.org.

Asthma

Asthma is an obstructive airway disorder characterized by reversible airway narrowing, mucus hypersecretion, chronic inflammation, and episodic shortness of breath. Severe recurrent bouts of asthma lead to airway scarring, termed *remodeling*, which is not reversible and can lead to an increased frequency and severity in asthma exacerbations and lung infections. Asthma can be allergen provoked, which is termed *atopic*, or asthma may be due to unknown factors and termed *intrinsic*. There is a broad consensus that the prevalence of asthma has increased dramatically in most industrialized countries over the past several decades, prompting the development of several asthma study cohorts that have attempted to address asthma incidence, etiology, prevention, and control. What has come to be appreciated in the past decade due to the use of conditional gene expression systems, and transgenic animals, is the true complexity of the multiple molecular pathways that are involved in the development and progression of asthma and the understanding that asthma is more accurately characterized as a collection of diseases comprising a syndrome rather than a distinct homogenous entity. Therefore, in terms of a cure, there is no single "magic bullet" with regard to the treatment or the prevention of asthma.

Incidence, Prevalence, and Cost Burden

In 2004, the incidence of asthma worldwide was estimated at 300 million people, and it was predicted that by the year 2025, this number would increase to 400 million. In the United States, it has been estimated that as many as 11% of the population may be afflicted. Asthma often develops in childhood, although incident adult cases are not uncommon. Estimates indicate that 28% to 78% of young children with asthma ultimately have symptom resolution once adulthood is reached, while 6% to 19% continue with severe forms of the disease. Asthma is the third leading cause of hospitalization among persons less than 18 years of age in the United States, and according to the Centers for Disease Control and Prevention, the prevalence of asthma among U.S. children increased from 3.6% in 1980 to 5.8% in 2003. Not surprisingly, there is a significant cost burden associated with health care for asthma. In 2002, the total direct medical costs relating to asthma were estimated to be $9.4 billion, with indirect costs comprising an additional $4.6 billion. Of these direct costs, 33.0% were due to hospitalization, and 39.4% were related to prescription drug costs.

Diagnosis

Asthma is a complex disease to diagnose with a single parameter. The definitive diagnosis of asthma is a clinical one made on the basis of a patient's medical history, physical examination, and assessment of the reversibility of airway obstruction. In most studies, questionnaires are used to assess whether subjects have had symptoms of asthma or have ever received a diagnosis of asthma from a physician. This type of assessment is highly subjective to a patient's and a physician's understanding and awareness of asthma and has brought speculation as to whether the number of asthma cases has really increased so dramatically or whether there is simply a higher public awareness of the disease coupled with a greater willingness on the part of physicians to diagnose patients as having asthma. One study that illustrates this conundrum well is a study in Scotland that showed that the proportion of children reporting asthma symptoms who also received a diagnosis of asthma increased from 28% in 1964 to 64% in 1999.

Risk Factors for Incidence and Exacerbations

The risk factors associated with asthma appear to involve both genetic and environmental components. Clinical population studies have shown that the risk of allergic disease such as asthma is inherited. In families in which one parent has allergic disease, including asthma, at least 30% of the children will

also be allergic; if both parents have the disease, the risk rises to 50%. In children, risk factors for asthma have been identified as wheezing, familial history of asthma, atopy, obesity, being male, an increase in eosinophils in the peripheral blood, severe infections of the lower respiratory tract, and increased IgE. In adults, the risk factors for asthma include cigarette smoking, rhinitis, atopy, familial history of asthma, and being female. The susceptibility to asthma that may be attributable to genetic predisposition has been estimated to be as high as half of all cases. Chromosome linkage studies have shown that regions on chromosome 6p and 12q are involved in susceptibility to allergy and asthma. However, studies involving monozygotic twins raised in different environments and having differing degrees of asthmatic disease have illustrated the complex interaction of the environment and genetics.

Several environmental factors have been studied for their impact on allergy and asthma sensitization. In urban areas, air pollution, especially diesel exhaust, nitrogen dioxide, and sulfur dioxide have been shown to be major contributors to the development and severity of asthma through the augmentation of previous allergen-specific IgE responses and the capacity to promote the primary sensitization to a new allergen. Exposure to tobacco smoke is highly associated with susceptibility to asthma, and a meta-analysis concluded that parental smoking is very likely to be causally related to both acute respiratory tract illnesses in infancy and the incidence of childhood asthma. There are also studies that have shown that active smoking is associated with the onset of asthma in adolescents and adults.

The environmental impact of living in rural areas has long been known to confer a protective influence against the development of asthma. This is thought to be due to the concurrent exposure of children to bacterial proteins called endotoxins, and potential allergens, while the immune system is still developing. This serves to confer an immunologic tolerance to classic allergens rather than promote an asthmatic response. The ''hygiene hypothesis'' relates to this phenomenon. According to this theory, our largely indoor and hygienic Western lifestyle has decreased our exposure to outdoor environmental allergens and bacterial proteins and, therefore, has increased our likelihood of developing asthma.

Psychological stress has also been shown to play a role in both the incidence and exacerbation of asthma, and there is an association between depressive disorders as well as anxiety disorders and asthma. The mechanism by which stress influences asthma is by altering hormonal components of the endocrine system, such as cortisol and catecholamines, which in turn alters the polarization of the immune response toward an allergic phenotype.

Epidemiologic studies continue to identify and evaluate risk factors for the development or exacerbation of asthma. Episodic exposures to toxicants will continue to be evaluated by documented studies of adverse health effects with comparisons to measured exposures, hospital admissions, or lost school days. Prospective cohort studies or population-based surveys will also continue to evaluate chronic exposures to toxicants and the incidence or prevalence of asthma. Future trends in asthma epidemiology are likely to focus on the combination of specific genetic and environmental factors in the etiology or contribution of the disease and will continue to be challenged by the difficulties of characterizing chronic exposures, multiple contaminant exposures, and study design limitations. It is clear that there are multiple contributing factors to asthma for which epidemiologic studies have been invaluable as far as their contribution to better understanding the syndrome as a whole.

—Janci L. Chunn and Meghan E. Wagner

See also Allergen; Child and Adolescent Health; Chronic Disease Epidemiology; Pollution; Tobacco

Further Readings

Delfino, R. J. (2002). Epidemiologic evidence for asthma and exposure to air toxics: Linkages between occupational, indoor, and community air pollution research. *Environmental Health Perspectives, 110*(Suppl 4), 573–589.

De Sanctis, G. T., Daheshia, M., & Daser, A. (2001). Genetics of airway hyperresponsiveness. *Journal of Allergy and Clinical Immunology, 108*(1), 11–20.

Eder, W., Ege, M. J., & von Mutius, E. (2006). The asthma epidemic. *New England Journal of Medicine, 355*(21), 2226–2235.

Global Initiative for Asthma. (2002). *Global strategy for asthma management and prevention.* N.I.o.H. Publication, Editor. Bethesda, MD: U.S. Department of Health and Human Services.

Marshall, G. D. (2004). Internal and external environmental influences in allergic diseases. *Journal of the American Osteopathic Association, 104*(5 Suppl 5), S1–S6.

National Heart Lung and Blood Institute. (2003). National Asthma Education and Prevention Program (NAEPP).

Expert panel report: Guidelines for the diagnosis and management of asthma. D.o.H.a.H. Services, Editor. Bethesda, MD: National Institutes of Health.

Skrepnek, G. H., & Skrepnek, S. V. (2004). Epidemiology, clinical and economic burden, and natural history of chronic obstructive pulmonary disease and asthma. *American Journal of Managed Care, 10*(5 Suppl.), S129–S138.

Strachan, D. P., & Cook, D. G. (1998). Health effects of passive smoking. 6. Parental smoking and childhood asthma: Longitudinal and case-control studies. *Thorax, 53*(3), 204–212.

Attack Rate

The attack rate is the proportion of people who become ill with a disease. These rates are used in the investigation of an acute outbreak of disease to determine what exposures contributed to the illness. It is calculated as the number of people who became ill divided by the number of people at risk for the illness:

$$\frac{\text{Number of people who became ill}}{\text{Number of people at risk of becoming ill}}.$$

A case definition must first be developed. Case definitions may be based on a constellation of clinical signs (e.g., fever with vomiting and/or diarrhea) or on serology (e.g., the presence of antibodies to the etiologic agent). Those people who meet the case definition are identified; the number of people who meet the case definition is the numerator of the attack rate. The denominator of the attack rate is the number of people at risk of becoming ill. As the attack rate is used in the investigation of an outbreak, those at risk are the people who had the opportunity to be exposed such as those who attended the same event as those who became ill.

The time over which cases are collected is defined by the specifics of the outbreak: those who develop the symptoms over a set period after an event. The period over which new cases occur can be a clue to the identification of the causative agent.

Those who did or did not get sick are interviewed to determine their exposures, and attack rates for those who were or were not exposed are calculated. It may be impossible to determine the causative exposure from exposure-specific attack rates alone, as many of the subjects will have exposure to multiple potential agents. A relative attack rate must be calculated for each exposure:

$$\frac{\text{Attack rate}_{\text{exposed}}}{\text{Attack rate}_{\text{unexposed}}}.$$

An Example of Analysis Using Attack Rates

An outbreak of Group A β-hemolytic streptococcal pharyngitis occurred among prisoners in a jail. Of 690 inmates, 325 were affected. A survey of 314 randomly selected inmates was conducted and a significant association between sore throat and both consumption of egg salad sandwiches and of a beverage was found.

As Table 1 shows, the attack rate among those who ate the egg salad is higher than among those who did not, but the attack rate among those who drank the beverage is also higher than among those who did not drink the beverage. One way to determine which is more likely to have been the vehicle is to examine the relative attack rates. For the beverage, the attack rate for those who drank compared with those who did not is 1.5, while for the egg salad the relative attack rate is 2.1.

A more definitive way to distinguish is to cross-classify the subjects by whether they did or did not eat egg salad and did or did not drink the beverage, as shown in Table 2.

From Table 2, it is clear that whether or not the subjects drank the beverage did not greatly alter the

Table 1 Attack Rates for Pharyngitis in Prisoners Who Consumed a Beverage or Egg Salad

	Ate			*Did Not Eat*			*Relative*
	Sick	*Total*	*Attack Rate (%)*	*Sick*	*Total*	*Attack Rate (%)*	*Attack Rate (%)*
Beverage	179	264	67.8	22	50	44.0	1.5
Egg salad	176	226	77.9	27	73	37.0	2.1

Source: Adapted from Saslaw et al. (1974).

Table 2 Cross-Classified Attack Rates for Pharyngitis in Prisoners Who Consumed Either Egg Salad, a Beverage, or Both

	Ate Egg Salad			*Did Not Eat Egg Salad*		
	Sick	*Total*	*Attack Rate (%)*	*Sick*	*Total*	*Attack Rate (%)*
Drank beverage	152	201	75.6	19	72	26.4
Did not drink beverage	12	15	80.0	7	28	25.0

Source: Adapted from Saslaw et al. (1974).

attack rates for those who ate the egg salad (75.6 vs. 80.0) or who did not eat the egg salad (26.4 vs. 25.0), whereas eating the egg salad did increase the attack rate both for those who drank the beverage (26.4 to 80.0) and for those who did not drink the beverage (25.0 to 80.). This table makes it clear that the egg salad was the most likely vehicle for the infection.

—Sydney Pettygrove

See also Case Definition; Incidence; Outbreak Investigation

Further Readings

Gordis, L. (2000). *Epidemiology* (2nd ed.). Philadelphia: W. B. Saunders.

Saslaw, M. S., Nitzkin, J. L., Levine, R., Hastings, G. E., Cominsky, S., Miller, D., et al. (1974). Outbreak of foodborne streptococcal disease: Florida. *Morbidity and Mortality Weekly Report, 23*, 365–366.

Attributable Fractions

The attributable fraction is one of a class of measures used by epidemiologists to quantify the impact of an exposure or intervention on the occurrence of a disease or other outcome events in a population. These measures can be used, for example, to examine the impact of obesity on disease risk and the potential or expected impact of weight-reduction programs on disease rates. The origin of this concept is a measure proposed by a cancer epidemiologist, Morton Levin, in 1953. In describing the association between cigarette smoking and lung cancer, Levin (1953) offers the following definition of the new measure:

> The third index (S) is the . . . maximum proportion of lung cancer attributable to smoking. This index is based on the assumption that smokers, if they had not become smokers, would have had the same incidence of lung cancer as that found among non-smokers. (p. 536)

Although Levin did not give this measure a name, many epidemiologists now call it the *attributable fraction.* It should be noted, however, that this term has several synonyms in the epidemiologic literature, including attributable risk,* attributable proportion, excess fraction, etiologic fraction,* impact fraction,* and Levin's measure.

This article conceptually defines the attributable fraction and other impact measures, shows how they are estimated from population data, discusses how they are interpreted and sometimes misinterpreted, and demonstrates their use in public health practice. Before discussing these concepts in more detail, it is necessary to provide some background on the theory of causal inference in epidemiology.

Causal Inference in Populations

Measures of Effect and Counterfactuals

Much of epidemiologic research is aimed at making inferences about the net (causal) *effect* of one or more exposures on disease occurrence in a population. The dominant paradigm for defining such effects is the *counterfactual* or potential-outcomes model in which we contrast the frequency of disease (usually incidence) in the population under two or more exposure conditions such as everyone being exposed versus everyone being unexposed. Since individuals cannot be both exposed and unexposed simultaneously, at least one of those conditions is counter to fact or counterfactual. For example, suppose a group of N_e exposed persons at risk are followed without

*These terms also have other meanings as noted later.

loss for a given period during which A new (incident) cases of the disease occur. Thus, the risk (R_e) of disease in this exposed group is A/N_e. The central causal question is how many cases would have occurred during that period if no one in that population had been exposed. Suppose that counterfactual number is A_0. Therefore, the counterfactual risk in the exposed group is A_0/N_e, and the *causal risk ratio* ($\mathcal{RR}_e$), a measure of the exposure effect, is $(A/N_e)/(A_0/N_e) = A/A_0$. If $\mathcal{RR}_e > 1$, the exposure is called a causal or positive *risk factor* for the disease in that population; if $\mathcal{RR}_e < 1$, the exposure is called a protective or negative risk factor. It should be noted that without knowledge of biological mechanisms, this distinction between a causal and protective risk factor is generally arbitrary since exposure statuses can be reversed. For example, inferring that an active lifestyle is a protective risk factor for coronary heart disease is equivalent to inferring that a sedentary lifestyle is a causal risk factor.

Measures of Association and Confounding

Unfortunately, the causal risk ratio cannot be observed or estimated directly because we cannot observe A_0. To deal with this inherent limitation in not being able to observe effects, epidemiologists compare the exposed group with another group of unexposed individuals to observe the statistical *association* between exposure status and disease risk. This approach is the epidemiologic method. Suppose, for example, that a comparison group of N_0 unexposed persons at risk is followed for the same period as the exposed group and B new cases occur. A measure of the exposure-disease association is the ratio of risks in the exposed (R_e) and unexposed (R_0) groups; that is, the risk ratio (RR_e) or relative risk is $R_e/R_0 = (A/N_e)/(B/N_0)$. Unlike the effect measure, this association measure is an observable quantity. Another measure of association is $R_e - R_0$, which is called the risk difference or attributable risk.

For the epidemiologist to infer that $RR_e = \mathcal{RR}_e$, a key assumption is needed: The risk of disease in the unexposed group (B/N_0) must be equal to the counterfactual risk in the exposed group had that group been entirely unexposed (A_0/N_e). The implication of this assumption is that the two groups are comparable or that RR_e is unconfounded. If the assumption is not true, RR_e is confounded (biased) and will not equal $\mathcal{RR}_e$.

Control for Confounders

In nonrandomized studies, it is important for researchers to examine the comparability assumption and, whenever possible, attempt to control for possible confounding. Since the no-confounding assumption is also not observable, epidemiologists address the problem indirectly by attempting to identify and control for *confounders*—that is, extraneous risk factors that are associated with exposure status in the population and that are not intermediate in the causal pathway between exposure and disease. For example, in a study of the association between high blood pressure and stroke, we would expect age to be a confounder because it is a risk factor for both stroke and high blood pressure. There are two statistical methods to control for confounders in the analysis: *stratification*, in which the exposure-disease association is examined within categories or strata of measured confounders, and *covariate adjustment*, in which a regression model, including predictors for the exposure and confounders, is fit to the data.

In stratified analysis, the stratum-specific measures are combined by computing a weighted average. The special stratification method used to estimate effect (a counterfactual contrast) in a particular target population is called *standardization.* In this method, the stratum-specific risks in the exposed and unexposed groups (R_{ej} and R_{0j}, where j refers to the jth stratum) are weighted separately by the numbers of persons in the strata of the target population (N_{sj}), called the standard population (s). That is, the risks in the two exposure groups are mutually standardized to the covariate distribution of the same population. The *standardized risk ratio* ($RR_{e.s}$) is the ratio of the two mutually standardized risks ($R_{e.s}$ and $R_{0.s}$). These operations can be expressed mathematically as

$$RR_{e.s} = \frac{R_{e.s}}{R_{0.s}} = \frac{\sum N_{sj}R_{ej}/\sum N_{sj}}{\sum N_{sj}R_{0j}/\sum N_{sj}} = \frac{\sum N_{sj}R_{ej}}{\sum N_{sj}R_{0j}} = \frac{\sum N_{sj}R_{0j}RR_{ej}}{\sum N_{sj}R_{0j}}, \qquad [1]$$

where $RR_{ej} = R_{ej}/R_{0j}$; and summation is across strata ($j = 1, \ldots, J$). Note that the standardized risk ratio can also be interpreted as a weighted average of the stratum-specific risk ratios (RR_{ej}). If the exposed group is chosen as the standard population (so that $N_{sj} = N_{ej}$) and there is no residual confounding (within strata), then

$$RR_{e.e} = \frac{R_{e.e}}{R_{0.e}} = \frac{R_e}{R_{0.e}} = \frac{\sum A_j}{\sum N_{ej} R_{0j}} = \frac{A}{A_0} = \mathcal{RR}_{\mathrm{e}}, \quad [2]$$

where $R_{e.e}$ is the risk in the exposed group standardized to its own covariate distribution, which is equivalent to the crude (unstandardized) risk in the exposed group $= R_e = A/N_e$. In other words, under these conditions, the risk ratio standardized to the covariate distribution of the exposed group (which is also called the standardized morbidity/mortality ratio, SMR) can be interpreted as the causal risk ratio in the exposed population.

Attributable Fractions for Causal Risk Factors

Impact measures reflect the number of new cases either attributable to the exposure or prevented by the exposure during a given period in a particular population. Thus, there are two basic types of impact measures: *attributable fractions*, which are used for causal risk factors, and *prevented fractions*, which are used for protective risk factors. Since these measures represent another way of expressing counterfactual contrasts, they are closely related to effect measures.

For a causal risk factor, where $\mathcal{RR}_{\mathrm{e}} > 1$, the *attributable number* (A^*) is the excess number of cases in the exposed population that would not have occurred during a given period in the absence of exposure (or exposed at some designated reference level), where $A^* = A - A_0$. It is customary to express this number as the proportion of exposed cases (A) in the population or as a proportion of all exposed plus unexposed cases (M). Thus, the *exposed attributable fraction* (AF_e) is A^*/A, that is, the proportion of exposed cases attributable to the exposure. The *population attributable fraction* (AF_p) is $A^*/M = (A/M)AF_e$, that is, the proportion of all cases attributable to the exposure.

Attributable fractions also have applied interpretations that make them very useful in public health practice. The AF_e is the probability that a randomly selected exposed case would not have developed the disease during the follow-up period in the absence of exposure. The AF_p is the proportion of cases that is potentially preventable, that is, the maximum proportion of cases that could be prevented during a given period if an intervention were implemented to make everyone in the population unexposed. For example, if the AF_p for the impact of cigarette smoking on lung cancer in the United States is 90%, the maximum proportional reduction in the incidence of lung cancer that could be achieved by eliminating smoking in the United States would be 90%.

Estimating the AF_e

Since A^* is a counterfactual quantity, attributable fractions cannot be observed or estimated directly from the conceptual formulas above. By comparing exposed and unexposed groups and assuming no confounding, however, the exposed attributable fraction can be expressed in terms of the observable risk ratio or exposure-specific risks:

$$AF_e = \frac{A^*}{A} = \frac{A - A_0}{A} = \frac{A/A_0 - A_0/A_0}{A/A_0} = \frac{\mathcal{RR}_{\mathrm{e}} - 1}{\mathcal{RR}_{\mathrm{e}}} = \frac{RR_e - 1}{RR_e} = \frac{R_e - R_0}{R_e}. \quad [3]$$

To control for confounders by stratification, the AF_e is standardized to the covariate distribution of the exposed group since the AF_e involves a contrast of the exposed population with that same population in the absence of exposure. Using the same principles as described for standardized risk ratios, the standardized AF_e is

$$AF_{e.e} = \frac{R_{e.e} - R_{0.e}}{R_{e.e}} = \frac{R_e - R_{0.e}}{R_e} = \frac{RR_{e.e} - 1}{RR_{e.e}}, \quad [4]$$

where $R_{e.e}$ is the risk in the exposed group standardized to its own covariate distribution = the crude risk in the exposed group (R_e).

Estimating the AF_p

Assuming no confounding and noting that $A^* = AF_e \times A$, the population attributable fraction can be expressed as a function of AF_e, RR_e, or risks:

$$AF_p = \frac{A^*}{M} = \frac{AF_e \times A}{M} = f_e\left(\frac{RR_e - 1}{RR_e}\right) = \frac{p_e(RR_e - 1)}{p_e(RR_e - 1) + 1} = \frac{R_p - R_0}{R_p}, \quad [5]$$

where $f_e = A/M =$ the proportion exposed among cases; $p_e = N_e/N_p =$ the proportion exposed in the total population (of size N_p); and $R_p = M/N_p =$ the risk in the total population $> R_0$. Note that the AF_p,

unlike the AF_e, is a function of two parameters: the magnitude of effect ($\mathcal{RR}_e = RR_e$, assuming no confounding) and the frequency of exposure in the total population (p_e) or in cases (f_e). In practice, these parameters are often estimated from different sources, for example, RR_e from observational studies of exposure effects and f_e from surveys of the target population. When the exposure is a polytomous variable (more than two categories), two of the expressions in Equation 5 can be extended as follows:

$$AF_p = 1 - \frac{1}{\sum_{i=0}^{I} p_i RR_i} = 1 - \sum_{i=0}^{I} f_i / RR_i, \qquad [6]$$

where i refers to the ith category of the exposure ($i = 0, 1, \ldots, I$); $i = 0$ is the reference category; p_i and f_i are the proportions exposed in the total population and cases, respectively; $RR_i = R_i/R_0 =$ the risk ratio for the ith exposure category; and $RR_0 = 1$. Equation 6 is equivalent to combining all nonreference categories ($i > 0$) into one exposed group and applying one of the expressions in Equation 5.

To control for confounders by stratification, the AF_p is standardized to the covariate distribution of the total target population since the AF_p involves a contrast of the total population with that same population in the absence of exposure. Using the same principles as described for standardized risk ratios, the standardized AF_p can be expressed in terms of standardized risks ($R_{p.p}$ and $R_{0.p}$), the standardized risk ratio ($RR_{e.e}$), or as a weighted average of stratum-specific population attributable fractions (AF_{pj}):

$$AF_{p.p} = \frac{R_{p.p} - R_{0.p}}{R_{p.p}} = \frac{R_p - R_{0.p}}{R_p} = f_e\left(\frac{RR_{e.e} - 1}{RR_{e.e}}\right) = \frac{\sum M_j AF_{pj}}{\sum M_j}, \qquad [7]$$

where j refers to the jth stratum; and R_p is the crude risk in the total target population = the risk standardized to its own covariate distribution ($R_{p.p}$). The target population in Equation 7 is the total population from which cases arose, but it could be another population of interest. With a polytomous exposure, one of the expressions in Equation 7 becomes

$$AF_{p.p} = 1 - \sum_{i=0}^{I} f_i / RR_{i.i}, \qquad [8]$$

where $RR_{i.i} = R_{i.i}/R_{0.i} =$ the risk ratio for the ith exposure group standardized to its own covariate distribution (analogous to $RR_{e.e}$); and $RR_{0.0} = 1$.

AF_p for Multiple Exposures

The methods described above can be extended to analyses in which the researcher's objective is to estimate the combined impact of two or more exposures on disease risk. Suppose, for example, a study objective is to estimate the proportion of oral cancer cases that is attributable to cigarette smoking and/or alcohol consumption in the population, that is, the proportion of cases that would not have occurred if no one in the population had smoked cigarettes or consumed alcoholic beverages. A common mistake is to assume that the population attributable fractions for these two exposures are additive. For example, if the unconfounded AF_p is 35% for smoking and 45% for alcohol, the AF_p for both risk factors is not necessarily 80%, and the AF_p for all remaining risk factors is not $100\% - 80\% = 20\%$. The main reason for these misconceptions is that there may be a biological interaction (e.g., synergy) between the two exposures; that is, the risks attributable to smoking and drinking are not additive, as in this example. In addition, attributable fractions for different risk factors are not additive whenever there are causal effects between the risk factors. In general, the sum of AF_p for multiple risk factors can be greater than 1, but the combined AF_p for any set of risk factors, by definition, cannot be greater than 1.

A correct method for estimating the combined population attributable fraction for K exposures (AF_{pK}) is to cross-classify the population by all K exposures, choose one joint category as the reference group ($i = 0$), and apply the method for polytomous exposures in Equation 6 or 8. Thus, in the previous example, if both smoking and alcohol are categorized into three groups, there would be nine joint categories with nonsmokers and nondrinkers treated as the reference group. This method is equivalent to collapsing all nonreference categories ($i > 0$) into one exposed group and applying Equation 5 or 7.

Prevented Fractions for Protective Risk Factors

For a protective risk factor, where $\mathcal{RR}_e < 1$, the *prevented number* (A^o) is the number of cases that

would have occurred in the exposed population during a given period in the absence of exposure but did not occur—that is, the number of cases prevented by the exposure, where $A^\circ = A_0 - A$. It is customary to express the prevented number as the proportion of potential exposed cases ($A_0 = A + A^\circ$) in the population or as the proportion of all potential cases ($M + A^\circ$). Thus, the *exposed prevented fraction* (PF_e) is $A^\circ/(A + A^\circ)$, that is, the proportion of exposed cases prevented by the exposure. The *population prevented fraction* (PF_p) is $A^\circ/(M + A^\circ)$, that is, the proportion of all cases prevented by the exposure. When the exposure is an intervention such as a vaccine, PF_e is also called the efficacy of that intervention, and PF_p is also called its effectiveness.

Although attributable and prevented fractions are analogous, note the difference in these two concepts: Whereas the denominators of attributable fractions are observable numbers of cases, the denominators of prevented fractions are nonobservable (counterfactual). The implication of this difference, for example, is that the proportion of all coronary heart disease cases prevented by active lifestyles in a particular population during a given period is not equal to the proportion of cases attributable to sedentary lifestyles in the same population.

Estimating the PF_e

By comparing exposed and unexposed groups and assuming no confounding, the PF_e can be expressed in terms of the observable risk ratio or exposure-specific risks:

$$PF_e = \frac{A^\circ}{A + A^\circ} = \frac{A_0 - A}{A_0} = 1 - RR_e = \frac{R_0 - R_e}{R_0}. \quad [9]$$

To control for confounders, the PF_e is standardized to the covariate distribution of the exposed group. Using the same principles as described for attributable fractions, the standardized PF_e is

$$PF_{e.e} = \frac{R_{0.e} - R_{e.e}}{R_{0.e}} = \frac{R_{0.e} - R_e}{R_{0.e}} = 1 - RR_{e.e}. \quad [10]$$

Estimating the PF_p

Assuming no confounding and noting that $A^\circ = A \times PF_e/(1 - PF_e)$, the population prevented fraction can be expressed as a function of observable parameters:

$$PF_p = \frac{A^\circ}{M + A^\circ} = \frac{f_e(1 - RR_e)}{f_e(1 - RR_e) + RR_e} = p_e(1 - RR_e) = \frac{R_0 - R_p}{R_0}, \quad [11]$$

where $R_p < R_0$. When the exposure is a polytomous variable, the expressions in Equation 11 become

$$PF_p = 1 - \sum_{i=0}^{I} p_i RR_i = 1 - \frac{1}{\sum_{i=0}^{I} f_i/RR_i}, \quad [12]$$

where i refers to the ith category of the exposure ($i = 0, 1, \ldots, I$); and $RR_0 = 1$.

Standardizing the population prevented fraction to the covariate distribution of the total population to control for confounders yields

$$PF_{p.p} = \frac{R_{0.p} - R_{p.p}}{R_{0.p}} = \frac{R_{0.p} - R_p}{R_{0.p}} = \frac{f_e(1 - RR_{e.e})}{f_e(1 - RR_{e.e}) + RR_{e.e}}, \quad [13]$$

where $R_p < R_{0.p}$. For a polytomous exposure, the last expression in Equation 13 becomes

$$PF_{p.p} = 1 - \frac{1}{\sum_{i=0}^{I} f_i/RR_{i.i}}. \quad [14]$$

Preventable Fraction

With protective risk factors, researchers and policy analysts are often interested in how much of the current disease risk in the total target population is potentially preventable if everyone in the population were exposed. This measure, called the *preventable fraction* (PaF_p), should be distinguished from the population prevented fraction (PF_p): Whereas the PF_p reflects the previous impact of being exposed in the population, the PaF_p reflects the potential impact in the future if everyone were to become exposed. The PaF_p would be of interest, for example, to predict the maximum proportion of coronary heart disease cases that could be prevented if everyone in the population were to have active lifestyles.

To estimate the PaF_p, it is helpful to recognize this measure as the proportion of cases that was attributable to being *un*exposed. Thus, the PaF_p is equivalent to the population attributable fraction (AF_p) in which "exposed" and "unexposed" categories are reversed; that is, the PaF_p is estimated with one of the expressions in Equation 5 or 7. Suppose, for example, that

the unconfounded risk ratio (RR_e) for coronary heart disease risk, comparing active with sedentary lifestyles, is 0.50, and 20% of the population is active (p_e). Therefore, from Equation 11, the PF_p is $0.20(1-0.50)=10\%$. Reversing the exposure categories, RR_e becomes $1/0.50=2.0$, and p_e becomes $1-0.20=0.80$. Therefore, from Equation 5, the PaF_p is $0.8(2-1)/[0.8(2-1)+1]\approx 0.44$. That is, approximately 44% of the cases that currently occur could potentially be prevented if everyone were to have an active lifestyle.

Covariate Adjustment by Model Fitting

A major limitation with stratification to control for confounders is that the method is inefficient when the sample size is small or when the number of strata is large—that is, when the data are "sparse." In this situation, estimation is imprecise, resulting in wide confidence intervals, and sometimes invalid. To deal with this problem, analysts usually rely on covariate adjustment to control for confounders by fitting regression models to the data. The advantages of model fitting are that continuous variables such as age need not be categorized, and model fitting involves additional parametric assumptions that improve estimation efficiency.

Unfortunately, covariate adjustment for confounders is often done improperly when estimating attributable or prevented fractions. For example, by fitting a logistic model that contains a dichotomous exposure (coded 0 or 1) and potential confounders as predictors, the adjusted odds ratio—an approximation of the risk ratio with a rare disease—is estimated by e^b, where b is the estimated logistic coefficient for the exposure. The incorrect method involves substituting this adjusted risk-ratio estimate for RR_e in one of the expressions in Equation 5 or 11. The main problem is that this method does not properly standardize for the covariates whenever RR_e is heterogeneous across covariate levels; that is, the risk ratios are not weighted to reflect the covariate distribution in the target population.

To standardize for confounders in a cohort study, the fitted logistic model can be used to estimate the disease risk (r_i) for the ith subject, that is,

$$\hat{r}_i = 1 - \frac{1}{1+e^{-(b_0+b_e x_{ei}+\sum_k b_k x_{ki})}}, \qquad [15]$$

where x_{ei} is the value of the dichotomous exposure for the ith subject (0 = unexposed and 1 = exposed); x_{ki} is the value of the kth covariate for the ith subject; and b_0, b_e, and b_k are the estimated logistic coefficients. Then, using the expression in Equation 7, the estimated population attributable fraction, standardized to the covariate distribution of the total population, is

$$\widehat{AF}_{p.p} = \frac{\hat{R}_p - \hat{R}_{0.p}}{\hat{R}_p} = \frac{\sum_{i=1}^{N_p}\hat{r}_i - \sum_{i=1}^{N_p}\hat{r}_{0i}}{\sum_{i=1}^{N_p}\hat{r}_i}, \qquad [16]$$

where $\hat{r}_{0i}$ is the estimated risk for the ith subject, setting exposure status (x_{ei}) equal to the reference value (0). It should be noted that this method may yield invalid results if the model is not properly specified or does not adequately fit the data—for example, if important interactions are ignored. The statistical method for estimating $AF_{p.p}$ from case-control studies differs from Equations 15 and 16 because this design does not include exposure data on the entire population at risk.

Methodological Issues of Interpretation

Risks Versus Rates

It is important to recognize that attributable and prevented fractions, such as risks, are cumulative measures; they have a well-defined period referent during which cases may occur. For example, we might be interested in the 1-year or 10-year AF_p for the impact of an exposure in a fixed cohort initially at risk for the disease and followed for 1 or 10 years. Thus, all observable expressions for the impact measures were expressed in terms of risks. In many epidemiologic studies, however, disease incidence is measured as a *rate*—that is, a person-time measure that has units of time^{-1} (e.g., 5 new cases occurring during 100 person-years of follow-up yield an incidence rate of 0.05 per year). When incidence or mortality rates and rate ratios are substituted for risks and risk ratios in the expressions for attributable fractions, the resulting measures have been called *rate fractions* (RF_e and RF_p). In general, the rate fraction will approximate the corresponding attributable fraction if the disease is rare in both exposure groups and the exposure has only a negligible effect on the changing number of people at risk during follow-up. (Note that the person-time denominator

of a rate changes during follow-up as the number of people at risk changes.)

Choice of the Reference Group

The choice of the reference group for estimating the impact of an exposure is critical to the interpretation of that estimate whenever the exposure is not an inherent dichotomy such as gender. For example, with obesity as the exposure, the proportion of all cases of coronary heart disease that is attributable to a body mass index (BMI) greater than 25 is likely to differ from the proportion of cases attributable to a BMI greater than 30. The choice of a reference group is also critical when comparing the proportions of cases attributable to different exposures. For example, a comparison of population attributable fractions for obesity and smoking in a particular population during a given period depends on how we define the reference categories or levels for these two exposures.

Etiologic Fraction Versus Attributable Fraction

It is common for attributable fractions to be misinterpreted as the proportion of cases *caused by* the exposure—that is, the proportion of cases in whom being exposed contributed to the development of the disease or when it occurs. This latter measure is known as the *etiologic fraction* (EF_e or EF_p). It is important to appreciate that the number of cases caused by the exposure (the etiologic number) is not the same as the number of cases attributable to the exposure (the attributable number). This distinction can be explained in counterfactual terms. The EF_e, for example, can be expressed as $(A_1 + A^*)/A$, where A_1 is the number of exposed cases that would have occurred later during follow-up in the absence of exposure; and A^* is the number of exposed cases that would not have occurred at all during follow-up in the absence of exposure (as previously defined). In contrast, AF_e is A^*/A. Thus, in general, the AF_e is less than or equal to the EF_e. Unfortunately, the EF_e cannot be readily estimated from epidemiologic data without making certain restrictive biological assumptions that cannot be assessed empirically.

The distinction between the etiologic fraction and the attributable fraction has important legal implications that can be used in favor of defendants in toxic tort litigation. In this type of civil case, the plaintiff is claiming injury that was "more likely than not" caused by an exposure for which the defendant is being held responsible. Therefore, using this criterion for legal causation, the court would like to know whether the probability of causation, EF_e, is greater or less than 50%, where e refers to the plaintiff's exposure level, compared with no exposure or some designated background exposure. This probability, however, is likely to be underestimated by the AF_e, as noted above. To the extent that A_1 is greater than zero, EF_e will be greater than AF_e. Suppose that RR_e is greater than 2. According to Equation 3, AF_e will therefore be greater than 50%; thus, it follows logically that EF_e is greater than 50%. On the other hand, if RR_e is less than 2, so that AF_e is less than 50%, it does not follow that EF_e will be less than 50%. Consequently, it is often to the benefit of defense attorneys to underestimate EF_e with AF_e.

Attributable Fraction for Covariates

The concept of the attributable fraction can be extended to "explain" a difference in disease risk or rate between groups in terms of differences in the distribution of one or more covariates (Objective 1) or to "explain" a trend in disease risk or rate over time in terms of changes in the distribution of one or more covariates (Objective 2). For example, an epidemiologist might want to know the extent to which the difference in lung cancer rates between whites and blacks is attributable to differences in cigarette smoking or the extent to which the increasing rate of infant mortality in a low-income population was attributable to a decline in prenatal care. A measure to address these objectives is called the *attributable fraction for covariates* (AF_c).

For Objective 1, the AF_c is the proportion of the difference in risk between groups that is attributable to a difference in the distribution of one or more covariates—that is, the proportion of the crude risk difference that is explained by the covariates. The AF_c can be expressed as a function of risks, risk differences (RD), or risk ratios (RR), as shown below.

$$AF_c = \frac{R_e - R_{e.0}}{R_e - R_0} = \frac{RD_e - RD_{e.0}}{RD_e} = \frac{RR_e - RR_{e.0}}{RR_e - 1}, \quad [17]$$

where e and 0 refer to the two groups (such as exposed and unexposed); RD_e is the crude risk

difference $= R_e - R_0$; and $RD_{e.0}$ is the risk difference standardized to the covariate distribution of the unexposed group $= R_{e.0} - R_0$. If $RD_{e.0} = RD_e > 0$, then $AF_c = 0$, meaning that covariate differences do not explain any of the crude risk difference between groups. If $RD_{e.0} = 0$ and $RD_e > 0$, then $AF_c = 1$, meaning that covariate differences completely explain the crude risk difference between groups.

For Objective 2, the AF_c is the proportion of the change in risk that is attributable to a change in the distribution of one or more covariates—that is, the proportion of the crude risk change that is explained by the covariates. AF_c is computed from one of the expressions in Equation 17, where time is treated as the exposure, letting the first follow-up period be the reference group (0) and the second period be the exposed group (e).

Impact Estimation

As noted earlier, population attributable fractions (and preventable fractions) can be interpreted in terms of the potential for prevention. For example, an AF_p of 80% for a causal risk factor implies that a maximum of 80% of all cases could be prevented if everyone in the population were unexposed. In practice, however, the expected proportional reduction in disease risk due to an actual intervention designed to eliminate the exposure is usually less than 80% because not every exposed (high-risk) person receives the intervention and because the intervention is not 100% efficacious in reducing risk. Impact estimation is a method for estimating the expected impact of a planned intervention on disease risk in a target population by changing the distribution of one or more risk factors. Specifically, the *impact fraction* (*IF*) is the expected proportional reduction in disease risk following the intervention, taking into consideration the limited effectiveness of the intervention. Expressed mathematically,

$$IF = \frac{R_p - R'_p}{R_p} = AF_p \times q_e \times E, \qquad [18]$$

where R_p is the risk in the total population before the intervention; R'_p is the risk in the total population after the intervention; q_e is the success rate of the intervention—that is, the proportion of exposed persons who receive the intervention; and E is the relative efficacy of the intervention—that is, the extent to which participation in the intervention results in a risk reduction to the level of those unexposed persons before the intervention $= (R_e - R'_e)/(R_e - R_0)$. Substituting this expression for E and the expression for AF_p from Equation 5 into Equation 18 yields

$$IF = \frac{p_e q_e RR_e(1 - RR')}{p_e(RR_e - 1) + 1}, \qquad [19]$$

where $RR_e > 1$; and RR' is the risk ratio corresponding to the effect of the intervention among the exposed < 1. Equation 19 assumes that participation in the intervention is unrelated to disease risk. This method has been used to estimate the expected impact of various cholesterol-lowering strategies on the reduction in risk from coronary heart disease. When all four parameters in Equation 19 cannot be reliably determined from existing data, the researcher may estimate *IF* for different sets of reasonable assumptions regarding the values of those quantities.

Impact estimation has recently become popularized as a component of *health impact assessment* (HIA), which is a systematic approach for assessing the expected health impact of a proposed project, program, or policy and communicating the results to decision makers and stakeholders. This approach is often applied to proposals for which health effects are not intended and typically involves identifying relevant health outcomes, defining the causal pathways linking the proposed action with those health outcomes, and developing an analytic strategy. For example, an HIA was conducted to assess the expected impact of a city living wage ordinance on total mortality.

Conclusion

Attributable fractions are important theoretical measures that underscore the counterfactual foundation of the epidemiologic method in populations. These and related impact measures discussed in this article can be applied in a number of informative ways to public health practice and policy making. The estimation of impact measures and their interpretation require explicit specification of a reference population and period referent, attention to the principles of confounder control by standardization, and careful assessment of other sources of bias.

—Hal Morgenstern

See also Admissibility of Scientific Evidence in Law; Causation and Causal Inference; Confounding;

Counterfactual Models; Direct Standardization; Incidence; Measures of Association; Stratified Methods

Further Readings

Benichou, J. (2001). A review of adjusted estimators of attributable risk. *Statistical Methods in Medical Research, 10*, 195–216.

Benichou, J., & Gail, M. H. (1990). Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models. *Biometrics, 46*, 991–1003.

Bruzzi, P., Green, S. B., Byar, D. P., Brinton, L. A., & Schairer, C. (1985). Estimating the population attributable risk for multiple risk factors using case-control data. *American Journal of Epidemiology, 122*, 904–914.

Bulterys, M., Morgenstern, H., Welty, T. K., & Kraus, J. F. (1990). The expected impact of smoking cessation program for pregnant women on infant mortality among Native Americans. *American Journal of Preventive Medicine, 6*, 267–273.

Cole, B. L., Shimkhada, R., Morgenstern, H., Kominski, G., Fielding, J. E., & Sheng, W. (2005). Projected health impact of the Los Angeles City living wage ordinance. *Journal of Epidemiology & Community Health, 59,* 645–650.

Flegal, K. M., Graubard, B. I., Williamson, D. F., & Gail, M. H. (2005). Excess deaths associated with underweight, overweight, and obesity. *Journal of the American Medical Association, 293*, 1861–1867.

Graubard, B. I., & Fears, T. R. (2005). Standard errors for attributable risk for simple and complex sample designs. *Biometrics, 61*, 847–855.

Greenland, S. (1999). Relation of probability of causation to relative risk and doubling dose: A methodologic error that has become a social problem. *American Journal of Public Health, 89*, 1166–1169.

Greenland, S. (2001). Estimation of population attributable fractions from fitted incidence ratios and exposure survey data, with an application to electromagnetic fields and childhood cancer. *Biometrics, 57*, 182–188.

Greenland, S., & Robins, J. M. (1988). Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology, 128*, 1185–1197.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association, 50*, 1168–1194.

Levin, M. L. (1953). The occurrence of lung cancer in man. *Acta Unio Internationalis Contra Cancrum, 9*, 531–541.

Morgenstern, H., & Bursic, E. S. (1982). A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. *Journal of Community Health, 7*, 292–309.

Robins, J. M., & Greenland, S. (1989). Estimability and estimation of excess and etiologic fractions. *Statistics in Medicine, 8*, 845–859.

Rockhill, B., Newman, B., & Weinberg, C. (1998). Use and misuse of population attributable fractions. *American Journal of Public Health, 88*, 15–19.

Siu, A. L., Beers, M. H., & Morgenstern, H. (1992). Choosing quality of care measures based on the expected impact of improved care on health. *Health Services Research, 27*, 619–650.

Thacker, S. B., Ikeda, R. M., Gieseker, K. E., Mendelsohn, A. B., Saydah, S. H., Curry, C. W., et al. (2005). The evidence base for public health: Informing policy at the Centers for Disease Control and Prevention. *American Journal of Preventive Medicine, 28*, 227–233.

Autism

Today, common usage of the term *autism* refers to the following subset of clinical diagnoses falling under the *Diagnostic and Statistics Manual*, fourth edition, revised (American Psychiatric Association) heading of *Pervasive Developmental Disorders*: autistic disorder, Asperger's disorder, and pervasive developmental disorders not otherwise specified (PDD-NOS). However, some still reserve usage of the term *autism* to refer only to the diagnosis of autistic disorder. Because of the potential for confusion, most epidemiologists now prefer to use the term *autism spectrum disorder* (ASD) to refer to the group of three diagnoses collectively and rely on the specific *DSM-IV-R* term when referring to a specific diagnosis. The use of the term *spectrum*, however, should not be taken to imply that this is one etiology with gradations of severity. It is quite possible that the ASDs actually comprise a number of etiologically distinct conditions.

ASDs are neurodevelopmental disorders characterized by deficits in reciprocal social interaction and communication along with the presence of restricted, repetitive, and stereotyped patterns of behaviors, interests, or activities. Individuals with a clinical diagnosis of autistic disorder must have 6 of 12 core symptoms with at least 2 in the social interaction domain, at least 1 in each of the other two domains, and emergence of impairment prior to age 3. A PDD-NOS diagnosis requires some impairment in each of the three core domains where impairment in at least one of the domains does not meet criteria for autistic disorder. An Asperger's disorder diagnosis requires impairment in social interaction and the presence of restricted and repetitive behaviors or interests without communication impairment. In practice, children receiving Asperger's diagnoses typically have no overt language

delays, but, in fact, they often have other deficits in communication. Approximately half of the children with ASDs have cognitive impairment ($IQ \leq 70$). Children with Asperger's disorder tend to have average to above-average IQs, though this is not part of the formal diagnostic criteria. ASDs are approximately four times more common in boys than in girls. Children with ASDs often present with other medical symptoms, including GI dysfunction, disordered sleep, as well as sensory and motor issues. Etiologic or pathophysiologic mechanisms underlying any of the ASDs are yet to be confirmed, and no biological test is available that can inform diagnosis.

In research and specialty clinic settings, two diagnostic tools have become increasingly accepted: the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview—Revised (ADI-R). The ADOS involves direct observation of the child interacting with an assessor who creates multiple, standardized stimuli to elicit behavioral responses from the child. The ADI-R is an in-person interview of the primary caregiver and focuses on the child's behaviors in various types of settings and responses to certain types of social and emotional stimuli. These tools are less reliable in children less than 3 years of age, and new tools are currently being developed and validated to identify signs of ASD in younger children. In most community clinical settings today, diagnosis, though informed by standard instruments, is still typically based on judgment of a clinician following *DSM-IV* diagnostic criteria.

Since clinical diagnoses of ASDs are based strictly on behavioral criteria, the prevalence of ASDs can be influenced by changes in the conventional wisdom surrounding the types and severity of behavior meeting criteria. Over the past two decades, general awareness of ASDs has increased markedly, and the conception of the behavioral deficits meeting diagnostic criteria has broadened. The prevalence of ASDs in the United States has recently been estimated to be approximately 65 per 10,000. This is considerably higher than estimates from past decades, although much of the early prevalence data were limited to estimates of autistic disorder. Undoubtedly, the increasing secular trend in ASD prevalence is due, at least in part, to this increasing awareness and changes in diagnostic tendency. However, it is exceedingly difficult to determine whether these factors account for all the increase in prevalence observed, and, consequently, there continues to be interest in exploration of environmental risk factors that could explain, in part, increasing ASD prevalence over time.

Although specific causes of ASDs are not yet known, it is understood that there is a strong genetic component to their etiology. The prevalence of autistic disorder among children with an older sibling who also has autistic disorder is several-fold greater than population prevalence, and autistic disorder concordance among monozygotic twins has been estimated to exceed 65%, which is far greater than concordance in dizygotic twins. Studies have also suggested that a less-severe broader autism phenotype appears more frequently in family members of individuals diagnosed with an ASD than in control families. ASDs also occur at greater than expected frequencies among individuals with other known genetic conditions such as fragile X, Down syndrome, tuberous sclerosis, Prader-Willi syndrome, and Angelman syndrome. Studies designed to identify autism risk genes have suggested chromosomal regions and candidate genes of potential interest, but as yet, no specific putative risk gene has been found. The lack of conclusive genetic findings most likely is attributable to complexity of the genetic mechanism, likely involving multiple interacting genes, combined with the fact that studies completed to date have not been able to account for etiologic heterogeneity because good markers of etiologically distinct groups are unknown.

Some features of the complex inheritance of autism, for example, the lack of 100% concordance among monozygotic twins and the wide variation in phenotype among concordant monozygotic twins, also point to the possible involvement of nongenetic factors. Neuropathologic studies of autism, although falling short of identifying anomalies in brain structure, organization, or function definitively linked to autism, do indicate that pathologic changes in the brains of individuals with ASDs likely begin prenatally suggesting that the critical window for any etiologically significant exposures may, in fact, be in utero. To date, there has been fairly limited exploration of nonheritable ASD risk factors in epidemiologic studies. Those factors that have been, or are currently being, considered include those related to infection and immune dysfunction, endocrine factors, obstetric factors, xenobiotic exposures, and lifestyle factors. Of course, obtaining valid and reliable measures of exposures and events during prenatal periods is quite challenging

in retrospective studies. At present, there are no nonheritable risk factors that have been conclusively linked to autism. Because of concerns about childhood immunization as a potential risk factor for ASDs, several epidemiologic studies have examined associations between vaccination and autism risk and are consistent in finding no evidence of an association. Studies are just now beginning that will investigate the potential role of environmental risk factors in the context of heritable mechanisms. These include studies of gene-environment interaction and investigations of how nonheritable risk factors may modify a gene's effect via epigenetic mechanisms, such as imprinting.

Although the cause or causes of ASD are unknown, treatment can be effective in ameliorating symptoms. Behaviorally based educational interventions are considered fundamental to an intervention strategy, although optimal approaches are still debated and many believe that different techniques may be more effective in certain, as yet undefined, subgroups of children with ASDs. Evidence does exist, however, indicating that the earlier the initiation of educational interventions, the greater the potential for impact. A variety of pharmacologic treatments are also used to address behavioral symptoms associated with autism, although the evidence supporting their effectiveness is generally limited. At present, only one drug is FDA-approved to treat behaviors (those related to irritability) in children with autism. A number of other intervention approaches are offered to families of children with ASDs in the absence of scientific evidence for their effectiveness and even their safety. These range from strict dietary modification to removal of heavy metals from the body by chelation. Effective provision of accurate information surrounding intervention strategies for ASDs remains a public health priority.

—Craig Newschaffer

See also Child and Adolescent Health; Maternal and Child Health Epidemiology; Psychiatric Epidemiology

Further Readings

Coon, H. (2006). Current perspectives on the genetic analysis of autism. *American Journal of Medical Genetics, Part C Seminars in Medical Genetics, 142C*, 24–32.

Newschaffer, C. J. (2007). The epidemiology of autism spectrum disorders. *Annual Reviews in Public Health, 28*, 21.1–21.24.

Rutter, M. (2006). Autism: Its recognition, early diagnosis, and service implications. *Journal of Developmental and Behavioral Pediatrics, 27*, S54–S58.

Avian Flu

Avian influenza, also known as avian flu, bird flu, or fowl plague, is an animal disease of viral etiology that ranges from a mild or even asymptomatic infection to an acute, rapidly fatal disease of chickens, turkeys, guinea fowls, and other domestic poultry, as well as wild birds, migratory waterfowl, and other avian species. Inasmuch as the avian influenza viruses can be occasionally transmitted to humans, avian flu is a zoonotic disease. Due to its potential to cause worldwide epidemics in humans (i.e., pandemics), and the current A/H5N1 avian flu outbreak—that is, an epizootic or *epornithic* (the nonhuman equivalent of an epidemic in bird populations) in some parts of the world, particularly Southeast Asia—avian influenza has been identified as a major public health concern worldwide. Indeed, under the revised International Health Regulations, any novel (i.e., different from currently circulating human flu H1 and H3 viruses) human influenza A virus infection must be reported immediately to the World Health Organization (WHO). In fact, the chief and foremost strategy in addressing the current pandemic threat entails diminishing the pandemic likelihood by controlling highly pathogenic influenza viruses in animals, expressly the epizootic caused by A/H5N1 virus in poultry, through improved detection, surveillance, and control by way of strengthening veterinary public health structures and competencies.

The natural reservoirs of influenza A viruses—the etiologic agents of avian flu—are the aquatic birds of the world, particularly ducks, in which the viruses appear to be in evolutionary stasis or equilibrium with their natural host, causing no disease. All known influenza A subtypes exist in the aquatic bird reservoir (i.e., the 16 hemagglutinin and 9 neuraminidase surface glycoprotein subtypes); due to this fact, influenza is reckoned as not an eradicable disease. In wild ducks, flu viruses replicate preferentially in the cells linings of the intestinal tract, cause no disease signs, and are excreted in high concentrations in the feces. As much as 30% of the large number of susceptible young ducks hatched each year can shed flu

virus before fall migration for as long as 30 days. Avian flu viruses have been isolated from fecal material and lake water, and it has been shown that viruses retained infectivity in fecal material for as long as 30 days at 4°C. Waterfowl, therefore, may have a very efficient mode of virus transmission: by fecal material in the water supply. Flu viruses of avian origin have been implicated in outbreaks of influenza in mammals, such as seals, whales, pigs, mink, and horses, as well as in domestic poultry. Ducks and wading birds may play a unique and very important role in the natural history of influenza.

As striking as the apparent genetic stability of avian flu viruses in aquatic reservoirs is another conspicuous characteristic: the continued evolution and extent of genetic variation of their mammalian strains. The gene pool of influenza A viruses in aquatic birds provides all the genetic diversity required for the emergence of annual epidemics and occasional pandemics of disease in humans, lower animals, and birds. In humans, pigs, and horses, influenza A viruses show both antigenic drift and genetic shift—that is, point mutations and gene reassortment, respectively—two mechanisms of molecular changes in the two surface glycoproteins and in the eight RNA segments of the viruses that keep accumulating their genetic variability. Another notable characteristic is the lack of proofreading among RNA polymerases, contributing to replication errors of the order of 1 in 10^4 bases (in contrast with the much higher replication fidelity found among DNA polymerases, with errors of the order of 1 in 10^9 bases). Antigenic and genetic evidence show that the 1957 H2N2 Asian and the 1968 H3N2 Hong Kong pandemic strains were generated by genetic reassortment between human and avian flu viruses. Pigs seem to play an important role in interspecies reassortment and subsequent transmission of influenza viruses.

Of the different subtypes of influenza A viruses, two of them H5 and H7 subtypes, are classified as highly pathogenic avian influenza (HPAI) viruses; the reason for this specificity remains unknown. Disease signs vary, depending mainly on the species and age of poultry, virus strain, and superimposed bacterial infections. Typical clinical signs of HPAI in chickens or turkeys are decreased egg production, ruffled feathers, respiratory signs, rales, excessive lacrimation, sinusitis, cyanosis of unfeathered skin, edema of head and face, diarrhea, and nervous disorder. HPAI viruses can cause quick death without clear previous clinical signs.

The Asian H5N1 avian flu virus that infected humans in 1997 acquired all eight gene segments from Eurasian avian sources. This virus was first detected in Guangdong Province, China, in 1996, when it killed some geese, but it received little attention until it spread through live poultry markets in Hong Kong to humans in May 1997, killing 6 of 18 infected persons. The culling of all poultry in that city ended the first wave of H5N1, but the virus continued to circulate among apparently healthy ducks and geese in southeastern China, and the 1997 H5N1 virus was soon replaced by additional genotypes. The most remarkable property of the H5N1 genotype from late 2002 was its high pathogenicity for ducks and other aquatic birds, a property hardly ever found in nature. In early February 2003, H5N1 virus reemerged in a family in Hong Kong and since then has produced 288 confirmed cases and 170 deaths in 12 countries (Azerbaijan, Cambodia, China, Djibouti, Egypt, Indonesia, Iraq, Lao People's Democratic Republic, Nigeria, Thailand, Turkey, and Vietnam), as reported by the WHO as of April 2, 2007. All available evidence on this ongoing situation suggests that human-to-human transmission of H5N1 virus is still highly inefficient and that close contact with sick birds has been a common epidemiologic feature among H5N1 human cases so far.

The unprecedented magnitude of the current avian flu epidemic has resulted in the destruction of hundreds of millions of poultry, mainly chickens, ducks, turkeys, and geese; coupled with export bans on affected countries, the disease is so far representing an enormous impact on the poultry industry and the economies of many countries. In the meantime, multiple opportunities for the successful mammalian transmission of H5N1 influenza viruses—including those taking place in huge live bird markets—are provided by their continuing evolution in Asia, their propensity for reassortment, the generation of multiple lineages and genotypes, and the acquisition of high pathogenicity for aquatic birds. If an opportunity for reassortment with human influenza strains occurs, then the likelihood of successful transmission between humans will be high.

—Oscar J. Mujica

See also Emerging Infections; Influenza; Veterinary Epidemiology; Zoonotic Disease

Further Readings

Horimoto, T., & Kawaoka, Y. (2001). Pandemic threat posed by avian influenza A viruses. *Clinical Microbiology Reviews 14*, 129–149.

Ito, T., & Kawaoka, Y. (1998). Avian influenza. In K. G. Nicholson, R. G. Webster, & A. J. Hay (Eds.), *Textbook of influenza* (pp. 126–136). Oxford, UK: Blackwell Science.

Lipatov, A. S., Govorkova, E. A., Webby, R. J., Ozaki, H., Peiris M., Guan, Y., et al. (2004). Influenza: Emergence and control. *Journal of Virology, 78*, 8951–8959.

Webster, R. G. (1998). Influenza: An emerging disease. *Emerging Infectious Diseases*, *4*, 436–441.

Webster, R. G., & Govorkova, E. A. (2006). H5N1 influenza: Continuing evolution and spread. *New England Journal of Medicine*, *355*, 2174–2177.

B

Bar Chart

A bar chart is a pictorial representation of the information in a frequency distribution. It displays the classes on the horizontal axis and the frequencies or relative frequencies of the classes on the vertical axis. The height of a bar represents how common a particular value or category is in a distribution. A frequency is the number of observations that fall in a class—that is, counts. A relative frequency is the ratio of the frequency of a class to the total number of observations—that is, the proportion or percentage of cases that fall into a class.

One may create either a *frequency* or a *relative frequency* bar chart. A *frequency* bar chart is a graph that displays the classes on the horizontal axis and the frequencies on the vertical axis. The frequency of each class is represented by a vertical bar whose height is equal to the number of times that class occurs in the data set. A *relative frequency* bar chart is a graph that displays the classes on the horizontal axis and the relative frequencies of the classes on the vertical axis. The only necessary difference between the two types of bar charts is the label on the *y*-axis: For a frequency bar chart, it will be counts, while for the relative frequency bar chart, it will be percentages.

It is also possible, although less common, to create a *horizontal bar chart*, which may display either frequency or a relative frequency. A horizontal bar chart displays the classes on the vertical axis and the frequencies or relative frequency on the horizontal axis.

Bar charts are distinguished in two primary ways from histograms:

1. Bar charts are customarily used for discrete or categorical data, while histograms are used only for quantitative data. When histograms are used for continuous data, it usually needs to be grouped into categories.
2. In a bar chart, space is left between the bars, that is, they do not touch each other, while in a histogram, bars may be connected to each other without space.

The second rule emphasizes the discrete or categorical nature of the data presented by the bar chart.

Consider the data given in Table 1, which were collected in a survey from a class of 26 students at Columbus State University.

The classes for grouping the data of sex are "Male" and "Female," while the classes for grouping the data of siblings are numbers from 1 to 5. Tallying the data sets in Tables 2 and 3, we obtain the frequency and relative frequency distributions for sex and siblings, respectively.

A frequency bar chart and relative frequency bar chart for sex and siblings are displayed in Figures 1 and 2.

—Renjin Tu

See also Histogram; Pie Chart; Proportion; Stem-and-Leaf Plot

Table 1 Survey Information Collected From a Class of Students at Columbus State University

Student	*Siblings*	*Sex*
1	1	Female
2	1	Female
3	1	Male
4	1	Male
5	2	Male
6	1	Male
7	1	Male
8	2	Male
9	2	Female
10	2	Male
11	1	Female
12	2	Male
13	2	Female
14	2	Female
15	2	Male
16	2	Male
17	1	Female
18	5	Male
19	4	Male
20	1	Female
21	1	Male
22	2	Male
23	1	Female
24	3	Female
25	5	Male
26	2	Female

Source: These data were collected in a survey by the author in an introductory stat class at Columbus State University.

Table 2 Frequency and Relative Frequency Table for Sex

Sex	*Frequency*	*Relative Frequency*
Male	11	42%
Female	15	58%
Total	26	100%

Table 3 Frequency and Relative Frequency Table for Siblings

Siblings	*Frequency*	*Relative Frequency*
1	11	42%
2	11	42%
3	1	4%
4	1	4%
5	2	8%
Total	26	100%

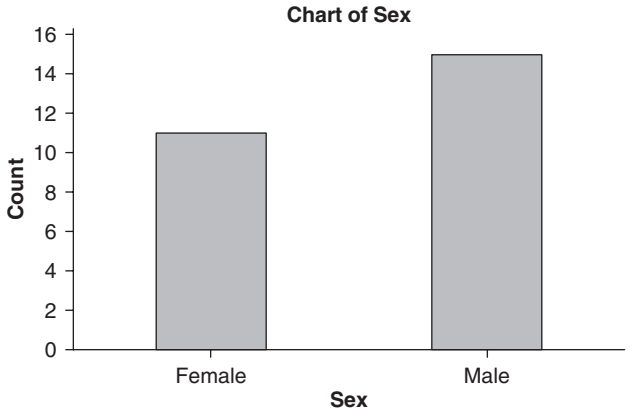


Figure 1 Number of Male and Female Students in a Class of Students at Columbus State University

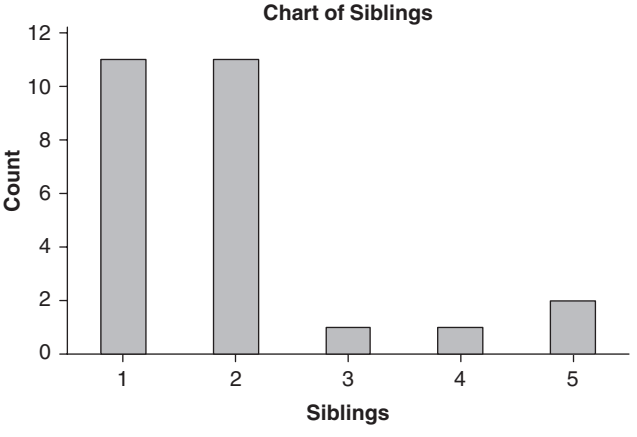


Figure 2 Number of Siblings for Students in a Class at Columbus State University

Further Readings

Ott, R. L., & Longnecker, M. (2001). *An introduction to statistical methods and data analysis* (5th ed.). Pacific Grove, CA: Duxbury.

Weiss, N. A. (2005). *Introductory statistics* (7th ed.). Boston: Addison-Wesley.

Barker Hypothesis

David Barker, of the University of Southampton, England, was among the first to argue that impaired fetal growth is an important contributor to coronary heart disease and its metabolic precursors. In the original formulation of the hypothesis, maternal malnutrition in pregnancy was the underlying determinant of suboptimal birth size, and Barker hypothesized that the effect of prenatal malnutrition was trimester-specific, producing different later metabolic consequences depending on when in pregnancy nutrition was impaired. Small size at age 1 year was also found to be associated with later coronary heart disease, implying a causal role in heart disease of both infant and fetal malnutrition. Barker and his colleagues made ingenious use of several old databases, most notably the records of birthweights and infant weights collected during the 1920s and 1930s in Hertfordshire by a dedicated nurse, which were then linked to contemporary death records. Critics of the Barker opus have asserted that large losses to follow-up; inadequate attention to socioeconomic, environmental, and genetic confounding; and other methodological problems weaken the evidentiary base of the hypothesis.

The hypothesis has broadened beyond its nutritional origins, but it still focuses principally on the relationship of impaired fetal growth to elevated blood pressure, glucose intolerance, and other contributors to coronary heart disease and stroke. The hypothesis has stimulated a great deal of research, both in the laboratory and in human populations. In animal studies, components of the Barker hypothesis have been confirmed; dietary restrictions in pregnancy producing smaller offspring have been accompanied by elevations in cardiovascular risk factors, although at times these nutritional effects have not been dependent on alterations in birthweight. A relatively common model involves halving caloric intake in pregnant animals. How closely these experimental interventions parallel the common human experience is uncertain. This laboratory work has in turn led to the concept of "fetal programming," that is, that some of the cardiovascular risk profile may be determined by subtle alterations in nutrition or metabolism occurring at critical periods in development, especially in fetal life. In human studies, the Barker hypothesis has been studied in many different settings across the globe. It has been broadened by several investigators to include other components of early social and environmental disadvantage, producing what has sometimes been termed the *life course* hypothesis that posits a major role for many kinds of experiences in fetal, infant, and child life in shaping risk of adult disease. Many of these critical experiences are hypothesized to be social and environmental in origin, with biological effects (such as small size) operating as mediators of subsequent cardiovascular risk.

The net contribution of the Barker hypothesis to public health has as yet been modest. Birthweight is notoriously resistant to change in human populations, and the effects of birthweight are modest in determining blood pressure and glucose levels, especially when compared with current weight. But the hypothesis has made many investigators more conscious of the possibility that fetal, infant, and child experiences may have long-lasting consequences for human health.

—Nigel Paneth

See also Cardiovascular Disease; Life Course Approach; Obesity; Reproductive Epidemiology

Further Readings

Barker, D. J. P. (1992). *Fetal and infant origins of adult disease.* London: British Medical Publishing.

Barker, D. J. P. (1998). *Mothers, babies and health in later life.* Edinburgh, UK: Churchill Livingstone.

Joseph, K. S., & Kramer, M. S. (1996). Review of the evidence on fetal and early childhood antecedents of adult chronic disease. *Epidemiologic Reviews*, *18*, 158–174.

Kuh, D., Ben-Shlomo, Y., Lynch, J., Hallqvist, J., & Power, C. (2003). Life course epidemiology. *Journal of Epidemiology and Community Health*, *57*, 778–783.

McMillen, I. C., & Robinson, J. S. (2005). Developmental origins of the metabolic syndrome: Prediction, plasticity, and programming. *Physiological Reviews*, *85*, 571–633.

Tu, Y. K., West, R., Ellison, G. T., & Gilthorpe, M. S. (2005). Why evidence for the fetal origins of adult disease might be a statistical artifact: The "reversal paradox" for

the relation between birth weight and blood pressure in later life. *American Journal of Epidemiology*, *161*, 27–32.

Bayesian Approach to Statistics

The Bayesian approach to statistics is a general paradigm for drawing inferences from observed data. It is distinguished from other approaches by the use of probabilistic statements about fixed but unknown quantities of interest (as opposed to probabilistic statements about mechanistically random processes such as coin flips). At the heart of Bayesian analysis is Bayes's theorem, which describes how knowledge is updated on observing data.

In epidemiology, diagnostic testing provides the most familiar illustration of Bayes's theorem. Say the unobservable variable θ is a subject's true disease status (coded as 0/1 for absence/presence). Let q be the investigator-assigned probability that $\theta = 1$, in advance of diagnostic testing. One interpretation of how this probability statement reflects knowledge is that the investigator perceives the pretest odds $q/(1-q)$ as the basis of a "fair bet" on whether the subject is diseased. If the subject is randomly selected from a population with known disease prevalence, then setting q to be this prevalence is a natural choice. Now say a diagnostic test with known sensitivity *SN* (probability of positive test for a truly diseased subject) and specificity *SP* (probability of negative test for a truly undiseased subject) is applied. Let q^* denote the probability that $\theta = 1$ given the test result. The laws of probability, and Bayes's theorem in particular, dictate that the posttest disease odds, $q^*/(1-q^*)$, equals the product of the pretest odds and the likelihood ratio (*LR*). The *LR* is the ratio of probabilities of the observed test result under the two possibilities for disease status, that is, $LR = SN/(1-SP)$ for a positive test, $LR = (1-SN)/SP$ for a negative test. Thus, postdata knowledge about θ (as described by q^*) is an amalgam of predata knowledge (as described by q) and data (the test result).

More generally, any statistical problem can be cast in such terms, with θ comprising all relevant unobservable quantities (often termed *parameters*). The choice of q in the testing problem generalizes to the choice of a *prior distribution*, that is, a probability distribution over possible values of θ, selected to represent predata knowledge about θ. A *statistical model* describes the distribution of data given the unobservables (e.g., *SN* and *SP* describe the test result given disease status). Bayes's theorem then produces the *posterior distribution*, that is, the distribution of θ given the observed data, according to

$$\frac{\Pr(\theta = a|\text{Data} = d)}{\Pr(\theta = b|\text{Data} = d)} = \frac{\Pr(\text{Data} = d|\theta = a)}{\Pr(\text{Data} = d|\theta = b)} \times \frac{\Pr(\theta = a)}{\Pr(\theta = b)},$$

for any two possible values a and b for θ. Succinctly, a ratio of posterior probabilities is the product of the likelihood ratio and the corresponding ratio of prior probabilities.

The specification of prior distributions can be controversial. Sometimes, it is cited as a strength of the Bayesian approach, in that often predata knowledge is available, and should be factored into the analysis. Sometimes, though, the prior specification is seen as more subjective than is desirable for scientific pursuits. In many circumstances, prior distributions are specified to represent a lack of knowledge; for instance, without information on disease prevalence, one might set $q = 0.5$ in the diagnostic testing scenario above. Or, for a continuous parameter (an exposure prevalence, say), an investigator might assign a uniform prior distribution, to avoid favoring any particular prevalence values in advance of observing the data.

Less controversially, the coherent summarization of postdata knowledge via the posterior distribution is generally perceived as a strength of Bayesian analysis. For instance, the best estimate for a parameter might be "read off" the corresponding posterior density curve, as a mean, median, or mode of the curve. Similarly, the width of the curve describes the accuracy of the estimate. The coherent summarization of uncertainty is particularly important in complex modeling situations, where multiple models and data sources feed into one another. The Bayesian approach avoids the pitfall of overconfident inferences arising from taking an estimate from one model and data source and "plugging it in" to another model as if it were a known value.

Since posterior distributions describe postdata knowledge about unobservables, they also describe postdata evidence about hypotheses concerning unobservables. In contrast to a frequentist hypothesis test leading to a p value, the Bayesian approach leads to the posterior probability that a hypothesis is true. Often this posterior probability for a particular null hypothesis is larger than the corresponding p value.

This practical difference, in tandem with the different foundations and interpretations of frequentist and Bayesian hypotheses testing, continues to generate much debate and discussion.

Computational requirements and lack of software have hampered the transfer of Bayesian technology into epidemiologic practice, although there have been improvements over the past 15 years. New algorithms (Markov chain Monte Carlo algorithms) can implement Bayesian analysis in complex problems with many parameters, and software has been developed to implement these algorithms while shielding the user from algorithmic details. These algorithms represent the posterior distribution via a computer-generated sample (of large, user-specified size) from this distribution. This yields a curious duality: "Nature" is assumed to generate the study data via random sampling according to the statistical model, whereas the investigator generates inferences by random sampling from the posterior distribution. It should be mentioned that there are also non-simulation-based computational strategies, and in some problems, Bayesian inference can be implemented with software for standard (non-Bayesian) techniques, via the addition of "pseudo data" that represent the prior knowledge.

Some of the best applications of Bayesian methods involve prior distributions that arise naturally as plausible scientific assumptions. In disease-mapping applications, for instance, the variation in population rates across regions is plausibly thought to be spatially smooth. Correspondingly, a prior distribution can be constructed such that rates for a pair of adjacent regions are likely more similar than for a pair of distant regions. Another good example arises in meta-analysis. A prior distribution can compromise between the implausible extremes of identical effects across studies or totally unrelated effects across studies. Also, Bayesian methods are well-suited for many epidemiological problems that are characterized by multiple sources of uncertainty and/or extreme magnitudes of uncertainty. In the diagnostic testing situation, for example, the test sensitivity and specificity may themselves not be known exactly, in which case prior distributions could be assigned to these quantities as well.

—Paul Gustafson

See also Hypothesis Testing; Likelihood Ratio; Point Estimate; Sensitivity and Specificity

Further Readings

Berry, D. A., & Stangl, D. (1996). *Bayesian biostatistics.* New York: Marcel Dekker.

Dunson, D. B. (2001). Commentary: Practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology*, *153*, 1222–1226.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis.* Boca Raton, FL: Chapman & Hall.

Greenland, S. (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology*, *35*, 765–775.

Woodworth, G. G. (2004). *Biostatistics: A Bayesian introduction.* New York: Wiley.

Bayes's Theorem

Bayes's theorem, also known as Bayes's rule, is a theorem of probability theory that originated from the works of the Reverend Thomas Bayes (1783). Bayes's theorem connects the conditional and marginal probability of events or probability distributions of random variables. In interpretations of probability, it can be seen as a way of understanding how a probability is updated or revised in light of a new piece of evidence. Bayesian analysis is built on Bayes's theorem and has been used in a wide variety of contexts, ranging from marine biology to the development of "Bayesian" spam blockers for e-mail systems. In the philosophy of science, it has been used to try to clarify the relationship between theory and evidence. The direct use of Bayes's theorem on epidemiology and health science is closely related to diagnostic and screening testing, while Bayesian analysis is gaining extensive application in estimation and statistical inference in epidemiology.

Statement of Bayes's Theorem

Bayes's theorem is a statement on probability. For any two events A and B, the Bayes's theorem is given by

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \qquad [1]$$

Here, $P(A|B)$ is the conditional probability of A given B, and $P(B|A)$ is the conditional probability of B given A. $P(A)$ and $P(B)$ are the marginal probabilities

of events A and B, respectively. To see this theorem expressing how a probability is updated in the presence of new evidence, $P(A)$ is considered as the *prior* probability of A, that is, the probability of A without taking any other information account. Then given the information on B, the updated probability of A, $P(A|B)$ is calculated. $P(A|B)$ is also called *posterior* probability because it takes information of B into account.

There is an alternative formulation of Bayes's theorem. Note that based on the rule of total probability,

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}), \qquad [2]$$

where $P(A)$ and $P(\bar{A})$ are the probabilities of event A occurring and *not* occurring, respectively, and $P(B|\bar{A})$ is the conditional probability of B given A *not* occurring. Then by substituting $P(B)$ with Equation 2 in Equation 1, Equation 1 could be rewritten as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}. \qquad [3]$$

The application of Bayes's theorem in diagnostics and screening tests is based on Equation 3.

The above formulations of Bayes's theorem are for discrete events. For continuous distributions, there is a version of Bayes's theorem based on probability densities:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int_{-\infty}^{+\infty} f(x|\theta)f(\theta)}. \qquad [4]$$

In this formula, $f(\theta)$ is the prior distribution of parameter θ; and $f(\theta|x)$ is the posterior distribution of θ, updated from $f(\theta)$ by incorporating information from x. Parameter θ is often the interest of inference. This formulation for continuous variable is widely used in Bayesian analysis.

Bayes's Theorem and Screening Test

In screening tests, there are several important quantities, including predictive value positive (PV^+), predictive value negative (PV^-), sensitivity, and specificity, that are expressed in terms of probabilities. These quantities are defined as follows:

- PV^+ of a screening test is the probability that a person has a disease given that the test is positive.
- PV^- of a screening test is the probability that a person does *not* have a disease given that the test is negative.
- Sensitivity of a screening test is the probability that the test is positive given the person has a disease.
- Specificity of a screening test is the probability that the test is negative given the person does *not* have a disease.

Let $A =$ disease, $B =$ test positive, $\bar{A} =$ no disease, and $\bar{B} =$ test negative, then predictive value positive $= P(A|B)$, predictive value negative $= P(\bar{A}|\bar{B})$, sensitivity $= P(B|A)$, and specificity $= P(\bar{B}|\bar{A})$. If the proportion of disease in the general population is $P(A)$, the relationship between predictive value positive, predictive value negative, sensitivity, and specificity can be expressed by using Bayes's theorem. With appropriate substitution of terms in Equation 3, PV^+ and PV^- could be given as

$$PV^+ = \frac{\text{sensitivity} \times P(A)}{\text{sensitivity} \times P(A) + (1 - \text{specificity}) \times (1 - P(A))} \qquad [5]$$

and

$$PV^- = \frac{\text{specificity} \times (1 - P(A))}{(1 - \text{sensitivity}) \times P(A) + \text{specificity} \times (1 - P(A))}. \qquad [6]$$

Note that $P(B|\bar{A}) = 1 - P(\bar{B}|\bar{A})$ and $P(\bar{A}) = 1 - P(A)$. In Equation 5, $P(A)$ is the proportion of disease in the general population, and predictive value positive PV^+ could be considered as the updated probability of disease given a positive testing result. Similarly in Equation 6, $1 - P(A)$ is the proportion of *not* having the disease in the general population, and predictive value negative PV^- could be considered as the updated probability of not having the disease given a negative testing result. An example of hypertension will illustrate the general concept of PV^+ and PV^- of a screening test.

Suppose that among people with hypertension, 84% of those are classified as hypertensive by an automated blood pressure machine; whereas among people with normal blood pressure, 23% are classified as hypertensive by the blood pressure machine. In other words, sensitivity $= P(B|A) = 0.84$ and specificity $= P(\bar{B}|\bar{A}) = 1 - 0.23 = 0.77$. Furthermore, suppose some epidemiological study showed that 20% of the general adult population is hypertensive.

Then, PV^+ and PV^- of the blood pressure machine are calculated as

$$PV^+ = \frac{0.84 \times 0.20}{0.84 \times 0.20 + 0.23 \times 0.80} = 0.48$$

and

$$PV^- = \frac{0.77 \times 0.80}{(1 - 0.84) \times 0.20 + 0.77 \times 0.80} = 0.95.$$

Therefore, a negative result from the machine is very predictive and the probability of *not* having hypertension given a negative result is 95%. However, a positive result from the machine is not very predictive with only 48% probability of having hypertension for a positive result.

The above results could also be interpreted in a Bayesian framework. For any person randomly selected from the general adult population, without any other information, the prior probability of having hypertension is 0.20, and equivalently, the prior probability of *not* having hypertension is $1 - 0.20 = 0.80$. With the test result from an automated blood pressure machine, the posterior probability of having hypertension increased to 48% given a positive test and the posterior probability of *not* having hypertension increased to 95% given a negative test. New information from the blood pressure test updates the probability of having hypertension.

As another example, Bayes's theorem is useful in evaluating the result of drug tests. Suppose that a certain drug test will correctly identify a drug user as testing positive 99% of the time, or sensitivity = 99%, and will correctly identify a nonuser as testing negative 99% of the time, or specificity = 99%. This would seem to be a relatively highly accurate test, but Bayes's theorem will reveal a potential flaw. Assume that a corporation decides to test its employees for drug use, and 0.5% of the employees use the drug. It is of interest to know that the probability is that, given a positive drug test, an employee is actually a drug user. It could be shown that

$$PV^+ = \frac{0.99 \times 0.005}{0.99 \times 0.005 + (1 - 0.99) \times 0.995} = 0.3322.$$

Despite the high accuracy of the test, the probability that the employee with a positive test is actually a drug user is only about 33%. The rarer the condition that is being tested, the greater the percentage of positive tests that will be false positives. This illustrates why it is important to do follow-up tests and to consider the issue of false positives when routinely screening for a rare event.

—Rongwei (Rochelle) Fu

See also Bayesian Approach to Statistics; Negative Predictive Value; Positive Predictive Value; Sensitivity and Specificity

Further Readings

Casella, G., & Berger, R. L. (2001). *Statistical inference* (2nd ed.). Belmont, CA: Duxbury Press.

Congdon, P. (2006). *Bayesian statistical modelling* (2nd ed.). West Sussex, UK: Wiley.

Daniel, W. W. (2005). *Biostatistics: A foundation for analysis in the health sciences* (8th ed.). West Sussex, UK: Wiley.

Rosner, B. (2006). *Fundamentals of biostatistics* (6th ed.). Belmont, CA: Thomson Brooks/Cole.

Behavioral Risk Factor Surveillance System

The Behavioral Risk Factor Surveillance System (BRFSS) is a survey conducted annually in the United States by the Centers for Disease Control and Prevention (CDC) in cooperation with state health departments. The BRFSS, whose primary focus is behavioral risks to health, began collecting data in 1984 with 15 states participating; since 1994, all 50 states, the District of Columbia, and the territories of Puerto Rico, Guam, and the U.S. Virgin Islands have been included in the BRFSS.

The BRFSS was the first ongoing national survey to collect data about health and health risk behaviors, and it was designed from the start to provide estimates of these behaviors at both the state and the national levels. Creation of the BRFSS was in large part due to the realization by epidemiologists and public health officials, in the early 1980s, of how much influence individual health and health risk behaviors such as smoking and exercise exerted on morbidity and mortality.

BRFSS data are collected through telephone surveys on a rolling basis throughout the year. The actual data collection is performed either by state health departments or by contractors, using a standardized questionnaire developed by the CDC in cooperation with the state health departments.

Households surveyed are selected through random-digit dialing, and one adult (defined as a person 18 years of age or older) in a selected household is surveyed. Information about health risk behavior of persons younger than 18 years is collected in a separate survey, the Youth Risk Behavior Surveillance System. Persons living in institutions or who do not have a home telephone are automatically excluded from the sample, as are people who are not willing and able to complete a telephone interview in English or Spanish. The states send the collected data to the CDC, who aggregates it, then returns it to the states, and publishes it on the BRFSS Web site.

The BRFSS questionnaire consists of three parts:

1. the *core* component, meaning those questions used by all states and territories; these questions are provided by the CDC and must be administered using exactly the wording and order in the questionnaire;
2. *optional* modules, meaning sets of questions on specific topics such as arthritis or exercise behaviors, which each state may choose to administer or not; these questions are provided by the CDC, and the states are required to use the module exactly as written and in its entirety (although once chosen, the module must be administered exactly as provided in the questionnaire); and
3. *state-added questions*, which are developed or otherwise acquired by an individual state in response to their specific concerns; these questions are not provided by the CDC, and their subject matter, wording, and order are determined entirely by the state using the questions.

The core component is further divided into *fixed* core component questions such as demographics and major risk behaviors such as smoking, which are asked every year; *rotating* core component questions that are included in the core every other year and as optional module questions in the alternate years; and *emerging issues* core questions that typically focus on "late-breaking" health issues. Optional modules are self-contained units that consist of questions focused on a particular topic: For instance, in 2005 there were 26 optional modules whose focus included diabetes, prostate cancer screening, indoor air quality, and intimate partner violence. Up to 10 emerging issues questions may be included in a given year: After that year, they are either discontinued or incorporated into the core or optional modules.

BRFSS data are collected using sampling procedures and weighted so that accurate estimates of risk behaviors may be made at the state and national levels. Some states also stratify their samples to allow them to estimate prevalence for specific metropolitan areas or regions within the state. The different weighting factors are combined into a single variable (*FINALWT* in 2005) that may be used to weight the data to represent the population of noninstitutionalized adults more than 18 years of age. Use of this weighting variable is sufficient to provide accurate point estimates (e.g., means), but complex survey software that can take into account the survey design must be used to account for the sampling procedures and produce accurate estimates of variability (e.g., standard deviations and confidence intervals).

BRFSS data and supporting materials, including the questionnaires used each year, are available for download from the BRFSS Web site. Data from the BRFSS are widely used to evaluate health risks within states and nationally and is extensively used in academic research as well. The BRFSS Web site maintains a searchable bibliography of scientific publications based on BRFSS data, state publications based on BRFSS data (some including the full text of the publication available online), and *Morbidity and Mortality Weekly Report* (MMWR) publications based on BRFSS data (all with full text available online). In addition, a bibliography of methodological papers relating to the BRFSS is available on the BRFSS Web site.

BRFSS data are subject to all the limitations of telephone survey data, including response bias, exclusion of persons without a home telephone, exclusion of people who are unable to conduct a telephone survey and potential for misunderstanding or lack of disclosure by the respondent. Because data are collected by different entities within each state, another potential limitation is that quality of the data may vary from state to state, and information about factors influencing quality may not be available to the individual researcher. However, these limitations are balanced by the fact that the range of data, both in years surveyed and in topics covered, is greater than that available from any similar survey. In addition, the fact that the BRFSS was conducted using scientific sampling procedures makes it possible to use the data to generalize about the prevalence of health risk behaviors at both the state and the national levels.

—Sarah Boslaugh

See also Centers for Disease Control and Prevention; Health Behavior; Public Health Surveillance; Youth Risk Behavior Surveillance System

Further Readings

Link, M. W., Mokdad, A. H., Stackhouse, H. F., & Flowers, N. T. (2005). Race, ethnicity, and linguistic isolation as determinants of participation in public health surveillance surveys [Electronic version]. *Preventing Chronic Disease*, *3*(1), A09.

Mokdad, A. H., Stroup, D. F., & Giles, W. H. (2003). Public health surveillance for behavioral risk factors in a changing environment. Recommendations from the Behavioral Risk Factor Surveillance Team. *MMWR Recommendations and Reports*, *52*(RR-9), 1–12.

Web Sites

Behavioral Risk Factor Surveillance System: http://www.cdc.gov/brfss/index.htm.

U.S. Department of Health and Human Services. Centers for Disease Control and Prevention. Coordinating Center for Health Promotion. Health risks in the United States: Behavioral risk factor surveillance system: http://www.cdc.gov/nccdphp/publications/aag/brfss.htm.

Bias

Error happens. Scales are set incorrectly. Memories fade. Study participants try to give the ''right'' answer. When these mistakes are systematic, that is, not random, they will likely cause a biased result. The term *bias* can be translated fairly as ''wrong'' with the additional refinement of ''wrong due to systematic error.'' Thus, in the general statistical and scientific languages, a biased estimate is an incorrect estimate. In epidemiology, biased estimates typically refer to the distortion of a measure of association between exposure and outcome. This entry describes these measures, including the rate ratio, relative risk, attributable risk, and odds ratio.

The following examples will help illustrate bias. Underestimation would be caused by a scale that always weighs people 10 lb less than their true weight. A survey with leading questions (e.g., ''Do you believe that smoking is bad?'') may draw the desired answers more frequently than the study population actually believes. Study participants with a condition linked to the research may be more likely to follow study procedures or be available for follow-up than participants without the condition.

Causes of biased estimates in epidemiologic studies are generally categorized as *selection* bias and *information* bias. Selection bias refers to ways an estimate may be incorrect due to how participants were enrolled in a study or dropped out of a study (e.g., lost to follow-up). Information bias refers to ways an estimate may be incorrect due to measures made during the study. More than 100 terms for information bias exist, many with similar or overlapping meaning. This entry describes common forms of bias that affect estimates in epidemiologic studies. Confounding can also bias measures of association and is discussed in a separate entry.

Selection Bias

To put selection bias into context, one must understand the logistical challenges of conducting an epidemiologic study. Most of the time, epidemiologists have the capacity (or funding) to study only a relatively small sample of people. However, their goal is to make inferences about a larger population, called the *target population.*

For instance, consider a study to assess the efficacy of a booster vaccine for mumps among people living in the United States. The target population would include individuals who have not had mumps and have had an initial mumps vaccination. Because the target population is so large, it is not feasible to study all of them, and it may also be difficult to obtain a true random sample of the target population.

The *study population* is defined as the group of individuals in the target population who are eligible to be in the study. In statistical parlance, these eligible people have a nonzero probability to be included in the study. For the vaccine research, a convenient study population may be individuals attending high schools in five selected cities. The study population is selected to provide a representative sample of the target population and is largely determined by logistical factors. For example, the investigators must consider the geographic locations of qualified investigators and school districts willing to participate in the study. The study sample may be a random sample of students in these districts who provide informed consent. A more likely (and logistically convenient) scenario would be a random sample of schools where either the whole school or all students in a specific grade are selected.

Multiple steps are usually required to select a study sample from the study population. Each step between the study population and study sample presents an opportunity for bias to be introduced and create a sample that no longer represents the study population, resulting in a biased measure of association. The following scenarios illustrate the complexity of capturing a sample without introducing selection bias.

Medical records may be incomplete for students who transferred in to the school from another district or state. If medical records are used to determine prior vaccination status, the unavailability of the records might make these students ineligible for the study. If these students are less likely to have been vaccinated, excluding them would overestimate the vaccination coverage of the community.

The community may have a large percentage of home-schooled children. These children would not be eligible for the study but may have a different risk of mumps due to lower opportunity for exposure. The study would underestimate the value of booster vaccination if the selection criteria are predicated on enrollment in public school.

Forms of Selection Bias

For selection bias to distort the measure of association, the probability of selection must be disproportionate across different combinations of exposure and disease. Well-designed studies try to minimize the likelihood that selection of study subjects will distort the exposure-disease relationship. Discussion of selection bias is often specific to study designs. The following are some of the more common forms of selection bias that occur. The focus here is on the core study designs used in epidemiology: cohort, case-control and cross-sectional designs, and randomized trials.

Nonresponse Bias

Almost everyone has contributed to nonresponse bias by declining to participate in a telephone survey. Busy people tend to decline to participate in surveys unless they have an inherent interest in the issue.

Epidemiologic studies are challenged by increasingly low response rates. Fewer and fewer people are willing to participate in research studies, perhaps related to the increased proliferation of telemarketing, privacy protection technologies, and regulations. Most national surveys conducted by randomly calling households (random-digit dialing or RDD) have response rates substantially less than 50% of the eligible study sample selected. Thus, there is a significant likelihood that the participants will differ markedly from the nonparticipants. This, in turn, increases the likelihood that the study sample will not adequately reflect the study population or target population.

Because almost every study population will have nonrespondents, it is important to characterize the degree and direction of bias likely to be introduced. Nonresponse by some eligible subjects, however, does not necessarily mean that the measure of association will be biased. It just means that the opportunity exists for bias to influence the outcome. If the distribution of disease (for cohort studies) or the distribution of exposure (for case-control studies) among participants (volunteers) is different from that of nonparticipants (nonresponders), a biased measure of association could result. Researchers interested in the impact of nonresponse on national health surveys conducted by RDD, such as the Centers for Disease Control and Prevention's (CDC's) Behavioral Risk Factor Surveillance System (BRFSS), have not identified substantial nonresponse bias for similar health outcomes when compared with much more expensive in-person interview studies with higher response proportions.

In case-control studies, nonresponse bias has been identified as an important issue. With this study design, the control group is used to estimate the distribution of exposure in the study population. Those who are cases (and thus have a particular interest in the research) are often more likely to respond than individuals who are randomly selected for the control group. Volunteers who do agree to participate are often healthier than the general population, so the control group may not be representative of the population from which the cases arose. If the exposure of interest is associated with healthy behaviors, for example, then the measure of association (e.g., odds ratio) will likely be biased. The following is an example:

> A case-control study is designed to examine the relationship between smoking and type 2 diabetes. Among cases (new cases of diabetes), smokers and nonsmokers are equally likely to agree to participate. Among controls, however, smokers are less likely to participate. A history of smoking among controls

would therefore be too low, and the effect of smoking on diabetes would be overestimated.

Selective Survival Bias (Neyman Bias)

Selective survival bias occurs in case-control studies when cases are selected from those who have disease, not exclusively from those who are newly diagnosed. The distinction here is between prevalent disease (those with an existing condition) and incident disease (those with a new diagnosis of the condition). Prevalent disease is influenced by factors that cause the disease as well as factors that influence cure or survival. If the study design allows individuals with established disease (prevalent cases) into the study, then the researcher may not be able to distinguish between effects related to cure or survival versus effects related to the likelihood of developing disease. To limit the risk of selective survival bias, most epidemiologists now limit study eligibility to incident cases when possible. For diseases with insidious onset, such as Alzheimer's disease, this may be difficult. The following is an example:

> Researchers enroll clients of a local psychology practice who are being treated for depression as cases in a case-control study designed to examine the effect of diet on the development of depression. If a well-balanced diet decreases the amount of time a client needs treatment for depression, then those with a well-balanced diet will leave the psychology practice sooner than those with poor diets. Using the diets of clients needing ongoing treatment will overestimate the effect of a poor diet because the effect of a poor diet on the development of depression is inadvertently linked to the effect of a poor diet on the need for ongoing treatment.

Berkson's Bias

This bias occurs in case-control studies when the probability of hospitalization for cases differs by whether or not they were exposed. If exposed cases are more likely to be hospitalized than nonexposed cases, then using only hospitalized cases in a case-control study will overestimate the prevalence of exposure among cases. The following is a well-known example of Berkson's bias:

> In hospital-based case-control studies of oral contraceptives and thromboembolism, early research found a strong association. However, because the medical community was aware of the risk of thromboembolism due to oral contraceptives, physicians were more likely to hospitalize women who had symptoms if they were taking oral contraceptives. The differential probability of hospitalization depending on oral contraceptive use resulted in a biased (overestimated) odds ratio.

Inclusion Bias

Inclusion bias occurs when controls selected for a case-control study are more likely to be exposed than the study population from which they arose. The most common examples are from case-control studies that use hospitalized patients as controls. If the exposure is associated with increased risk of diseases other than the one being studied, then the hospital-based controls may have higher history of exposure than the general population. Since this will make the exposure of the controls similar to the exposure of the cases, inclusion bias typically results in a bias toward the null. When hospital-based controls are used, researchers can try to avoid this bias by excluding controls with conditions known to be associated with the exposure of interest.

For example, consider a case-control study of alcohol use and liver cancer. Hospital-based controls may be more likely to drink alcohol than nonhospitalized, healthy people, because alcohol use is associated with many diseases resulting in hospitalization (e.g., cardiovascular disease, injuries).

Exclusion Bias

Case-Control Studies

Exclusion bias is a term used to describe two different forms of bias. For case-control studies, exclusion bias occurs when different exclusion criteria are applied to cases and controls. For example, cases may be eligible even if they have other health conditions (comorbidities), but controls are excluded unless they are healthy. As an example, consider the case-control study of alcohol use and liver cancer described above.

Researchers want to select only hospitalized controls with diseases unrelated to alcohol, so they exclude patients with a history of heart disease, even if the patients are being admitted for other conditions. Some

of the cases, however, will also have a history of heart disease, so the excluded controls do not represent the same underlying population as the cases. To avoid this, the exclusion criteria should have been related to the reason for admission to the hospital rather than prior history of disease.

Randomized Trials

Exclusion bias also describes biased results from randomized trials. The bias arises when intent-to-treat analysis is not used. Intent-to-treat analysis involves analyzing the data based on the original random assignment. Study participants are included in the denominator based on their initial assignment into treatment groups; participants are not pulled out of the denominator due to early departure from the study, noncompliance with the assigned treatment, or loss to follow-up, among other issues. Because early departure and noncompliance may be related to the toxic effects of the treatment under study, lack of good data for these individuals should be considered in the analysis to best describe the effects of the intervention or treatment. For example, consider the following:

> A rare side effect for a new drug is severe headache. In a clinical trial, patients receiving the drug and experiencing the headache are likely to stop taking the drug. Patients receiving the placebo do not experience the headache and do not discontinue participation in the study. Excluding patients who stop taking the drug will make the treatment group less comparable with the placebo group.

Healthy Worker Effect

The healthy worker effect is a bias that occurs in cohort studies and arises from the fact that those who work tend to be healthier than those who do not. Thus, when comparing the mortality of workers in a high-risk industry (e.g., potentially toxic chemicals) with the mortality experience of the general population, the harmful effects of the exposures may be underestimated.

For example, consider a cohort study of exposure to toxic fumes on lung function that compares the lung function among firefighters with that among office workers. Firefighters must be extremely physically fit to perform their jobs; office workers do not. Firefighters who have poor lung function stop being firefighters and get office jobs. The firefighters who remain at their jobs will have better lung function than the office workers, and the effect of the toxic fumes will be underestimated.

Matching

Matching cases and controls in a case-control study can introduce selection bias. Fortunately, problems with the estimation can be corrected by statistical analysis and are thus rarely an issue if the proper techniques are used for adjustment.

Overmatching refers to inappropriate use of matching when it is either not necessary or causes bias. The purpose of matching is to improve statistical efficiency in the control of confounding by assuring that each strata of the confounder has enough controls. Matching on a factor that is not a confounder will introduce confounding by that factor and decrease statistical efficiency, but the bias due to confounding can be accounted for in the analysis. A greater problem with overmatching occurs when cases and controls are matched on factors that are affected by the exposure and disease. If cases and controls are matched on intermediaries between the exposure and disease, the resulting bias cannot be accounted for in the analysis. The following is an example:

> In a study of occupational exposure to radiation and the risk of leukemia, researchers matched cases and controls based on when they started working and length of employment at the nuclear facility. The purpose of matching on these employment-related factors was to control for the effects of differences in safety practice over time. Length of employment is also correlated with radiation dose, however, since longer work history means a higher cumulative dose of radiation. Cases and their matched controls, therefore, had the same radiation doses, and matching on employment date obscured the effect of dose on leukemia risk.

Loss to Follow-Up

Loss to follow-up can cause bias in both clinical trials and cohort studies. If it seems counterintuitive to consider loss to follow-up as a type of selection bias, think of it as differential selection *out* of the study. Typically, participants who leave a study differ from those who remain enrolled in important ways. For

clinical trials, individuals who are not getting "better" on the treatment may disproportionately decide to leave the study. This difference in attrition can bias the study results, as in the following example:

> In a cohort study of breastfeeding and risk of infections in the infant, researchers enroll women giving birth at a local hospital, some who are breastfeeding and some who are not. They count the number of infections the infants have in their first year. Women with less stable economic situations may be less likely to breastfeed and also harder to track if their living situation is not stable. Other factors related to infection risk, such as day care attendance, may also differ.

Missing Information

This type of selection bias occurs at the analysis stage. There are many potential reasons for missing information, such as participants refuse to answer some survey questions or medical records cannot be found. The most common method for dealing with missing data in epidemiology studies is to exclude the individuals with missing information based on key factors. Thus, they are selected out of the study in the analysis phase. Another rarely used technique to deal with missing information is estimation of the values using statistical imputation. The following is an example of missing information bias:

> A study was conducted to assess the racial disparities that exist for hypertension during pregnancy. All hospital records for delivering a baby during a year period in a state were identified in a computerized database. If information on race was missing, the record was excluded. An association between race (black, white) and hypertension during pregnancy was identified. Black women were more likely to be hypertensive during pregnancy than white women in the state. It was then noted that some hospitals did not record race on computerized records so all women who delivered their baby at one of these hospitals were missing. If women served by these hospitals were different on both race and hypertension, then the measure of association computed would be biased due to the deletion of missing data.

Estimating the Effect of Selection Bias

It is rarely possible to estimate the effect of selection bias mathematically. To do so requires the following information: an accurate study participation rate, and knowing the exposure and outcome status for both participants and nonparticipants. When this information is available, the percentages can be used to estimate the amount and direction of bias. Also, the biased measure can be adjusted to provide a more accurate estimate of the true measure of association.

For most studies, the data described above are not available to compute the mathematical adjustments for selection bias. Nevertheless, it is important to contemplate the types of selection bias that can occur with different study designs. The researcher can then choose the study design that best minimizes all forms of bias, including selection bias. With careful consideration, one can often identify the *direction* that selection bias will pull the measure of association, but the *magnitude* of the bias is more difficult to conceptualize. Experience in the topic area and study population improves the ability to predict both the magnitude and the direction of bias.

Understanding selection bias is critical to strategic study design. Experienced epidemiologists usually design studies to keep the magnitude of bias as small as possible and point the direction of bias *toward* the null. This conservative approach means that the measure of association can be interpreted as an underestimate of the true strength of association. If the direction of the bias is *away* from the null, then computing the actual magnitude of the bias becomes essential to the determination of whether an association exists at all.

Reducing the Effect of Selection Bias

Case-control studies are particularly vulnerable to selection bias. Cases are often drawn from hospitals or clinical practices that do not have well-defined populations from which to select controls. Even people hospitalized in the same facility may not be from the same source population because the catchment areas for hospitals often vary by diagnostic group.

Carefully delineating the source population from which the cases arose is imperative to develop a good framework for selecting controls to minimize selection bias. Integrating the suggestions below in the study design phase may help:

- For case-control studies, select controls that best represent the source population for the cases. Controls should come from a population that is similar in every way to the cases, except for the outcome of

interest. In other words, if the control had gotten the disease, they would have been a case in the study.
- Reduce loss to follow-up with regular outreach to study participants. In clinical trials, you can attempt to measure the impact of unavoidable loss to follow-up with a small pilot study among those who dropped out of the study.
- Anticipate which factors may be intermediate between the exposure and disease and avoid matching cases and controls according to these factors. This will reduce the likelihood of overmatching.
- Consider using statistical imputation to minimize the bias introduced by missing information in the analysis stage.

Information Bias

Information bias is caused by errors in the measurement of study factors among participants. Misclassification occurs when study subjects are assigned to the wrong disease or exposure group. In a cohort study or clinical trial, information bias primarily results from misclassification of disease outcome (e.g., disease was not diagnosed or the subject was mistakenly determined to have disease), but misclassification of exposure is also possible. In a case-control study, information bias primarily results from misclassification of exposure, but misclassification of disease status is also possible. Misclassification is directly related to sensitivity and specificity of the measurements. If the sensitivity and specificity are both 100%, then there is perfect measurement and no misclassification will occur. However, if either sensitivity or specificity is lower, misclassification will likely occur. It is inevitable that misclassification will happen in epidemiology research. Even when measures are validated, there will never be perfect measurement.

The type of bias that results from misclassification depends on how the misclassification happened. When mistakes occur on the classification of responses to one factor independent of the other factors, it is called *nondifferential misclassification.* More specifically, nondifferential misclassification of disease occurs when the errors are not linked to exposure status. For instance, if disease status is determined by laboratory tests that have 98% sensitivity and 90% specificity, then all study participants have the same chance of being incorrectly classified. Similarly, nondifferential misclassification of exposure means that mistakes made by categorizing participants by exposure status are not related to their disease status. Nondifferential misclassification usually results in a bias toward the null; that is, the measure of association is underestimated. In very rare situations, the bias results in an overestimate.

Differential misclassification occurs when the error in classification is somehow linked to both the subject's exposure status and the disease status. Differential misclassification can result in an over- or underestimate of the measure of association, so careful understanding of the type of bias and potential effect of the bias is needed to interpret the study findings.

Interviewer Bias

When the interviewer knows the study hypothesis or goals of the study, a conscious or unconscious attempt may be made to elicit responses consistent with the desired information. This potential bias particularly exists if the interviewer knows the study hypothesis and specific exposure, intervention, or disease status of the study participants. Interviewer bias is generated by using intonation (emphasis) or gestures when questions are asked or using prompts to elicit certain responses differentially depending on the subject's status. It is important to note that even if the study design blinded interviewers to participants' exposures and diseases, it is possible that the participants will spontaneously share this information with the interviewer. The following is an example:

> In a case-control study of diet and heart disease, the interviewer asks participants about dietary habits, exercise, and lifestyle choices. The interviewer knows which participants have heart disease and which do not. When the survey turns to questions about diet, the interviewer inadvertently smiles and rushes through the questions about snack foods for controls who are obviously lean and healthy, assuming their diet would not contain snack foods.

Recall Bias

People's ability to remember accurately depends on a number of factors, including the significance of the exposure, the degree of ruminating about the exposure, prompts (e.g., calendars, pictures) to improve memory, and motivation. Recall bias occurs

when groups are different in their ability to recall information. Investigators designing case-control studies must be particularly cognizant of the potential for recall bias to be an issue.

Consider, for example, a case-control study of maternal health behaviors during conception and early first trimester and the risk of birth defects. The recall may be very different for mothers of newborns with a birth defect compared with mothers of healthy newborns. Mothers of newborns with a congenital health condition may ruminate on possible exposures to isolate a cause. In contrast, mothers of healthy newborns are much more focused on adjusting to the current needs of the baby and pondering the future of their new child. Case mothers are likely to be more accurate than control mothers, and control infants would be more likely to be misclassified as unexposed. If the behavior is harmful, the differential misclassification of exposures based on case-control status would result in an overestimate of the odds ratio.

Telescoping

Telescoping occurs when individuals are asked to report exposures or diseases that occurred within a specified time period (e.g., during the past year). It can also occur when participants are asked when symptoms began. In case-control studies, controls are often asked about their exposures at the time the case was diagnosed. Unlike the cases, controls have no inherent meaning assigned to this particular day or month in the past and therefore have difficulty pinpointing it in relation to their own life events. Telescoping typically occurs because the person has difficulty bounding the time period of the inquiry, as in the following example:

> In a cross-sectional study of eye infections among contact lens wearers, the participants are asked about their lens hygiene and previous eye conditions. The participants are asked to report on these factors for the previous 6 months. Some participants have difficulty remembering and report eye infections that occurred more than 6 months previously.

Reporting Bias

Reporting bias occurs when the participant, consciously or unconsciously, provides incorrect information. This can happen when the participant tries to please the investigators by supplying the presumed "correct" answer. If the respondents are aware of the hypothesis being studied, reporting bias is likely to cause differential misclassification resulting in an overestimate or underestimate of the measure of association. In a case-control study, this misclassification occurs because cases will be more likely to report being exposed while the converse will be true for the controls. Another form of reporting bias may be due to underreporting of exposures or experiences. Underreporting is a particular issue when collecting sensitive information (e.g., illegal drug use, high-risk sexual activities). Underreporting may be differential or nondifferential, depending on the study being conducted.

In a cross-sectional study of occupation and sexually transmitted disease (STD), participants in high-status occupations (e.g., executives, physicians) are less likely to honestly report the occurrence of an STD.

Detection Bias

Detection bias occurs primarily in cohort studies where the ability to detect disease is different for the exposed compared with the unexposed. If those with an exposure are more likely to be seen regularly by a clinician, they will also be more likely to be diagnosed. Diseases with an indolent progression are especially vulnerable to this issue. The following is an example of detection bias:

> In studies of high-risk materials (e.g., chemical, radiation), the disease experience of cohorts working with the material is compared with the disease experience of the general population with similar characteristics (e.g., males). These studies are often performed using retrospective cohort designs and the outcome status (e.g., thyroid cancer) is based on medical record reviews. Thyroid cancer is rarely fatal and is often an incidental finding on medical examination. In a study of radiation and thyroid cancer, it is possible that those who work in occupations with exposure to radiation have more frequent and extensive medical examinations and thus are more likely to have a thyroid cancer diagnosis than the general population.

Hawthorne Effect

The Hawthorne effect describes the change in normal behavior typical of individuals who know

they are being observed as part of the research process, as in the following example:

> A field trial is designed to examine the effect of a new hand-washing campaign among hospital workers. The new campaign is introduced among employees of five hospitals. Five other hospitals are used as a comparison (usual hand-washing behavior) group. At the comparison (usual hand washing) hospitals, the staff knows that hand washing is being monitored, and the amount of hand-washing behavior increases and improves. The amount of hand washing at the comparison hospitals does not reflect the norm and the effect of the new campaign is underestimated.

Panel Effect

The panel effect is caused by repeated interviews about exposures and diseases during the study period. Participants tend to report fewer events over time, thus misclassification occurs as a result of lowering sensitivity. Panel effect may produce differential or nondifferential misclassification depending on the study design. An example is the following:

> In a cohort study of the relationship between diet and birthweight, pregnant women are asked to complete a dietary recall at each prenatal visit. As time goes on, participants lose enthusiasm and the completeness and accuracy of the diet recall decreases. The effect causes nutrient intake to appear lower as pregnancy progresses.

Estimating the Effect of Information Bias

Fortunately, it is often possible to estimate the degree of information bias in epidemiologic studies. For example, when nondifferential misclassification occurs due to imperfect sensitivity and specificity, measures of association can be adjusted for the misclassification. When sensitivity and specificity are known, it is possible to estimate the magnitude as well as direction of misclassification bias.

Using a measure with high sensitivity and specificity is ideal, but sometimes not practical for research purposes. To evaluate the potential role of information bias in the study, investigators might design a pilot study to measure the sensitivity and specificity of the questionnaire or measuring instrument. For types of information bias that cannot be so directly evaluated, and thus not amenable to mathematical adjustment, the investigator must carefully consider the magnitude and direction of the information bias. Examples of such deliberation may be found in the discussion sections of published research studies.

Reducing the Effect of Information Bias

Epidemiologists use multiple methods to reduce the opportunity for information bias to occur and reduce the magnitude of the bias. Several suggestions are provided here:

- Interviewers should be blinded to the research question and hypothesis and, if possible, to the exposure or disease status. Train interviewers to conduct structured interviews, test interviewers to establish interrater reliability before the study begins, and provide periodic refresher sessions. Anticipate participant questions and provide structured responses for interviewers.
- Select questionnaires or measurement instruments with established validity and reliability.
- Recall bias is reduced with the use of calendars or temporal landmarks to improve the memory. Pictures and other memory aids may help improve the recall of medications or other exposures. Furthermore, a checklist may elicit a more complete recall of medications taken than would open-ended questions.
- For cohort studies, use similar methods of detecting diseases among exposed and unexposed populations.

—Louise-Anne McNutt, Allison Krug, and Colleen McLaughlin

See also Confounding; Hawthorne Effect; Healthy Worker Effect; Study Design

Further Readings

Delgado-Rodriguez, M., & Llorca, J. (2004). Bias. *Journal of Epidemiology and Community Health*, *58*, 635–641.

Hennekens, D. H., & Buring, J. E. (1987). *Epidemiology in medicine.* Boston: Little, Brown.

Rekers, H., Norpoth, T., & Michaels, M. A. (1996). Oral contraceptive use and venous thromboembolism: A consideration of the impact of bias and confounding factors on epidemiological studies. *European Journal of Contraception & Reproductive Health Care*, *1*, 21–30.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott, Williams & Wilkins.

Schoenbach, V. J. (2000). *Understanding the fundamentals of epidemiology: An evolving text.* Chapel Hill: University of North Carolina. Retrieved March 19, 2007, from http://www.epidemiolog.net.

Binomial Variable

Binomial variables are frequently encountered in epidemiological data, and the binomial distribution is used to model the prevalence rate and cumulative incidence rate. Binomial variables are created through repeated trials of a Bernoulli process, often called Bernoulli trials. Daniel Bernoulli (1700–1782) was the first mathematician to work out the mathematics for Bernoulli trials. *Bernoulli trials* must satisfy the following three conditions:

1. The experiment has two possible outcomes, labeled *success* (S) and *failure* (F).
2. The trials are independent.
3. The probability of a *success* remains the same from trial to trial; this is called the *success probability* and is denoted with the letter p.

The word *success* as used here is arbitrary and does not necessarily represent something good. Either of the two possible categories may be called the success S as long as the corresponding probability is identified as p.

A random variable for the number of successes in a sequence of Bernoulli trials is called a *binomial variable.* The probability distribution for a binomial variable is called the *binomial distribution.* The binomial probability formula for the number of successes, X, is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

where the *binomial coefficient* $\binom{n}{k}$ is defined as

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

and $k!$ is the product of the first k positive integers and is called k factorial. In symbols,

$$k! = k \cdot (k-1) \cdot \cdots 2 \cdot 1.$$

Consider the following example. Epidemiological surveys have determined that 9% of men and 0.25% of women cannot distinguish between the colors red and green. This is the type of color blindness that causes problems reading traffic signals. If six men are randomly selected for a study of traffic signal perceptions, the probability that exactly two of them cannot distinguish between red and green can be calculated using the binomial distribution formula as follows.

We must carefully define which outcome we wish to call a success. For convenience, we define a *success* as a randomly selected man who cannot distinguish between the colors red and green, so $p = 0.09$. Let X denote the number of men of the six who cannot distinguish between the colors red and green. The number of trials is the number of men in the study, so that $n = 6$. Using the binomial probability formula for $k = 2$ yields

$$\begin{aligned} P(X = 2) &= \binom{6}{2}(0.09)^2(1-0.09)^{6-2} \\ &= \frac{6!}{2!(6-2)!}(0.09)^2(0.91)^4 \\ &= (15)(0.09)^2(0.91)^4 = 0.0833. \end{aligned}$$

Therefore, the probability that exactly two of the six men cannot distinguish between red and green is 8.33%.

—Renjin Tu

See also Incidence; Prevalence; Random Variable

Further Readings

De Veaux, R. D., Velleman, P. F., & Bock, D. E. (2006). *Intro stats* (2nd ed.). Boston: Addison-Wesley.
Triola, M. F. (2001). *Elementary statistics* (8th ed.). Boston: Addison-Wesley.

Biomarkers

A biomarker is broadly defined as a substance that can be measured and evaluated as an indicator of normal or pathogenic biologic processes or of a biologic response to a therapy or intervention. In epidemiology, biomarkers are often used to measure internal dose, biologically effective dose, early

biologic response, altered structure or function, and susceptibility. By incorporating biomarkers into epidemiologic assessments, researchers may more precisely measure exposures or outcomes, reduce exposure and/or disease misclassification to produce less biased estimates of association, and elucidate biologic processes underlying exposure-disease relationships. The increasing integration of biomarkers in epidemiologic studies has fostered the creation of a distinct multidisciplinary subspecialty called molecular epidemiology in which molecular, cellular, and other biologic measurements are incorporated into cross-sectional, retrospective, and prospective observational studies and clinical trials.

Types of Biomarkers

Biomarkers may take the form of exogenous compounds (e.g., absorbed chemicals, pesticides, food derivatives, metals), whose presence can be detected and quantified in biologic media, as well as measures of endogenous biologic substances, such as nucleic acids, proteins, and lipids. Biomarkers may be measured from any number of biologic media, including blood, urine, hair, feces, sputum, nails, and other body fluids and tissues. Table 1 lists the types of different biomarkers that have been used in epidemiologic assessments.

Uses of Biomarkers in Epidemiologic Studies

For purposes of epidemiologic research, biomarkers can be broadly grouped into three categories: (1) biomarkers of exposure, (2) biomarkers of effect, and (3) biomarkers of susceptibility. These categories are defined by their position in the spectrum of the exposure-outcome relationship. It is important to note that many biomarkers may play multiple roles, depending on the particular research question being posed. For example, circulating low-density lipoprotein (LDL) may be both a measure of exposure (as a risk factor for coronary heart disease) and a measure of outcome (in a trial of a lipid-lowering therapeutic) within two different studies.

When evaluating exposures, a distinction can be made between the external dose, or the amount that the subject is physically exposed to in their immediate environment, and internal dose, or the total amount of exposure that the subject has absorbed internally over a period of time. Depending on the measure, *biomarkers of exposure* can reflect the internal dose at any stage along its pathway, from the initial unaltered but absorbed exposure to a biologically altered and metabolized form delivered to tissues, and finally to the chemically and/or structurally altered biologic substance as a result of exposure.

Table 1 Types of Biomarkers

Type of Biomarker	*Examples*
Exogenous compounds	
Chemicals (including pesticides)	Polychlorinated biphenyl (PCBs)
Metals	Aluminum, chromium
Food derivatives	Isoflavones (e.g., genistein, daidzein)
Endogenous compounds	
Nucleic acids	DNA, mRNA
Proteins (including antibodies, some growth factors, and hormones)	c-Reactive protein, estrogen receptor
Lipids (such as cholesterol, some steroid hormones, and growth factors)	Triglycerides, estradiol
Molecular characteristics	
DNA sequence variation	Genetic polymorphisms
Cellular characteristics	
Morphologic changes	Sperm motility

Biomarkers identified and measured in body tissues or fluids that are either unchanged or metabolically altered are considered biomarkers of internal dose, while biomarkers of biologically effective dose are markers measured in target or surrogate tissue that reflect the interaction of the absorbed exposure with a subcellular target. Examples of internal and biologically effective dose biomarkers are given in Table 2.

Biomarkers of effect include those that measure biological or biochemical changes in target cells or tissues that occur as result of exposure (Table 3). These may include preclinical biologic effects, such as elevated tumor-specific antigen levels (e.g., prostate-specific antigens or PSA) produced in response to tumor presence, and early-stage disease, such as preneoplastic tissue that may progress to cancer. Molecular markers are also useful in differentiating diseases with the same appearance into different individual, such as estrogen receptor positive (ER+) versus estrogen receptor negative (ER−) breast cancers, thus reducing outcome heterogeneity, improving the precision of effect estimates, and guiding medical treatment.

Biomarkers of susceptibility are similar to effect modifiers in traditional epidemiologic terms. Even when individuals are similar in their environmental exposures, inherent differences in biological responsiveness can produce markedly different doses at the target site and, therefore, different effects. Biomarkers of susceptibility are frequently acquired or genetic factors that influence the response to exposure, yet are preexisting individual characteristics that are independent of exposure. Biomarkers of susceptibility can identify subpopulations of individuals who have a different response to the effects of exposure. Among the more widely studied biomarkers of susceptibility are genetic polymorphisms in enzymes involved in drug and xenobiotic metabolism. Interindividual differences in rates of metabolism will affect the distribution and persistence of different xenobiotic metabolites, which can have downstream implications on biologically effective dose and outcome. For example, individuals with slow *NAT2* acetylator phenotypes have been observed to have higher risk of bladder cancer, relative to fast acetylators, particularly among smokers and those exposed to bladder carcinogens.

Methodologic Considerations in Epidemiologic Studies

Although similar to more traditional measures of exposure, disease, and susceptibility, specific methodologic considerations must be considered when using biomarkers in epidemiologic studies. When selecting an appropriate marker to study, three primary considerations include (1) the feasibility of sample collection, (2) the reduction of systematic error and bias, and (3) ethical issues. Many of these are also applicable to general biomonitoring and exposure assessment studies that use biomarkers.

When selecting a biomarker for incorporation in an epidemiologic study, researchers must consider the feasibility of collecting the type of tissue required (e.g., the invasiveness of the procedure and the availability of the tissue), cost, and ease of assays that will be used for analysis. For example, for measures of certain tumor protein markers, fresh frozen tumor tissue is preferred, yet is procedurally and

Table 2 Examples of Biomarkers of Exposure

Marker	*Exposure*	*Biological Media*
Internal dose		
Cotinine	Nicotine in cigarette smoke	Body fluids
HIV antibodies	HIV virus	Sera or plasma
Asbestos particle	Inhaled asbestos	Lung parenchyma
Biologically effective dose		
DNA adducts	Benozy(a)pyrene	White blood cells
Protein adducts	Ethylene oxide	Red blood cells

Source: Adapted in part from Hulka (1990).

Table 3 Examples of Biomarkers of Effect

Marker	*Disease*
Elevated PSA	Prostate cancer
Dysplastic cervical cells	Cervical cancer
Microsatellite instability (tumor DNA)	MSI + vs. MSS colorectal cancer

logistically difficult to collect. Paraffin-embedded fixed tissue may be more easily obtained from tissue blocks, but some markers may be altered by or degraded prior to the fixative process. In general, it is preferable to collect tissues by the least invasive method possible. Biomarkers that can be measured from tissues such as fingernails, hair, sputum, urine, feces, and expired air are less invasive to the study subject when collecting, while markers requiring the collection of blood, tumor and normal organ tissue, adipose tissue, and bone marrow are considerably more invasive and difficult. Analytic techniques are that automated, nonlabor intensive, and, for large studies, have high throughput can increase the reliability and precision of measures, as well as reduce the study costs.

Measurement error may induce bias in a study involving biomarkers if appropriate measures are not taken to avoid or reduce its effects. The standardization of protocols for biospecimen collection, processing, and storage can help minimize measurement error, as well as procedures to ensure uniformity in the analytic procedures (e.g., using the same assay, technicians). The incorporation of quality assurance and quality control measures into study protocols can also help reduce error. Specific knowledge of the behavior and modifying influences on the biomarker of interest is also important in improving accuracy and validity of measures. Understanding of the normal distribution of a biomarker measure can provide insight into the accuracy of a diagnostic technique (e.g., Are obtained values within the range of expected values?). Knowledge of how biomarker measures might differ if obtained from different tissues (e.g., PCBs measured in serum vs. adipose tissue), using different techniques (e.g., growth factors measured using ELISA vs. IRMA), at different times (diurnal or seasonal variation), under different storage conditions (e.g., -70° C or room temperature, degradation of marker over time), and in relation to subject characteristics, such as disease state, treatment effects, diet, or physical activity.

Finally, as rapid advances in biotechnology increase the accessibility and feasibility of incorporating biological markers into epidemiologic studies, the ethical issues and legal considerations confronting epidemiologists have become particularly relevant. Topics of current interest to policymakers, ethicists, researchers, and clinicians include disclosure of research results (Should researchers inform participants of results, particularly if results have yet to be confirmed or validated?), confidentiality (Who will have access to a subject's biological information? Will information affect a subject's insurability?), banking of samples for future research (Will stored samples be linked by identifying information to subject? Are researchers obligated to recontact subjects to inform them of results of future research?), and the psychological and social risks (Should adjunct services such as genetic counseling referrals be required to communicate results? Is it ethical test for disease, or susceptibility to diseases, for which there is no effective treatment?).

The rapidly evolving technology in cellular and molecular biology and genetic research promises to identify many new biomarkers that will offer both exciting opportunities and technical, ethical, and other challenges in epidemiologic research. This field should lead to improvements in epidemiologic research, better understanding of the natural history of disease, characterization of risk factors for disease, and the development of more effective and targeted medical treatments.

—Libby M. Morimoto and Michael A. Kelsh

See also Environmental and Occupational Epidemiology; Exposure Assessment; Molecular Epidemiology; Pollution

Further Readings

Armstrong, B. K., White, E., & Saracci, R. (1992). *Principles of exposure measurement in epidemiology.* New York: Oxford University Press.

Biomarkers Definition Working Group. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, *69*, 89–95.

Holtzman, N. A., & Andrews, L. B. (1997). Ethical and legal issues in genetic epidemiology. *Epidemiologic Reviews*, *19*, 163–174.

Hulka, B. S., Wilcosky, T. C., & Griffith, J. D. (Eds.). (1990). *Biological markers in epidemiology.* New York: Oxford University Press.

Schulte, P. A., & Perera, F. P. (Eds.). (1993). *Molecular epidemiology: Principles and practices.* San Diego, CA: Academic Press.

Vineis, P., Malats, N., Lang, M., d'Errico, A., Caporaso, N., Cuzick, J., et al. (Eds.). (1999). *Metabolic polymorphisms and susceptibility to cancer* (Vol. 148). Lyon, France: IARC Press.

Biomedical Informatics

Biomedical informatics is broadly defined by the American Medical Informatics Association (AMIA) as the study of "effective organization, analysis, management, and use of information in health care" ("About Informatics," AMIA Web site). Although not requiring the inclusion of computers, the field has grown parallel to the explosive growth of the computer industry, and the two are often associated. Biomedical informatics deals with all aspects of information processing and communication toward a common goal of providing better health care. Its foundations stem from the intersection of computer science, clinical medicine, biomedical engineering, mathematics, and cognitive science. Most health care research today touches some aspect of biomedical informatics, typically through the use or creation of electronic databases, searching online resources, or management of electronic monitoring systems. This entry provides an overview of the history and scope of this field, current research techniques, and applications in both clinical medicine and epidemiology.

The Field of Bioinformatics

The field of bioinformatics can be divided into three broad categories: clinical informatics, public health/population informatics, and translational bioinformatics. Within each of these are potential theoretical and applied divisions. The aims of each branch, however, is the same: to find better ways to research, share, and use the fast-growing fund of medical knowledge.

Clinical informatics is the umbrella term surrounding the information needs of clinicians, medical staff, and patients. It is a broad term covering both inpatient and outpatient medicine and has been the subject of recent state and federal legislation. There are several areas of research within clinical informatics. The goals of each of these areas are to improve patient care through better information management.

Public health or population informatics is similar to clinical informatics except in scope. Population informatics encompasses biosurveillence, preventative medicine research, and disease incidence data.

Translational bioinformatics includes research on tools for acquiring, storing, analyzing, and sharing information. Although not limited to genetic information, this area includes technologies behind the Human Genome Project (1991–2003) that have allowed for processing and storage of large quantities of genetic data and has led to an explosion of information on gene expression, protein synthesis, and biomarker determination. Much of the rapid growth in our understanding of genetics stems from the development of microarrays. A single microarray (a silicon wafer on which genes or gene fragments are attached) can yield 4,000 to 50,000 measurements of gene expression, and many studies use multiple microarrays. Although the biology behind these technologies is beyond the scope of this entry, research in this area continues with the creation and analysis of large genetic databases to facilitate discovery of linkages between genetic information and disease processes. Several new prognostic tests and pharmaceutical agents have been developed through analysis of these data that target specific genes or gene products.

History of Informatics

Some of the early notable achievements in informatics occurred in the 1960s. The first version of the National Library of Medicine's MEDLARS/MEDLINE bibliographic database was released in 1963. In 1966, Massachusetts General Hospital Utility Multiprogramming System was created in Octo Barnett's laboratory as one of the first systems designed specifically for hospital staff to interact with a clinical database. Several large companies such as IBM, 3M, and Hewlett-Packard created medical systems in the 1970s to 1990s with variable success. From 1982 to 1985, the Veterans Administration (VA) began a project to create a decentralized computer network throughout its hospital system, which in 1994 was renamed VistA (Veterans Health Information System and Technology Architecture). It is the current system used to integrate and manage all the clinical information systems throughout the VA network.

The World Wide Web began to connect people internationally in the early 1990s, and approximately 50% of the U.S. households had personal computers by the year 2000. A study by the American Medical Association in 2001 showed that 79% of clinical practices used the Internet to research medical information, and 63% used the Internet to search for drug information. Surveys of medical providers during that same year, however, showed that only 31% of emergency rooms, 29% of outpatient departments, and 17% of physician practices used electronic health records. Thus, around the turn of the century, online resources were in use, but personal records and health systems were mostly paper based. In 2004, President Bush issued an executive order creating the Office of the National Coordinator for Health Information Technology with a goal of nationwide implementation of electronic medical records by 2014.

Much of the recent growth of the use of clinical information systems and of legislation surrounding their implementation stems from two reports released by the Institute of Medicine (IOM) from 1999 to 2001. These reports catalyzed change in the medical community and brought clinical informatics to the forefront by identifying areas where modern medicine was failing to provide safe and effective care. The first report, "To Err Is Human," discussed the prevalence of medical errors, and in particular preventable medical errors. Basing their findings on two studies from different parts of the United States, the IOM estimated that 44,000 to 98,000 deaths occurred each year to medical errors alone. Subsequent reports found that more than 50% of these errors took place during clinician ordering and transcription of these orders. Difficulties in interpreting handwriting, acronyms, and time delays were just a few of the underlying reasons outlined by these reports. Another study on the analysis of handwritten prescriptions in 2005 found that 7.6% contained errors and 2% contained potentially life-threatening errors. Difficulty arises, however, in discussing medical errors in that several definitions of errors have been used. In addition, many studies do not establish reliably of their measures of error, and often no efforts are made to determine which errors led to adverse outcomes.

The second report, "Crossing the Quality Chasm," outlined clinical inequities for various populations as well as disparities between medical knowledge and clinical practice. This report and others found an approximate 5-year delay between evidence-based guidelines and their subsequent widespread use in clinical practice. This delay was more striking in smaller and rural practices.

These two reports stressed that safety was a systems problem and helped to spearhead new initiatives in clinical informatics to study and improve health care information use and delivery. Several key research areas of clinical informatics, including computerized physician order entry (CPOE), electronic prescribing, and the use of electronic medical records, have grown in response to theses concerns.

CPOE attempts to reduce errors at the point of ordering. Clinicians type their orders directly into a computer alleviating transcription errors and handwriting issues. The additional advantage to this method is that a sophisticated computer system can also perform checks on the item ordered ensuring completeness, accuracy, and safety. This type of feedback is called clinical decision support and involves anything from checking prescriptions against patient allergies or formularies to providing clinicians with cost estimates of the procedure or medication they are prescribing. Although the exact prevalence of physician order entry in the United States is unknown, one published survey estimated that only 9.6% of hospitals in the United States have fully implemented CPOE as of 2002. Multiple studies of CPOE with clinical decision support have shown significant reductions in targeted medical errors.

Electronic prescribing (e-prescribing), where the prescription is sent from an ordering clinician to the pharmacy electronically, is seen as one method for eliminating errors of prescription interpretation. Current issues surrounding e-prescribing include security, verifiability of the prescriber, and network creation between patients, pharmacists, and physicians.

Patient Records

Several exciting new developments are under way to allow patients greater access to their medical records, increase the accuracy of these records, and maintain an individual's privacy and security of their medical data.

Many hospitals have established "patient portals" where patients can view parts of their medical records, request prescription refills, schedule appointments, and communicate electronically with their health care provider. Several studies have suggested that such portals could increase patient participation in their medical care and ultimately help maintain the accuracy of their medical record.

In 1996, the Health Insurance Portability and Accountability Act, also known as the Kennedy-Kassebaum bill, was passed into law as a measure to increase access, portability, and security of a person's medical records. The law outlined procedures and circumstances where one's medical records could be shared and established a requirement that the amount of information disclosed be the "minimum necessary." Although initially envisioned as a blueprint for facilitating mobility of patient information between providers, the largest effect of the bill has been in establishing security protocols for the storage and dissemination of personal medical data. These safety protocols cover clinical as well as research use of data.

Text Mining and Natural Language Processing

Unfortunately, most of the clinical records in the United States are not computerized. Of the records on computer systems today, much of the information is not stored in a structured, machine-readable way. Lab reports, admission and discharge summaries from hospital stays, and medication lists all contain useful information that could be used by a computer or clinician to make future decisions about a patient's care. However, parts of the report are often not identified specifically; as a result, a large amount of time is required for chart review and data extraction. Text mining is the process of information identification and extraction through pattern recognition. An example might be in searching for patient charts to find all smokers. Charts can be searched to identify phrases about smoking such as "packs per day" or "nicotine," and specific charts can be tagged for further analysis. Natural language processing goes one step further in that it analyzes the meaning of a text using techniques such as identifying contextual cues, isolating parts of speech, and phrase recognition. A very simple example of this is the recognition of negative phrases in a sentence: "Formerly a two-pack per day smoker, Mr. Smith has been controlling his cravings for nicotine successfully for the past 3 months with gum and patches." Searching for "packs per day" alone might suggest that the patient is a current smoker, although phrase analysis would show that the main idea of the sentence is that Mr. Smith has been controlling his smoking habit. Methods such as these are being used to better isolate and tag key medical information for use in clinical systems today.

Biosurveillance

Biosurveillance is the study of disease outbreaks that occur naturally or as part of a bioterrorist attack. In 1998, the Clinton administration authorized a system to provide early warning against bioterrorist attacks. Currently, the government is investing heavily in building the infrastructure necessary to detect outbreaks, share information, and analyze data of this type. Research in this area is focused on interoperability between diverse systems, signals detection of new outbreaks, and measures for hospitals to manage outbreaks.

Syndromic surveillance is the detection of health indicators before official diagnoses are made through pattern recognition and signal variance detection. The process involves mapping of natural fluctuations of disease prevalence and monitoring for changes in these patterns. An example of this might be analysis of respiratory syndromes in emergency rooms. These syndromes typically have a peak in late fall or early winter. By monitoring emergency room visits using patient complaints and symptoms as markers of respiratory illness, comparisons of the current findings can be made with previous years, or between emergency rooms themselves. This type of data has also been used in studies to determine vaccination timing of different age groups.

Informatics and Cognitive and Social Science

Equally as important as the tools we use is how we interact with them. Cognitive studies in informatics are helping us understand how we process information, human-computer interactions, and insights into how clinicians make decisions based on available, and often incomplete, information. Topics within the purview of cognitive science include the study of interface design, such as the best ways to present trends of lab results, or sophisticated analysis of ICU decision making.

There have also been several studies of successful and failed implementations of electronic medical record systems. Implementation within a hospital system is a large financial and operational commitment; yet even after investing millions of dollars in a system, implementations may fail. Two notable failures include the University of Virginia's CPOE system in 1993 and Cedar-Sinai's CPOE system in 2003, both of which

were stopped due to complaints by physicians using the system. Typically, the reason for success or failure is a systems management issue surrounding training on the new system, workflow changes, or participation of various departments in the planning stages. The University of Virginia failed to involve the physicians in the creation and implementation of the system and subsequently mandated its use without adequate understanding of the changes it would require, resulting in an organized physician boycott. At Cedars-Sinai Medical Center, physician buy-in was solicited from only a few users prior to implementation, and the system required significant workflow changes and time for users to perform common tasks. Current implementation efforts, having learned from studies outlining these difficulties, now spend considerable resources to understand how new information systems interact and affect other hospital resources.

Biomedical informatics is a diverse and growing field. It provides tools for patients, nurses, doctors, pharmacists, and researchers to use in order to manage their information needs. It will also help us ask better questions, find better answers, and share what we learn as we continue to build on the information technologies we have today.

—Michael Jernigan

See also Clinical Epidemiology; Genetic Epidemiology; Outbreak Investigation; Public Health Surveillance

Further Readings

Ash, J. S., Gorman, P. N., Seshadri, V., & Hersh, W. R. (2004). Computerized physician order entry in U.S. hospitals: Results of a 2002 survey. *Journal of the American Medical Informatics Association*, *11*, 95–99.

Bates, D. W., Kuperman, G. J., Wang, S., Gandhi, T., Kittler, A., Volk, L., et al. (2003). Ten commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. *Journal of the American Medical Informatics Association*, *10*, 523–530.

Bates, D. W., Leape, L. L., Cullen, D. J., Laird, N., Petersen, L. A., Teich, J. M., et al. (1998). Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *Journal of the American Medical Association*, *280*, 1311–1316.

Brownstein, J. S., Kleinman, K. P., & Mandl, K. D. (2005). Identifying pediatric age groups for influenza vaccination using a real-time regional surveillance system. *American Journal of Epidemiology*, *162*(7), 686–693.

Collen, M. F. (1994). The origins of informatics. *Journal of the American Medical Informatics Association*, *1*(2), 91–107.

Cullen, D. J., Sweitzer, B. J., Bates, D. W., Burdick, E., Edmondson, A., & Leape, L. L. (1997). Preventable adverse drug events in hospitalized patients: A comparative study of intensive care and general care units. *Critical Care Medicine*, *25*, 1289–1297.

Ford, E. W., Menachemi, N., & Phillips, M. T. (2006). Predicting the adoption of electronic health records by physicians: When will health care be paperless? *Journal of the American Medical Informatics Association*, *13*, 106–112.

Hoffer, T. P., & Kerr, E. A. (2000). What is an error? *Effective Clinical Practice*, *6*, 261–269.

Jha, A. K., Kuperman, G. J., Teich, J. M., Leape, L., Shea, B., Rittenberg, E., et al. (1998). Identifying adverse drug events: Development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *Journal of the American Medical Informatics Association*, *5*, 305–314.

Massaro, T. A. (1993). Introducing physician order entry at a major academic medical center: I. Impact on organizational culture and behavior. *Academic Medicine*, *68*(1), 20–25.

Web Sites

American Medical Informatics Association (AMIA): http://www.amia.org/informatics.

Bioterrorism

Bioterrorism can be defined as the deliberate release of biological agents with the intention of infecting civilians for belligerent purposes. These agents are usually naturally occurring microorganisms such as bacteria, viruses, and fungi; however, they can also be engineered to be more deadly, drug resistant, or easier to transmit. Bioterrorists may spread disease through air, water, and food, and from person to person, infecting humans, animals, or plants. Such an attack on humans with an agent causing a disease such as smallpox or botulism could lead to severe illness and possibly death and could result in widespread fear among the public. Furthermore, it could result in significant economic, social, and political damage as critical infrastructures struggle to handle the effects of disease on the population. Experts are concerned about the possibility of bioterrorist agents falling into the hands of groups or individuals who could use them to inflict devastation in the name of political, personal, or religious beliefs.

Development and Use of Bioterrorist Agents

In 1984, members of the Rajneeshees, a religious cult, sprayed *Salmonella typhimurium* on salad bars in The Dalles, Oregon, causing 750 cases of food poisoning, though no deaths occurred. In the 1990s, the Aum Shinrikyō extremist cult in Japan attempted to use bioterrorism agents but were unsuccessful because they could not overcome the technical hurdle of dissemination. They later released sarin gas in the Tokyo subway. During the anthrax attacks of September 2001, *Bacillus anthracis* spores sent through the U.S. Postal Service resulted in 22 cases of the disease and five deaths.

Russia, the United States, and Iraq ran biological weapons programs during the 20th century, but these were dismantled by the end of the century. Before closing, these programs employed thousands of scientists, and concern remains among experts that some of the scientists might have taken their bioweapons knowledge and possibly samples of biological agents to sell to the highest bidder.

Rationale for Using Biological Weapons

While concern over the likelihood of bioterrorist attacks has risen in light of recent attacks that indicate an erosion of taboos against mass killings, there are still some barriers to obtaining and using the necessary agents. For example, bioterrorist agents are difficult to produce in large quantities, and they can be difficult to aerosolize for maximum spread. But while bombs and guns may still be favored by terrorist organizations because of their ease of use and accessibility, advances in biotechnology, access to information on the Internet, and the availability of dual-use equipment and technologies increase the probability that biological weapons will be used in the future.

At present, terrorist organizations can obtain small amounts of biological agents from hundreds of germ banks worldwide and use them to produce large quantities of potential bioweapons. Agents could also be stolen from research laboratories—a possible explanation for the anthrax letters of 2001. They might be purchased from a former bioweapons scientist or a rogue nation. With the proper information, terrorists could isolate and grow agents from natural sources; anthrax can be found in animal hides and soil, while tularemia is found in soil. Of special concern, the developing field of genetic engineering makes it possible to create new or genetically modified biological agents that are more deadly than those presently available, and for which there are no cures. This was the goal of the scientists in the bioweapons program in the former Soviet Union.

Once they are available, these biological agents offer terrorists many advantages. In 1999, the Gilmore Commission identified five possible reasons why terrorists might use weapons of mass destruction, including bioweapons:

1. to kill as many people as possible;
2. to exploit the classic terrorist weapon—fear;
3. to allow terrorist groups to negotiate from a position of strength;
4. to take advantage of the fact that a biological attack could go undetected for some time, allowing terrorists to escape; and
5. to cause economic and social damage by targeting the agricultural sector.

In addition, experts calculate that the cost per life taken is considerably less for biological agents than for standard terrorist weapons, and biological agents are becoming increasingly easy to produce.

Dissemination and Detection of Biological Agents

More than 60 potential biological agents are presently available to bioterrorists. In liquid or powder form, the agents can be dispersed as an aerosol made up of particles small enough to enter the lungs. The North Atlantic Treaty Organization has identified 31 bioterrorist agents they believe are most likely to be used, and the U.S. Army Medical Research Institute of Infectious Diseases has further narrowed this list to 6, based on availability, ease of production, lethality, stability, and infectivity: anthrax, smallpox, plague, tularemia, botulinum toxin, and agents of viral hemorrhagic fever.

Dissemination of biological agents can occur either indoors or outdoors, though an outdoor attack would require a larger quantity of the agent. Terrorists might create an aerosol cloud of dried anthrax spores and spread the disease by equipping low-flying airplanes or trucks with sprayers. The agent could be spread indoors using a small aerosol

canister equipped with a remote device or by a suicide terrorist or someone who has received prophylactic treatment against the agent.

Early detection of bioterrorist attacks requires a universally available surveillance network that can provide the necessary information accurately and in a timely manner. However, of those systems presently available, some use unreliable diagnostic codes that may fail to determine the exact nature of an agent or that may fail to detect it at all when there are too few pathogens. Syndromic surveillance, the most likely to be effective, is so labor intensive that its use remains limited. At present, discovery of a bioterrorist attack is likely to be made by an alert physician who diagnoses a rare disease or identifies a suggestive pattern of disease. However, since most physicians have not seen diseases caused by bioterrorist agents, they will not be able to determine when an attack has occurred.

The federal government uses a surveillance system called BioWatch, a network of detection equipment set up around the country to detect the release of biological agents. However, biological agents must be released in large quantities or in close proximity to the detectors for the system to recognize them. The equipment is also subject to frequent false positives, so it is useful only in certain high-risk areas.

Protection, Prophylaxi, and Treatment

Many countries, including the United States and the United Kingdom, have developed national response plans for dealing with natural disasters and acts of terrorism. These plans will play an important role in the response to bioterrorism and, together with plans for global infectious disease outbreaks, will provide effective response mechanisms that will require only limited adaptations to meet the challenge of bioterrorist attacks. Responders will have to modify existing laws to maximize effectiveness of responses. The Model Emergency Health Powers Act, developed by U.S. lawyers and adopted by some states, can be used to enhance state governments' response capability.

In the event of a bioterrorist attack, the National Bioterrorism Response Plan and planning documents from local and state health departments and emergency management agencies would establish protocols for mass prophylaxis distribution. Hospitals are currently working to increase their surge capacity to meet the challenges of an attack, and the Strategic National Stockpile (SNS) contains vaccines, medications, and equipment for use during a mass casualty event and is making efforts to increase that supply. However, levels of preparedness vary widely across the nation, and the SNS will be inadequate in the event that simultaneous attacks are launched against a number of targets.

—R. Gregory Evans and Rachel D. Schwartz

See also Plague; Smallpox; War

Further Readings

Alexander, Y., & Hoenig, M. (Eds.). (2001). *Super terrorism: Biological, chemical, and nuclear.* Ardsley, NY: Transnational.

Bronze, M. S., & Greenfield, R. A. (Eds.). (2005). *Biodefense: Principles and pathogens.* Norfolk, UK: Horizon Bioscience.

Ursano, R. J., Norwood, A. E., & Fullerton, C. S. (2004). *Bioterrorism: Psychological and public health interventions.* Cambridge, UK: Cambridge University Press.

Bipolar Disorder

Bipolar disorder is a mental illness characterized by drastic mood swings from very high (manic phase) to very low (depressive phase). The diagnosis of bipolar disorder is usually based on the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association (fourth edition) criteria, a tool developed by the American Psychiatric Association for clinicians.

The *DSM-IV* includes two types of bipolar disorder diagnoses, Bipolar I and Bipolar II. Bipolar I disorder is identified when a person presents with one lifetime occurrence of either a manic episode or a mixed episode. Although bipolar disorder is known for the polarization of moods, a diagnosis of Bipolar I does not require the presence of a depressive state. The manic episode alone can satisfy the criterion for Bipolar I, if the mania is not caused by another mental disorder. A manic person can be described as a normal person in fast-forward. When a patient is experiencing a manic episode, he or she experiences

an elevated, irritable, or expansive mood. To meet the diagnostic criterion specified by the *DSM-IV*, the mood must occur for longer than 1 week and must be accompanied by three or more other symptoms not caused by mood-altering substances, medications, or other medical conditions. Other symptoms may include any of the following: inflated self-esteem or grandiosity, decreased need for sleep, being more talkative than usual or experiencing pressure to keep talking, flight of ideas or subjective experience that one's thoughts are racing, distractibility, increase in goal-directed activity or psychomotor agitation, or excessive involvement in pleasurable activities that have a high potential for painful consequences.

In the other form of Bipolar I disorder, mixed bipolar episodes—the other criterion for Bipolar I—the symptoms of mania and depression occur simultaneously; symptoms include agitation, trouble sleeping, and significant change in appetite, psychosis, and suicidal ideation. A person may feel sad and hopeless, yet extremely energized.

The diagnosis of Bipolar II disorder, according to the *DSM-IV*, is a clinical course characterized by one or more episodes of major depression and at least one hypomanic episode. In contrast to the manic episode, the depressed pole of bipolar disorder is characterized by extreme sadness. A depressive episode (depression) of bipolar disorder is diagnosed with the same criteria as major depressive disorder. The criteria for diagnosis are the presence of five or more symptoms lasting most of the day for 2 weeks or longer. At least one of the symptoms must include depressed mood or loss of pleasure or interest, and symptoms must not be due to other physiological conditions or substances. The symptoms may include any of the following in addition to depressed mood or loss of interest and pleasure in all or almost all activities: decreased or increased appetite/change in weight; insomnia or hypersomnia; psychomotor agitation or retardation; fatigue or loss of energy; feelings of worthlessness, or excessive or inappropriate guilt; diminished ability to think or concentrate, or indecisiveness; or recurrent thoughts of death, suicidal ideation with or without a plan or suicide attempt.

Hypomania, the other element of Bipolar II, is defined as a mild to moderate level of mania. Hypomania may feel good to a person and may be accompanied by good functioning and enhanced productivity, but without treatment it may lead to mania or swing into depression. Also, psychosis or psychotic symptoms may occur in severe cases of mania or depression. Common symptoms include auditory and visual hallucinations and delusions.

Most researchers now agree that there is no single cause for bipolar disorder. Etiology of the disease may be attributed to three factors: genetic vulnerabilities, biological vulnerabilities, and levels or styles of coping with socioenvironmental stress. Also, other factors such as environment, life events, and individual attributes may contribute to the cause.

Throughout history and dating back to ancient Greece, the presence of bipolar disorder, formally known as manic-depressive disorder, has captivated scholars. More than 80 years ago, the course of this illness was described systematically, and the heterogeneity in the types of symptoms, the pattern of episodes, and the level of functioning was noted. Those in literature, the arts, and history have been inspired by the creativity of individuals with bipolar disorder, including Vincent Van Gogh, Martin Luther, Robert Schumann, Pytor Illyich Tchaikovsky, and the Pulitzer Prize winners John Berryman, Amy Lowell, and Anne Sexton.

Two major community surveys in the United States of the lifetime prevalence of bipolar disorder indicate that from 1.0% to 1.6% of adults and 1.2% of children and adolescents (9 to 17 years) are affected by this illness. Misdiagnoses and underdiagnoses of bipolar disorder are ongoing concerns, especially among adolescence and children, and may contribute to incorrect prevalence of the disorder. Community studies of this issue are difficult to conduct, however, as standard epidemiological surveys may fail to reliably diagnose as many as 50% of cases. Researchers have identified several features of bipolar depression that may distinguish it from unipolar depression, including longer episode duration, increased probability of psychotic symptoms, and limited efficacy of antidepressant medications.

With regard to gender and bipolar disorder, Bipolar I affects men and women equally, but Bipolar II disorder is more common in women. Males may also present differently than females with the disorder. Commonly, the first episode in males is a manic. Females are more likely to be depressive.

The onset of bipolar disorder may occur at anytime during the life span, but the median age of onset for the disorder is 18 years of age. The Epidemiologic Catchment Area study reported a mean age of 21 years for bipolar disorder. Although bipolar disorder can be present before the onset of puberty, the

available evidence suggests that full-blown mania in children is relatively rare. At the opposite end of the age spectrum, newly diagnosed mania in patients more than the age of 65 is uncommon. There are no known significant differences among racial groups in prevalence of either Bipolar I or Bipolar II disorder.

—Keneshia Bryant

See also Child and Adolescent Health; Psychiatric Epidemiology

Further Readings

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (1995). *Practice guideline for treatment of patients with bipolar disorder.* Washington, DC: Author.

Hofmann, S., & Tompson, M. (2002). *Treating chronic and severe mental disorders: A handbook of empirically supported interventions.* New York: Guilford Press.

Johnson, S., & Leahy, R. (2004). *Psychological treatment of bipolar disorder.* New York: Guilford Press.

Nemeroff, C. (2001). *The Medical clinics of North America: Advances in the pathophysiology and treatment of psychiatric disorders; implications for internal medicine.* Philadelphia: W. B. Saunders.

Web Sites

National Institute of Mental Health. Bipolar Disorder: http://www.nimh.nih.gov/healthinformation/bipolarmenu.cfm.

Birth Certificate

Birth certificates provide important data about a newborn's status, the child's parents, the mother's use of prenatal care, and complications and events during pregnancy. As such, they are a key resource for maternal and child health epidemiology. They also serve as the basis for legal identity. The history, organization, availability, content, uses, and accuracy of birth certificates are summarized in this entry.

In the United States, the annual collection of birth statistics on a national basis began in 1915, with 10 reporting states and the District of Colombia; by 1933, it included the entire country with at least 90% coverage in each state. The standard certificate of live birth was developed in 1900 and has since undergone 12 revisions, with the 1989 and 2003 revisions expanding the content of medical information. National statistics on fetal deaths have been compiled annually since 1922.

The National Center for Health Statistics (NCHS) collates national and state data on live births, deaths, and fetal deaths based on vital records filed in state registration offices. States generally adhere to the standard certificates developed by the NCHS but can choose to add or delete items. The NCHS regularly publishes reports on birth certificate statistics and prepares public use data sets for further analysis. These include a Linked Birth/Infant Death File (linking live birth and infant death [0 to 365 days] certificates for all infants born during a cohort year) that may be used to help explain trends in infant heath and mortality. Technological advances further allow linkages with data for Medicaid enrollment and other publicly funded health programs to assess quality of perinatal outcomes.

Birth certificates have two parts. The first provides demographic data on the parents and infant; the second, completed from hospital records, gives data on maternal and infant health. Birth certificates offer useful information for researchers, policymakers, and state officials to evaluate trends in maternal and infant health and the quality of care delivered to pregnant women. Birth certificate data allow for the creation of many potential health status outcome and health risk indicators.

The reliability and validity of birth certificate data vary considerably by item, but their quality and completeness are reasonable for population-based analyses. Basic demographic characteristics (e.g., maternal age) and maternal data tend to be more accurately and completely reported than those for fathers or social traits (e.g., education). Medical information about the birth or the newborn is typically least adequately reported. Data are more likely to be missing for very low birthweight infants, teenage and unwed mothers, and those with less than a high school education. Absent accurate data, however, high-risk groups are less likely to be understood or helped, and inaccurate conclusions may be drawn about underlying variables of interest, such as substance abuse during pregnancy.

—Andrzej Kulczycki

See also Maternal and Child Health Epidemiology; National Center for Health Statistics

Further Readings

Northam, S., & Knapp, T. R. (2006). The reliability and validity of birth certificates. *Journal of Neonatal, Gynecologic, and Neonatal Nursing*, *35*, 3–12.

Birth Cohort Analysis

A birth cohort is a group of individuals born during a given calendar time period within a specified geographical region. For example, the U.S. 1950 annual birth cohort refers to the group of people born in the United States during the calendar year 1950, and the 1950 to 1952 birth cohort identifies those born during the period covering the three consecutive calendar years 1950, 1951, and 1952. Birth cohort analysis is an observational cohort analysis (as opposed to an experimental cohort analysis such as clinical trials) of an entire birth cohort or of a selected sample of the birth cohort.

Data for birth cohort analysis are best depicted by a Lexis diagram with the horizontal axis denoting calendar time and the vertical axis denoting age wherein each individual live birth is represented by a lifeline with 45° inclination from the horizontal axis. Starting at birth on the horizontal axis, each individual is followed continuously as he or she passes through different ages during all or part of the life span, that is, until death or censoring (such as out-migration), whichever occurs first. Events of interest occurring to an individual during his or her lifetime are marked on his or her lifeline. The events of interest typically include exposure to risk factors (such as start of smoking) and occurrence of health outcomes (such as incidence or recurrence of a disease or medical condition) as well as vital events (such as giving birth, change of marital status, and death). Each marked lifeline represents the complete life history of an individual and constitutes a sample path. So the life histories for all individuals (all marked lifelines) in the birth cohort constitute the sample space.

Since we start at the beginning of life at which time one is susceptible to almost all risk factors, events, and health outcomes, almost any of them can be studied by birth cohort analysis. In particular, the recorded birth cohort data as described above allow calculation of both cumulative incidence and incidence density and so can be used to perform the following cohort analyses retrospectively.

Event History Analysis Using Survival Time Data. When sample size is not too large, this can be done by applying likelihood methods and martingale theory to produce statistical inference (maximum likelihood or martingale estimates and hypothesis testing comparing different birth cohorts) of disease incidence rates and of effects of risk factors on disease incidence.

Construction of Cohort Life Tables Using Age Group Data (for Large Populations). Such tables include attrition life tables and the most general increment-decrement life tables to obtain transition probabilities from one state to another and to obtain an estimate of the expected duration of stay in each state. In addition to comparing two different cohort life tables, one can also compare a cohort life table with a period life table, both constructed on the same base period, to see if there exists any period-cohort effect or birth cohort effect.

Comparing the Health Outcome or Vital Event of Interest (Say, Mortality) of the Birth Cohort With That of the Corresponding General Population to See if the Special Life Event Experienced by the Birth Cohort Has Caused a Significant Change in Mortality Level. This may be done by calculating the standardized mortality ratio (SMR) and testing the null hypothesis $SMR = 1$ using the asymptotic unit normal test statistic $Z = \ln SMR/\sqrt{D}$, where $SMR = D/E$ and D and E are the observed deaths and expected deaths (based on the mortality of the corresponding general population), respectively, in the cohort.

Comparing Birth Cohorts of Different Calendar Periods for Period-Cohort Effects, Using Logrank and Related Tests for Survival Time Data and Logistic Regression for Binary Outcome Data. Period-cohort effect arises because different birth cohorts may experience different levels of health outcomes as they grow up exposed to differing environmental and societal changes, which in turn lead to differing behavior changes. For example, age-sex patterns of lung cancer mortality between the 1900 birth cohort and the 1950 birth cohort differ greatly because prevalence of smoking, amount smoked per day, age at initiation, and lifetime smoking duration are all influenced by the calendar year of birth. These covariate data can be used to investigate the effects of different kinds of risk using the extended Cox

regression model, Poisson regression, logistic regression model, or structural equations models.

Analysis of Panel Data and Data on Repeated Measures of an Outcome Variable. Methods for analyzing longitudinal data such as generalized estimating equations and random coefficients analysis may be used to account for intraperson time series correlation due to repeated observations on the same individual as well as missing data.

Finally, when birth cohort data are not available but a series of consecutive cross-sectional data are available, one can still form birth cohorts by concatenating from serial cross-sectional age-specific data. What has often been done in this case is to construct an open birth cohort by concatenating a series of period/age-matched cross-sectional data. For example, Kemm determined for Great Britain the percentage of current smokers and ever smokers by age in successive birth cohorts and the percentage of ever smokers who continue, as well as alcohol consumption, by analysis of data from serial cross-sectional surveys of smoking status.

—John J. Hsieh

See also Clinical Trials; Cohort Effects; Descriptive and Analytic Epidemiology; Event History Analysis; Life Tables

Further Readings

Breslow, N., & Day, N. (1987). *Statistical methods in cancer research: Vol. 2. The design and analysis of cohort studies*. IARC Scientific Publication 88. Lyon, France: IARC Scientific.

Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society of London, Series A*, *332*, 487–509.

Kemm, J. R. (2001). A birth cohort analysis of smoking by adults in Great Britain, 1974–1998. *Journal of Public Health Medicine*, *23*(4), 306–311.

Woodward, M. (1999). *Epidemiology: Study design and data analysis*. Boca Raton, FL: Chapman & Hall.

Birth Defects

Birth defects, also known as congenital malformations or anomalies, are abnormalities in the structure or function of organs that are present at birth. They are the leading cause of infant death and contribute to morbidity and long-term disability in the population. Birth defects are a heterogeneous group of outcomes that are not always apparent at delivery. Major birth defects are structural malformations that are either lethal, require medical or surgical treatment, or are of cosmetic importance.

The frequency of congenital anomalies is highest in pregnancies that result in miscarriages or stillbirths. Major birth defects occur in 2% to 3% of live births in the United States. Some of the most common types of birth defects are heart defects (1 in 100 to 200), neural tube defects (1 in 1,000), orofacial clefts (cleft lip and cleft palate; 1 in 700 to 1,000), and hypospadias (abnormal development of the urethra; 1 in 200 to 300 males). Down syndrome, a chromosomal disorder known as trisomy 21, occurs in about 1 in 800 births.

Risk Factors

While approximately 10% of birth defects are attributed to environmental factors and 20% are attributed to single-gene or chromosomal defects that may be inherited or represent new mutations, the causes of the remaining 70% of birth defects remain unknown. Most birth defects are believed to arise from interactions between genes or interactions between environmental factors and genes.

Teratogens are agents that can cause birth defects. The timing of exposure to a teratogen during pregnancy is important: The greatest risk for structural malformations exists for exposures occurring between the third and eighth week of gestation (the embryonic period), when most organ systems are developing (organogenesis). The period in which an organ or system is at greatest risk of damage from a teratogen is referred to as the critical period: For some organs and systems, this period extends beyond the eighth week into the fetal period. However, structural birth defects are less likely to occur after the eighth week because most organ systems have already been established. Some well-known human teratogens include thalidomide (a sedative), isotretinoin (an antiacne medication sold under the brand name Accutane), valproic acid (an antiseizure medication sold under the brand name Depakote), warfarin (an anticoagulant sold under the brand name Coumadin), and mercury (a heavy metal).

Other factors have been identified as risk factors for birth defects, such as dietary deficiencies (vitamin A), maternal behaviors (smoking and alcohol

consumption), maternal illnesses (diabetes mellitus and rubella infection), and family history of a similar defect (hypospadias). Some defects vary by maternal age (Down syndrome), race (tetralogy of Fallot), ethnicity (spina bifida), or infant sex (anencephaly). Epidemiologic research has demonstrated that taking a multivitamin containing folic acid, a B vitamin, before conception and in early pregnancy lowers the risk of having a baby with a neural tube defect. Other research suggests that folic acid may reduce the risk of other types of birth defects. Since the mandated folic acid fortification of cereal grains and flour in the United States in 1998, the occurrence of neural tube defects has decreased and reductions in other birth defects have been observed.

Methodologic Challenges

Researching the causes of birth defects is challenging because specific birth defects are rare, the developing embryo and fetus are exposed to a variety of genetic and environmental factors during pregnancy, and the biologic mechanisms that cause most birth defects are unknown. Due to pregnancy losses, the defects identified at birth represent only the birth prevalence, not the true incidence of the condition. The retrospective ascertainment of exposure in epidemiologic studies is a concern because mothers of infants with defects may more accurately recall exposures during pregnancy than mothers of healthy infants, a bias known as recall bias. Adding further complexity to the study of birth defects is the interplay between genetic and environmental factors in the etiology of many anomalies.

Ongoing Surveillance and Research

Public health surveillance systems are important in collecting and analyzing data on birth defects in human populations. The systematic and ongoing monitoring of births of malformed infants in the population allows for the description of birth defect patterns that may suggest environmental causes, such as infections, drugs, other chemicals, or physical agents. The Metropolitan Atlanta Congenital Defects Program was created in 1967 in response to the thalidomide tragedy. Surveillance data from the program have been used to describe the epidemiology of birth defects and evaluate possible etiologic factors. The program also served as the source of data for the Atlanta Birth Defects Case-Control Study, which helped increase the understanding of risk factors associated with birth defects. The National Birth Defects Prevention Network was organized in 1997 with the goal of establishing a network of population-based birth defects surveillance and research programs to assess the impact of birth defects, identify risk factors for targetting primary prevention activities, and assist in the prevention of secondary disabilities.

To help reduce the burden of birth defects in the United States, Congress passed the Birth Defects Prevention Act of 1998 (Public Health Law 105–168). The bill authorized the Centers for Disease Control and Prevention to

1. collect, analyze, and make birth defects data available;
2. operate regional centers for applied epidemiologic research on the prevention of birth defects; and
3. educate the public about the prevention of birth defects.

As a result, the Centers for Disease Control and Prevention established the Centers for Birth Defects Research and Prevention, which funded centers around the United States. The main activity of each center is to participate in the National Birth Defects Prevention Study, the largest collaborative birth defect study in the United States. Approximately 35 categories of birth defects are included in this ongoing case-control study. The study seeks to improve the study of birth defects by including a large ethnically and geographically diverse birth population that will provide unprecedented statistical power to evaluate potential risk factors, more etiologically homogeneous case definitions for specific birth defects groups, an interview with questions on a wide array of exposures and potential confounders, and the collection of DNA for the study of genetic susceptibility and interactions between genes and the environment. Cases are identified from population-based birth defect surveillance systems, and controls are randomly selected from birth certificates or birth hospital records. Mothers of infants are interviewed and parents are asked to collect cheek cells from themselves and their infants for DNA testing. Because of the large sample sizes, scientists will be able to study the epidemiology of some rare birth defects for the first time. The combined interview data and banked DNA will enable

future research as new hypotheses and technologies develop.

—Alissa R. Caton

See also Genetic Epidemiology; Mercury; Newborn Screening Programs; Teratogen; Thalidomide

Further Readings

Canfield, M. A., Collins, J. S., Botto, L. D., Williams, L. J., Mai, C. T., Kirby, R. S., et al. (2005). Changes in the birth prevalence of selected birth defects after grain fortification with folic acid in the United States: Findings from a multi-state population-based study. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, *73*(10), 679–689.

Canfield, M. A., Honein, M. A., Yuskiv, N., Xing, J., Mai, C. T., Collins, J. S., et al. (2006). National estimates and race/ethnic-specific variation of selected birth defects in the United States, 1999–2001. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, *76*(11), 747–756.

Correa-Villasenor, A., Cragan, J., Kucik, J., O'Leary, L., Siffel, C., & Williams, L. (2003). The Metropolitan Atlanta Congenital Defects Program: 35 years of birth defects surveillance at the Centers for Disease Control and Prevention. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, *67*(9), 617–624.

Yoon, P. W., Rasmussen, S. A., Lynberg, M. C., Moore, C. A., Anderka, M., Carmichael, S. L., et al. (2001). The National Birth Defects Prevention Study. *Public Health Reports*, *116*(Suppl 1), 32–40.

Web Sites

International Clearinghouse for Birth Defects Monitoring Systems (for an international perspective): http://www.icbdsr.org.

National Center on Birth Defects and Developmental Disabilities of the Centers for Disease Control and Prevention: http://www.cdc.gov.

Bloodborne Diseases

Bloodborne diseases are caused by pathogens such as viruses or bacteria that are carried in the blood. In the United States, the most common bloodborne diseases are hepatitis B, hepatitis C, and HIV/AIDS. Hemorrhagic fevers, including Ebola, are not a health threat in the United States, but sporadic outbreaks have occurred in Africa and other parts of the world since 1976. Common routes of infection with bloodborne diseases include unprotected sexual activity, contact with blood through needles or other sharps, and transmission from mother to child during the birth process.

Common Bloodborne Diseases

Hepatitis B is caused by infection with the hepatitis B virus (HBV). In about 30% of cases, the person experiences no symptoms; others may experience jaundice, fatigue, abdominal pain, loss of appetite, nausea, vomiting, and joint pain. Infection may become chronic, particularly to infants infected at birth, and may lead to death from chronic liver disease (15% to 25% of all cases). A hepatitis B vaccine has been available since 1982 and is recommended for people in high-risk groups, including health workers, household members and sexual partners of persons infected with HBV, injection drug users, and persons traveling to or living in parts of the world where HBV infection is endemic. According to the Centers for Disease Control and Prevention (CDC), about 300,000 cases of HBV occur in the United States annually.

Hepatitis C is caused by infection with the hepatitis C virus (HCV). Persons at highest risk include injection drug users, persons treated for blood-clotting problems before 1987 or who received a blood transfusion before 1992, and hemodialysis patients. Eighty percent of individuals infected with HCV have no symptoms, but hepatitis C causes liver damage and is a leading indicator for liver transplants. About 70% of infected individuals develop chronic liver disease and 5% to 20% will develop cirrhosis. There is no vaccine for hepatitis C but several drug treatments are available. According to the CDC, about 26,000 new cases of HCV were identified in the United States in 2004, and 4.1 million people are living with the disease.

HIV (human immunodeficiency virus) is a virus transmitted primarily through blood; typical routes of infection include unprotected sexual activity and use of unsterilized needles (and outside the industrialized world, use of other unsterilized sharps and transfusion of contaminated blood). Acquired immunodeficiency syndrome (AIDS) is a label given to the advanced stages of HIV infection, when the person's immune system starts failing and he or she is subject to many opportunistic infections as well as

unusual cancers such as Kaposi's sarcoma. There is no vaccine for AIDS but a number of medical treatments are available. The CDC estimates that there are approximately 44,000 new infections annually; approximately 1.1 million people in the United States were living with HIV/AIDS in 2003.

Viral hemorrhagic fevers (VHFs) are a group of illnesses caused by bloodborne viruses, which damage the vascular system and cause hemorrhage (bleeding). Most VHFs are zoonotic, with rodents and arthropods the main reservoirs; however, the hosts of Ebola and Marburg viruses, two of the best-known VHFs, are unknown. Symptoms of infection with a VHF include fever, fatigue, dizziness, muscle aches, loss of strength, exhaustion, and bleeding under the skin, in internal organs, and from body orifices. There are no vaccines for VHFs and treatment is primarily supportive. Prevention of VHF infection is based on preventing contact with host species, for example, controlling rodent populations, keeping them out of homes and workplaces, and cleaning up rodent nests and droppings. VHFs can be spread through physical contact, so isolation of infected individuals is recommended, with Universal Precautions (defined below) observed by health care workers treating VHF patients.

Prevention and Control

The group at greatest risk for bloodborne infection, worldwide, is health workers who are exposed to blood and other body fluids in the course of their work. In the industrialized world, most other cases of bloodborne disease are caused by injection drug use (with unsterile needles) or are sexually transmitted. In the developing world, ordinary medical care such as receiving an injection may carry a high probability of infection due to lack of sterilization. Use of barrier methods (condoms and dental dams) can prevent or reduce the transmission of most sexually transmitted diseases, and sterilization or use of disposable sharps (e.g., needles and scalpels, which are intended to penetrate the skin) can sharply reduce the risk of infection during medical care.

The World Health Organization estimates that 3 million health workers annually experience percutaneous (needle stick or other sharps injury) exposure to bloodborne pathogens, about two thirds of those to hepatitis B. Most of these exposures are preventable, and a set of procedures known as Universal Precautions has been developed to minimize infection. Immunization against hepatitis B and postexposure management such as the provision of prophylactic medication are also recommended for health workers.

Eight types of activities are recommended in the Universal Precautions: hand washing after patient contact; no needle recapping; establishing a system for safe disposal of sharps; use of gloves when contact is anticipated with body fluids, broken skin, or mucous membranes; use of mask, eye protection, and gown if splashing with body fluids is anticipated; covering broken skin; cleaning up spills of body fluids; and establishing a system for disposal of hospital wastes. These are called universal because they should be practiced in all health care circumstances, without making a judgment about whether a particular individual or blood sample might be infected.

—*Sarah Boslaugh*

See also Epidemiology in Developing Countries; Hepatitis; HIV/AIDS; Sexually Transmitted Diseases

Further Readings

Kane, A., Lloyd, J., Zaffran, M., Simonsen, L., & Kane, M. (1999). Transmission of hepatitis B, hepatitis C and human immunodeficiency viruses through unsafe injections in the developing world: Model-based regional estimates. *Bulletin of the World Health Organization*, *77*(10), 801–807.

World Health Organization. (2003). *Aide-Memoire for a strategy to protect health workers from infection with bloodborne viruses.* Retrieved February 28, 2007, from http://www.who.int/injection_safety/toolbox/en/AM_HCW_Safety_EN.pdf.

Web Sites

Centers for Disease Control and Prevention, HIV/AIDS Web page: http://www.cdc.gov/hiv.

Centers for Disease Control and Prevention, Viral Hepatitis Web page: http://www.cdc.gov/ncidod/diseases/hepatitis/index.htm.

Body Mass Index (BMI)

Body mass index (BMI) is a calculated measurement of a person's height and weight for the purpose of classifying individuals as underweight, normal weight, overweight, or obese. BMI is sometimes referred to as the *Quetelet Index*, named after Lambert

Adolphe Jacques Quítelet (1796–1874), the Belgian sociologist and statistician who developed it. Quítelet, who wrote a book about the "average man," was interested in developing a simple method for classifying an individual's weight scaled according to height. His Quetelet Index, which later became the modern BMI measurement, is the most recognized calculation for obesity statistics. The BMI calculation is frequently used by national and international organizations in health policy discussions concerning obesity-related issues.

Reliability

BMI is generally used as an indicator of body adiposity, although it does not measure body fat directly. However, research has demonstrated the BMI to be a reliable indicator of body fatness when correlated with direct measurements such as underwater weighing, skin fold thickness measurements, computed tomography (CT), and dual-energy X-ray absorptiometry (DEXA). The BMI is a simple, inexpensive method for estimating body adiposity and is a more practical method for health care practitioners and the general public to use.

Calculation

BMI can be calculated using metric units or by an adapted version with imperial units. The same formula is used for both adults and children.

BMI Calculated With Metric Units

Formula: weight (kg)/[height (m)]^2

Calculation: [weight (kg)/height (m)/height (m)]

BMI Calculated With Imperial Units

Formula: weight (lb)/[height (in.)]$^2 \times$ 703

Calculation: [weight (lb)/height (in.)/height (in.)] $\times$ 703

Interpretation

Although the BMI calculation for adults and children is the same, the interpretation of the calculation is different for each age group. Adults, 20 years of age or older, use the standard weight categories for all ages and both genders. *Underweight* adults have a BMI of less than 18.5; *healthy weight* adults have a BMI of at least 18.5 but less than 25; *overweight but not obese* adults have a BMI of at least 25 and less than 30; and *obese* adults have a BMI of 30 or more. BMI and weight status categories can vary slightly depending on the reporting agency.

BMI interpretation for children and adolescents are age and gender specific and use the BMI percentile. The calculated BMI for children and adolescents is plotted on the BMI-for-age growth chart to obtain the BMI percentile. Percentiles are commonly used in the United States to assess the growth of an individual child as compared with children of the same gender and age. Gender and age are significant when assessing a child's growth pattern because the amount of body fat changes with age and the amount of body fat differs between genders. Weight status based on the BMI percentile for children and adolescents are as follows: *underweight* children or teens with a BMI-for-age that is less than the 5th percentile; *normal weight* children or teens with a BMI-for-age that is at least the 5th percentile but less than the 85th percentile; *at risk for overweight* children or teens with a BMI-for-age that is at least the 85th percentile but less than the 95th percentile; and *overweight* children or teens with a BMI-for-age greater than the 95th percentile.

Clinical Uses

BMI is used as a screening tool to compare an individual's weight status with that of the general population. It is also used to identify possible weight problems in both adults and children. The BMI ranges for the adult population are based on the relationship between weight and morbidity and mortality. Overweight and obese adults are at an increased risk for a variety of diseases and health conditions such as coronary artery disease, hypertension, type 2 diabetes, dyslipidemia, stroke, gallbladder disease, and osteoarthritis.

The BMI ranges for children aged 2 to 19 are based on the relationship of weight and risk for developing weight-related health problems. Children with a BMI-for-age percentile in the overweight range are at risk for developing hypertension, hyperlipidemia, hyperglycemia, sleep apnea, and other respiratory disorders. More than 50% of all overweight children have at least one cardiac risk factor, and 25% of overweight children have two or more cardiac risk factors.

Other developmental and social problems include low self-esteem, social discrimination, and poor performance in school. Furthermore, overweight adolescents are more likely to become overweight or obese adults with various chronic diseases, including cardiovascular disease and type 2 diabetes.

It is important to remember that the BMI calculation is not a diagnostic tool. A child with a BMI-for-age in the 85th percentile and higher or an adult with a BMI greater than 25 should be assessed further to determine potential health risks. A detailed health history, family history, diet evaluation, and exercise assessment are warranted. A direct method of adiposity measurement, that is, skin fold thickness measurements or underwater weighing, would provide an accurate and reliable indicator of actual body adiposity. Appropriate therapeutic interventions can be made based on the individual's comprehensive assessment.

Limitations

Although the correlation between the adult BMI and estimated body adiposity is relatively strong, there are some variations based on sex and age. For example, older adults tend to have more body fat than younger adults with the same BMI, and women tend to have more body fat than men with the same BMI. Furthermore, highly trained athletes will have a higher BMI that reflects an increase in muscularity and not in true body adiposity. These individuals should be assessed using other direct methods, as previously noted, for measuring their body fatness. The limitations mentioned above should be taken into account when using the BMI calculation under these circumstances.

—Darlene McPherson

See also Obesity

Further Readings

Calle, E. E., Thun, M. J., Petrelli, J. M., Rodriguez, C., & Heath, C. W. (1999). Body-mass index and mortality in a prospective cohort of U.S. adults. *New England Journal of Medicine*, *341*(15), 1097–1105.

National Task Force on the Prevention and Treatment of Obesity. (2000). Overweight, obesity, and health risk. *Archives of Internal Medicine*, *160*, 898–904.

Schoenborn, C. A., Adams, P. F., & Barnes, P. M. (2002). Body weight status of adults: United States, 1997–98. *Advance Data*, *330*, 1–16.

Box-and-Whisker Plot

The box-and-whisker plot, also called a boxplot, was invented by John Tukey. It is a graph of a data set that consists of a line extending from the minimum value to the maximum value, and a box with lines drawn at the first quartile Q_1; the median, the second quartile; and the third quartile Q_3, with outliers plotted as individual data points. It is useful for revealing the central tendency and variability of a data set, the distribution (particularly symmetry or skewness) of the data, and the presence of outliers. It is also a powerful graphical technique for comparing samples from two or more different treatments or populations.

Although boxplots are usually generated using statistical software, they also may be constructed by hand, using the following steps:

1. Draw a rectangular box whose left edge is at Q_1 and the right edge is at Q_3. The box width is therefore the interquartile range $IQR = Q_3 - Q_1$. Draw a vertical line segment inside the box at the median.
2. Place marks at distances 1.5 times the *IQR* from either end of the box: These are the inner fences. Similarly, place marks for the outer fences at distances 3 times *IQR* from either end.
3. Extend a horizontal line segment (''whiskers'') from each end of the box out to the most extreme observations that are still within the inner fences.
4. Represent values for mild outliers or observations between the inner and outer fences by circles. Represent values for extreme outliers or observations beyond the outer fences by asterisks.

The median is the middle value in the ordered data list. It is the number that divides the bottom 50% of the data from the top 50%. The median is also the second quartile Q_2. Use the following steps to find the median of a data set:

1. Arrange the data from smallest to largest.
2. If the number of observations is odd, then the median is the observation exactly in the middle of the ordered list.
3. If the number of observations is even, then the median is the average of the two middle observations in the ordered list.

The first quartile Q_1 is the median of the lower half of the ordered data, and the third quartile Q_3 is the median of the upper half of the ordered data. If the number of observations is odd, the median of the entire data is included in both halves.

Example

Biological disturbances that are closely associated in adults suffering from endogenous depression (depression with no obvious external cause) are cortisol hypersecretion and shortened rapid eye movement (REM) period latency (the elapsed time from sleep onset to the first REM period). In a paper titled "Plasma cortisol secretion and REM period latency in adult endogenous depression," Gregory Asnis and colleagues reported on a comparison of REM period latency for patients with hypersecretion and patients with normal secretion. The data values are given below.

Hypersecretion Sample (n = 8)

0.5, 1.0, 2.4, 5, 15, 19, 48, 83

minimum = 0.5

maximum = 83

median = 10

$Q_1 = 1.7$

$Q_3 = 33.5$

$IQR = 31.8$

$1.5(IQR) = 47.7$

Normal Secretion Sample (n = 17)

5, 5.5, 6.7, 13.5, 31, 40, 47, 47, 59, 62, 68, 72, 78, 84, 89, 105, 180

minimum = 5

maximum = 180

median = 59

$Q_1 = 31$

$Q_3 = 78$

$IQR = 47$

$1.5(IQR) = 70.5$

Figure 1, the boxplot representing these data, displays several interesting features. Each sample has a mild outlier and an upper tail rather longer than the corresponding lower tail. Normal secretion REM period latency values appear to be substantially higher than those for hypersecretion; this was confirmed by a formal analysis.

—Renjin Tu

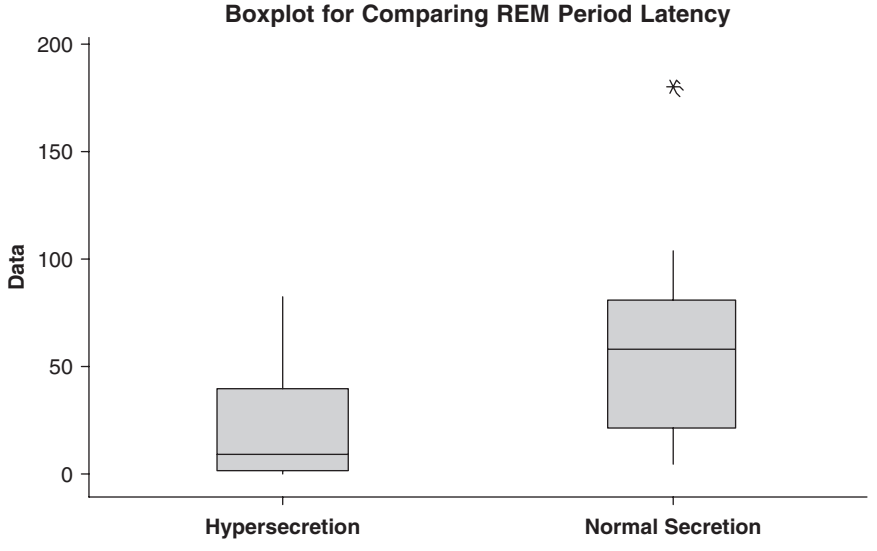


Figure 1 REM Period Latency for Patients With and Without Hypersecretion

See also Bar Chart; Histogram; Stem-and-Leaf Plot; Tukey, John

Further Readings

Asnis, G. M., Halbreich, U., Sachar, E. J., Nathan, R. S., Ostrow, L. C., Novacenko, H., et al. (1983). Plasma cortisol secretion and REM period latency in adult endogenous depression. *American Journal of Psychiatry*, *140*(6), 750–753.

Tukey, J. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Weiss, N. A. (2005). *Introductory statistics* (7th ed.). Boston: Addison-Wesley.

Breastfeeding

Human milk is the appropriate nutrition for infants. Exclusive breastfeeding for the first 6 months and continued breastfeeding with appropriate introduction of solids at 6 months reduces the risk of many illnesses and chronic diseases. The *Healthy People 2010* breastfeeding goals are to increase to 75% the proportion of women who initiate breastfeeding, to 50% the proportion of women who are breastfeeding at 6 months, and to 25% the proportion of women who breastfeed for 1 year. The World Health Organization (WHO) recommends 6 months of exclusive breastfeeding, followed by the timely introduction of appropriate complementary foods, and continued breastfeeding for 2 years and beyond.

Despite documentation and public knowledge of the health, social, and economic benefits of breastfeeding and recommendations from numerous organizations, including the American Academy of Pediatrics, the American Public Health Association, and the WHO, breastfeeding incidence, exclusivity, and duration are well below the goals identified in *Healthy People 2010.* National data also reveal disparities in breastfeeding practices, primarily associated with economic and education status. The primary demographic factors associated with not breastfeeding or breastfeeding for a short duration include being nonwhite, poor, unmarried, or younger than 25; completing 12 or less years of education; and living in the southeastern United States. To improve breastfeeding practices and to reduce existing disparities in infant and young child feeding, interventions need to target not only individuals but also organizations and communities.

Defining Breastfeeding

Breastfeeding behaviors are neither clearly defined nor consistently operationalized in much of the literature. Some studies of feeding practices define breastfeeding as exclusive (infants receive only human milk, with no water, other liquids, or solids), others permit water to be included in "exclusive" breastfeeding, and others allow any amount of human milk to constitute the equivalent of "breastfeeding," even if human milk substitutes (HMS, also referred to as "formula") or other liquids and solids are also part of the child's diet. Moreover, some studies define successful initiation of breastfeeding as having "ever breastfed," even just once, while others require that breastfeeding lasts a specific number of days or weeks to meet the criteria for successful initiation, and others fail to define how "breastfeeding" is measured. Another consideration is the way in which the milk is being given, and whether the child is fed directly from the breast or fed human milk from a bottle.

Inconsistent definitions of breastfeeding make it difficult to compare and interpret study findings. If breastfeeding is not clearly defined, breastfeeding once a day could be grouped in the same category as exclusive breastfeeding, despite the very different amounts of human milk and exposure of the infants to other nutrients. Additionally, carefully identifying when and why women discontinue exclusive breastfeeding may permit development of specific interventions to prevent untimely weaning. Women who discontinue exclusive breastfeeding in the first week postpartum may have experienced feeding technique difficulties, whereas women who stop exclusively breastfeeding later may do so because they are returning to work or school and have no place to express their milk. Different strategies are needed to address each of these situations.

Risks and Costs of Not Breastfeeding

Mothers and children who do not breastfeed lose the physiological, immunological, and psychological benefits that breastfeeding confers and face increased risk for a number of acute and chronic diseases. Women who do not breastfeed may experience more postpartum bleeding, increased risk of breast and ovarian cancers, as well as increased risk of osteoporosis and rheumatoid arthritis, short intervals between births, and longer time to return to prepregnancy weight than women who breastfeed. Child

health risks associated with not breastfeeding or not receiving human milk include, but are not limited to, increased morbidity from gastrointestinal, respiratory, and middle-ear infections, more atopic illness and allergic disease, as well as increased risk of childhood obesity and type 1 and type 2 diabetes.

In addition to health benefits, breastfeeding has economic benefits. A 1997 study estimated that HMS provided by the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) cost more than $2.6 million annually. Furthermore, the cost of 1 year of HMS for families not enrolled in WIC is estimated to be as much as $1,500. Apart from savings resulting solely from not purchasing HMS, breastfed infants generally have fewer illnesses and lower health care costs. In a study of medical care costs incurred in the first 12 months of life, expenditures for infants not breastfed were estimated to be $200 higher than for breastfed infants, part of which may be attributable to significantly lower incidence of otitis media (ear infection) among breastfed infants. By increasing U.S. breastfeeding rates to the *Healthy People 2010* goal of 50% breastfeeding at 6 months, estimated national savings of more than $3.6 billion would be realized from reduced costs related to hospital care, parents' lost wages, and premature deaths, just considering otitis media ($3.6 million), gastroenteritis ($9.9 million), and necrotizing enterocolitis ($3.2 *billion*).

Promoting, Protecting, and Supporting Breastfeeding

To effectively promote, protect, and support breastfeeding, action at multiple levels is required at several levels. The WHO developed the International Code of Marketing of Breast-Milk Substitutes (the Code), which prohibits advertising of HMS to the public, and requires that only accurate, scientific information about HMS be given to health care providers. The WHO/UNICEF Ten Steps to Successful Breastfeeding state that all mothers should have access to skilled support to initiate and continue breastfeeding; health care providers must be adequately trained to provide clinical care; and continued health care provider support should be augmented with trained community lay or peer counselors. The Ten Steps, as listed on the UNICEF Web site describing the Baby Friendly Hospital, Initiative include the following:

1. Have a written breastfeeding policy that is regularly communicated to all health care staff.
2. Train all staff in skills necessary to implement this policy.
3. Inform all pregnant women about the benefits and management of breastfeeding.
4. Help mothers initiate breastfeeding within half an hour of birth (in the United States this timeframe is expanded to 1 hr).
5. Show mothers how to breastfeed and how to maintain lactation, even if they should be separated from their infants.
6. Give newborn infants no food or drink other than breast milk, unless medically indicated.
7. Practice rooming-in, which allows mothers and infants to remain together 24 hr a day.
8. Encourage breastfeeding on demand.
9. Give no artificial teats (bottle nipples) or pacifiers (also called dummies or soothers) to breastfeeding infants.
10. Foster the establishment of breastfeeding support groups and refer mothers to them on discharge from the hospital or clinic.

Government can monitor and enforce industry compliance with the Code, mandate paid maternity leave, and protect the breastfeeding rights of working women through legislation or regulation. The media can portray breastfeeding as the norm.

Breastfeeding Practices in the United States

Data from the Ross Laboratories Mothers Survey indicate that the proportion of women who initiate breastfeeding in the United States has increased steadily, from a low of 24.7% in 1971 to 70.1% in 2002. Data from the 2003 National Immunization Survey indicate similar findings (70.3%). The percentage of WIC-eligible women who initiate breastfeeding is substantially lower than that of the general population. In 2002, 58.8% of WIC mothers initiated breastfeeding compared with 79.2% of non-WIC mothers. Disparities are also evident in duration data. Breastfeeding promotion is a mandated part of WIC programming, but researchers suggest that WIC sends mixed messages by providing vouchers for HMS,

thereby appearing to endorse it as an equivalent or acceptable form of infant nutrition.

Influences on Infant and Young Child Feeding Practices

Interactions with health care providers have been identified as an important influence on women's infant feeding decisions and practices. While a variety of personnel can provide effective breastfeeding support, International Board Certified Lactation Consultants have specific education and training to help mothers establish lactation and continue breastfeeding. Mother-to-mother support such as that offered in groups such as La Leche League International, Nursing Mothers Counsels, and lay or peer counselors may help women sustain breastfeeding beyond the early postpartum period.

Hospital policies and procedures developed around an HMS-feeding paradigm can be detrimental to breastfeeding. For example, early separation of mother and newborn, ''test-feeding'' using a bottle of sterile water, and imposing a schedule on the newborn (vs. feeding on cue) are some of the traditional hospital policies that thwart mothers' efforts to initiate and exclusively breastfeed. Not only WIC's distribution of vouchers for HMS but also samples distributed by hospitals and physicians' offices and WIC, as well as use of items such as pens with HMS manufacturers' logos, act as a silent, yet powerful, endorsement of the brand as well as the product.

A mother's childbirth experience can also be potentially harmful to breastfeeding. Some studies have found that women who give birth by cesarean section are less likely to successfully initiate and continue breastfeeding compared with women who have vaginal births, possibly due to surgery-related delays in the time between birth and mother-child skin-to-skin contact and delays in first breastfeeding.

—Deborah L. Dee and Mary Tully

See also Child and Adolescent Health; Healthy People 2010; National Immunization Survey; United Nations Children's Fund; World Health Organization

Further Readings

International Lactation Consultant Association. (2005). *Clinical guidelines for the establishment of exclusive breastfeeding.* Raleigh, NC: Author.

U.S. Department of Health and Human Services, Office on Women's Health. (2000). *HHS blueprint for action on breastfeeding*. Washington, DC: Author.

U.S. Department of Health and Human Services. (2000). *Healthy people 2010: Understanding and improving health* (2nd ed.). Washington, DC: Government Printing Office.

World Health Organization. (1981). *International code of marketing of breast-milk substitutes*. Retrieved March 2, 2007, from http://www.who.int/nutrition/publications/code_english.pdf.

World Health Organization and United Nation's Children's Fund. (2002). *Global strategy on infant and young child feeding*. 55th World Health Assembly. Geneva, Switzerland: Author.

Web Sites

International Lactation Consultant Association: http://www.ilca.org.

UNICEF: http://www.unicef.org.

United States Breastfeeding Committee: http://usbreastfeeding.org/breastfeeding/compend-babyfriendlywho.htm.

Budd, William
(1811–1880)

Dr. William Budd is considered a pioneer in the development of the germ theory of disease and waterborne transmission. He is best known for identifying water as the source of transmission in typhoid fever.

Budd was born in Devon, England, into a medical family. His father was a physician, and 6 of his 10 brothers studied medicine. After initially completing an apprenticeship with his father, he spent 4 years training in Paris, where he was a student of Pierre Charles Alexandre Louis, who is often referred to as the ''father of epidemiology.'' In 1841, Budd settled in Bristol, England, where he worked as a physician at St. Peter's Hospital and the Bristol Royal Infirmary. It was during his time in Bristol that Budd developed his theory regarding the transmission of typhoid fever.

In 1853, Budd recorded an outbreak of typhoid fever in the Welsh town of Cowbridge. Local celebrations during this time involved two parties on successive nights at a town inn. Eight of those who attended the parties died of typhoid fever. Budd noticed the close proximity of a local well that was located next to the septic tank of the inn. Given this, he suggested that water may have been the source of the infection. This theory was further developed after

noting that a person recovering from typhoid fever had left the inn before the parties began and also that all the eight individuals who became ill had had the same lemonade at the party, made with water from the well.

This theory was later reinforced in 1866 when Budd and a colleague traced a similar outbreak in a group of farm cottages. A father from one of the cottages had become infected with typhoid fever from elsewhere and then returned home to one of the cottages. Several days later, people in the neighboring cottages also became ill with typhoid fever. Budd noted that the drains of the cottages with infected people were linked to the same stream and that those who became infected lived downstream from the original outbreak. Budd concluded that water had been the source of transmission of infection.

In addition to suggesting typhoid was waterborne, Budd also argued that the mode of transmission was fecal-oral. Given this, Budd suggested that poor hygiene and living conditions contributed to its spread and recommended improved sanitary measures, including hand washing and boiling water, to slow and prevent transmission. It was thought that this application of preventive measures helped reduce the spread of cholera in Bristol during this time. In this way, Budd was a great contributor to the public health sanitation movement.

In 1873, Budd's classic paper on typhoid fever, "Typhoid Fever: Its Nature, Mode of Spreading, and Prevention," was published. Although his primary research focused on typhoid fever, he also suggested, along with Dr. John Snow, that cholera was a waterborne disease. He died in 1880, the same year that the typhoid bacillus, *Salmonella typhi*, was isolated.

—Kate Bassil

See also Public Health, History of; Snow, John; Waterborne Diseases

Further Readings

Moorhead, R. (2002). William Budd and typhoid fever. *Journal of the Royal Society of Medicine*, *95*(11), 561–564.

C

Cancer

Cancer occurs when abnormal cells grow out of control. Normal cells in the body grow, divide, and die, and as we become older, our cells divide at a decreasing rate. Cancer cells, however, continue to grow and divide unchecked by the body, and they outlive normal cells. Cancers are capable of both invasion, or spreading into adjacent tissue, and metastasis, or traveling and settling into new, noncontiguous parts of the body. Cancer cells commonly form a tumor, or mass of cells, but they may also circulate in the bloodstream. *Carcinogenesis* describes the transformation of normal cells to neoplastic cells or abnormal cells that grow uncontrollably. *Neoplasia* is the process by which neoplasms develop from normal tissue. Cancers generally develop from a single neoplastic cell, commonly referred to as clonal expansion. *Dysplasia* refers to the early stages of neoplasia in which clonal expansion of abnormal cells occur.

Cancers are usually named for the site in which the neoplastic cells originated, even in cases of metastasis to other organs. Cancers are also usually named for their histology or morphologic features. *Carcinomas* are cancers of the epithelium and consist of cells found on external surfaces, internal cavity linings, and glandular linings, with cancers named *squamous cell carcinoma, transitional cell carcinoma*, and *adenocarcinoma*, respectively. *Sarcomas* are cancers of the mesenchyme, from which supporting tissue, including connective tissue, bone, cartilage, and muscle, is derived, and are often used with a prefix denoting the tissue of origin. Some cancers are given eponymous names, including Hodgkin lymphoma and Burkitt lymphoma.

Multistage Process of Carcinogenesis

Carcinogenesis involves two major phases, initiation and promotion. Initiation is the first stage and usually involves the interaction between a carcinogen (cancer-causing agent) and DNA that is permanent and irreversible. The promotion phase, on the other hand, is generally reversible and instable; it progresses when the abnormal cell is stimulated to grow and divide. Initiation and promotion generally occur as a multistage process that involves a series of genetic and molecular alterations and events.

Cancers arise as a result of damage to the DNA—mutations of critical genes involved in the regulation of cell growth and division. Often, the body has the capability of repairing DNA; the process of carcinogenesis begins when the body is unable to repair the damage. Two major classes of genes involved in carcinogenesis are oncogenes and tumor suppressor genes. Oncogenes generally have important roles in cell growth and differentiation. When a single copy of an oncogene is altered, leading to inappropriate or overexpression, neoplasia may result. The function of tumor suppressor genes is to prevent the development of cancers, generally by maintaining the integrity of the genome through cell cycle control and apoptosis, or signaled cell death; carcinogenesis, however, may occur on the loss or damage of both copies of the genes. A cancer develops generally when several mutations have occurred in the same

cell. In addition to being caused by exposure to carcinogens, some critical mutations may be inherited, giving the individual a genetic predisposition for cancer.

Staging Classification

Tumors are classified by stage, which defines the extent to which a particular cancer has grown and spread. Staging is useful in indicating the potential prognosis of an individual case of cancer and helps in the selection of the most appropriate treatment methods. Many different staging methods are in use to describe tumors, but the one most commonly used is the TNM system. This system assesses tumor size (T), involvement of lymph nodes (N), and distant metastasis (M). Tumors are then classified into Stage I (early-stage tumor), II, III, or IV (advanced tumor) based on the TNM description. Tumors may also be described as in situ (no invasion), localized, invasive, or distant (distant metastasis present).

Survival

Survival is the time between cancer diagnosis and death. It is commonly expressed as a 5-year survival, which is the percentage of patients alive after 5 years of follow-up from the diagnosis date. Relative survival takes into account the background risk of dying among those without cancer given the same age, sex, and race. Survival varies greatly depending on the type of cancer and the stage of the cancer at diagnosis. Survival has been generally increasing in the United States. Increases in survival reflect detection and diagnosis of cancers at earlier stages and improvements in treatment methods.

Treatments

Different ways in which cancer may be treated include surgery, radiation, chemotherapy, and biological therapies. Generally, surgery and radiation are used for localized cancers, whereas chemotherapy and biological therapy are used for metastatic cancers. Chemotherapy involves anticancer drugs that are intended to target cancer cells. Biological therapy, also known as immunotherapy or biotherapy, is based on the idea of using the immune system to control cancer cells. Often, combinations of these methods are used to treat cancer. For example, some chemotherapies or radiation may also damage normal cells, and biological therapies may be used in conjunction to help repair normal cells damaged by these other methods.

Current Cancer Burden

The Global Impact

As the control of infectious diseases in the world improves, the burden, or impact, of chronic diseases, such as cancer, is expected to rise. According to GLOBOCAN 2002, a global cancer database produced by the International Agency for Research on Cancer (IARC), an estimated 11 million new cancers and 7 million cancer deaths occurred in the world in 2002. The most common new cancers for both sexes combined were of the lung (12% of all cancers), breast (11%), colon/rectum (9%), and stomach (9%), and the most common cause of death from cancer were of the lung (18%), stomach (10%), liver (9%), and colon/rectum (8%). Among males, the most new cases of cancer were of the lung (17%), prostate (12%), stomach (10%), and colon/rectum (9%) in terms of new cases; the top cancers in terms of new deaths among males were lung (22%), stomach (12%), liver (11%), and colon/rectum (7%). Among females, the most common new cancers were breast (23%), cervix (10%), colon/rectum (9%), and lung (8%); the most common causes of new deaths among women were breast (14%), lung (11%), cervix (9%), and stomach (9%) cancers.

Cancer in the United States

According to the American Cancer Society (ACS) *Cancer Facts & Figures 2002* report, the United States was estimated to have nearly 1.3 million new cancer cases and more than 555,000 deaths from cancer in 2002. Cancer risk is strongly associated with age; 77% of cases were diagnosed among people aged 55 years or older. The lifetime risk of developing cancer among men and women was 1 in 2 and 1 in 3, respectively. Cancer was the second leading cause of deaths after heart disease. Based on the ACS report, the most commonly occurring new cancers in the United States were breast (16%), prostate (15%), lung (13%), and colon/rectum (12%). The top cancer killers included lung (28%), colon/rectum (10%), breast (7%), and prostate (5%). As expected, rankings of the cancers in males were

different from the rankings in females. The most common new cancers among men were prostate (30%), lung (14%), colon/rectum (11%), and urinary bladder (7%), whereas among women, the top new cancers were breast (31%), lung (12%), colon/rectum (12%), and uterine corpus (6%). The most common causes of deaths from cancer among men were lung (31%), prostate (11%), colon/rectum (10%), and pancreas (5%) and among women were lung (25%), breast (15%), colon/rectum (11%), and pancreas (6%).

Specific Cancers

Worldwide data on the following cancers are based on the most recent information available from GLOBOCAN 2002. Current trends may differ from these 2002 estimates by varying degrees depending on changes in risk or prevention strategies (incidence) and improvements in cancer detection or treatment (mortality). Incidence and mortality statistics of the United States are also based on 2002 estimates to allow for comparability with the worldwide data from GLOBOCAN 2002. More recent estimates for the United States are available at the ACS Web site. Survival estimates for the United States are based on data from the ACS *Cancer Facts & Figures 2006* report.

Breast Cancer

Worldwide, 1.15 million females were estimated to be newly diagnosed in 2002 with invasive and in situ breast cancer, and more than 411,000 females were expected to die annually from breast cancer, making it the cancer with the most new cases and deaths among females. The age-standardized incidence and mortality rates of invasive and in situ breast cancer among females were 37.5 and 13.2 per 100,000 females, respectively. The high ratio of incidence to mortality indicates an overall good prognosis throughout the world. Areas of high risk are usually the more developed parts of the world, including the United States, Canada, parts of Europe, and Australia and New Zealand.

In the United States, 203,500 females were diagnosed with invasive breast cancer and an additional 54,300 were diagnosed with in situ breast cancer in 2002. The age-standardized incidence rate of invasive and in situ breast cancer in the United States was 101.1 per 100,000 females, more than 2.7 times that of the world. Forty thousand women were estimated to die from breast cancer annually, with an age-standardized mortality rate of 19.0 per 100,000. In the United States, the average 5-year survival for breast cancer is roughly 88%. Survival, however, is very much dependent on tumor stage. Localized cancers are associated with a 98% survival, whereas cancers with distant metastasis are associated with a 26% survival.

Risk Factors

The most important risk factors for breast cancer are those that affect reproductive and hormonal patterns. For example, factors associated with increased levels of endogenous estrogens, such as early menarche, late age at first birth, low parity, and late menopause, increase risk for breast cancer. Obesity and alcohol consumption also increase risk. Genetic or familial factors also affect risk: Women with a family history of breast cancer have an increased risk, and when mutations to the *BRCA1* gene are involved, risk is very high and the cancers occur early in life.

Prospects for Prevention

At present, the most practical approach to improving the burden of breast cancer is by decreasing the mortality rate through early detection by screening. The ACS recommends regular annual mammographic screening and clinical breast examinations (CBE), as well as monthly breast self-examinations (SBE), for women aged 40 years and older. Women aged 20 to 39 years are recommended to have a CBE every 3 years and monthly SBE. Although recent evidence indicates that magnetic resonance imaging (MRI) may be more sensitive than mammography in detecting tumors in women with an inherited breast cancer susceptibility, mammography, CBE, and SBE are still part of the standard recommended screening guidelines according to the ACS, especially for women of the general population.

Prostate Cancer

Worldwide in 2002, 679,000 males were estimated to be newly diagnosed with prostate cancer, and 221,000 were estimated to die from the cancer. The age-standardized incidence and mortality rates were 25.3 and 8.2 per 100,000 males, respectively. Areas of high incidence include North and South

America, most of Europe, Australia, and New Zealand, whereas areas of low incidence include Asia and North Africa.

In the United States, 189,000 males were estimated to be newly diagnosed with prostate cancer in 2002. The age-standardized incidence rate of prostate cancer in the United States was 124.8 per 100,000 males, nearly five times the world average. This difference has been in large part attributed to the use of prostate-specific antigen (PSA) blood testing, which is more common in the United States than in other countries in the world; PSA blood testing is effective in detecting latent cancers, and most prostate cancers in the United States are detected at that stage—the great majority of which would never progress to invasive cancer. The age-standardized mortality rate was 15.8 per 100,000 with 30,200 estimated annual deaths from prostate cancer. The incidence among white, African American, and Asian males was 169.0, 272.0, and 101.4, respectively, demonstrating a racial component in conferring risk. In the United States, the survival rate for all prostate cancers is nearly 100%. Separating by stage, localized cancers are associated with a 100% survival, whereas cancers with distant metastasis are associated with a 34% survival.

Risk Factors

Age, ethnicity, and family history are the most established risk factors. In 2002, more than 70% of all prostate cancers in the world were estimated to have been diagnosed in men aged 65 and older. As previously mentioned, African American men have the highest incidence in the United States. Five percent to 10% percent of prostate cancers may be attributed a strong family history of the cancer. Although prostate cancer has been extensively studied, the environmental risk factors for it remain unclear. Thus far, diets high in fat, meat, and dairy products have been associated with increased risk.

Prospects for Prevention

Regular exercise; a high-vegetable diet that is low in fat, red meat, and dairy products; and avoidance of obesity are recommended. Additionally, the ACS recommends that the PSA blood test, as well as the digital rectal examination, should at least be offered annually to men aged 50 or older. High-risk males, including African Americans and those with strong family histories of prostate cancer, should be offered the tests annually at age 45 and older. Men being tested, however, should be informed of the benefits as well as the limitations, with overtreatment or unnecessary treatment a likely consequence of a positive PSA test result.

Lung Cancer

The world was estimated to have more than 1.35 million new cases (965,400 males, 386,800 females) and 1.18 million deaths (848,300 males, 330,700 females) annually from lung cancer in 2002. Lung cancer was the top-ranked cancer in both new cases and deaths among males. The age-standardized incidence rates among males and females were 35.5 and 12.1 per 100,000, respectively, and the age-standardized mortality rates among males and females were 31.2 and 10.3 per 100,000, respectively. A ratio of incidence to mortality close to 1 indicates an overall poor prognosis throughout the world. Areas of high risk among males include the United States, Canada, most of Europe, and parts of East Asia. High-risk areas for females include the United States, Canada, parts of Europe, Australia, New Zealand, and China.

In the United States, 169,400 new cases (90,200 males, 79,200 females) and 154,900 deaths (89,200 males, 65,700 females) from lung cancer were estimated to have occurred in 2002. The age-standardized incidence rates of lung cancer among males and females in the United States were 61.9 and 36.1 per 100,000, respectively, and the mortality rates were 48.7 and 26.8, respectively. Survival of lung cancer is among the lowest of all types of cancers. In the United States, the survival rate of all lung cancers is nearly 15%. Separating by stage, localized cancers are associated with a 50% survival, whereas cancers with distant metastasis are associated with a 2% survival.

Risk Factors

The most important risk factor for lung cancer is clearly tobacco smoking. Passive exposure to environmental tobacco smoke (ETS) is believed to increase risk among nonsmokers. Other risk factors include occupational exposure to harmful materials, such as asbestos and rubber exposure, and air pollution. Diets high in vegetables, especially green vegetables and carrots, and fruits may decrease risk.

Prospects for Prevention

The promotion of smoking cessation and programs oriented to persuade adolescents not to start smoking are the most cost-effective campaign against lung and other smoking-related cancers and diseases. Social pressure to make smoking socially unappealing and legislation to make smoking financially less accessible are important measures for prevention. These preventive measures have had great success in the United States.

Early detection using sputum cytology and chest radiographs have shown no favorable impact on mortality and, therefore, are not recommended for screening of lung cancer. Spiral computerized topography and detection of molecular markers, such as p53 mutations, in sputum, however, have received increasing interest and are being evaluated as possible screening methods.

Cancer of the Colon or Rectum

More than 1.02 million new cases of colorectal cancer (550,500 males, 472,700 females) and 529,000 deaths (278,400 males, 250,500 females) were estimated to have occurred worldwide in 2002. The age-standardized mortality rates (10.2 and 7.6 per 100,000 for males and females, respectively) were nearly half the incidence rates (20.1 and 14.6 per 100,000 for males and females, respectively), indicating a relatively good prognosis throughout the world. Areas of high risk for both males and females include the United States, Canada, most of Europe, Australia, New Zealand, and Japan.

In the United States, 148,300 new cases (72,600 males, 75,700 females) and 56,600 deaths (27,800 males, 28,800 females) from colorectal cancer were estimated to have occurred in 2002. The age-standardized incidence rates of colorectal cancer among males and females in the United States were 44.6 and 33.1 per 100,000, respectively, and the mortality rates were 15.2 and 11.6, respectively. In the United States, the average survival for colorectal cancers is 64%. Localized colorectal cancers are associated with a 90% survival, whereas cancers with distant metastasis are associated with a 10% survival.

Risk Factors

Diets rich in vegetables and unrefined plant foods, such as cereals and legumes, protect against colorectal cancer, whereas diets rich in red meat increase risk. The mechanism by which these foods confer risk, however, remains unclear. Alcohol also increases risk for colorectal cancer. Regular physical exercise may decrease risk, and obesity may increase risk.

Prospects for Prevention

Given the dietary and behavioral risk factors for colorectal cancer, improvements in diet and frequency of exercise are a prospect for prevention. Diets rich in vegetables and unrefined plant foods and moderate in red and processed meat, as well as regular exercise and weight control, should be promoted. Another potential prevention route is screening. Colorectal cancer can be detected early, by screening using tests for occult blood in stool, sigmoidoscopy, or colonoscopy. Any of these methods can be used for early detection of colorectal adenomatous polyps, which are precursors of colorectal cancer and localized cancers.

—Binh Y. Goldstein, D. Maxwell Parkin, and Zuo-Feng Zhang

See also Cancer Registries; Carcinogen; Chronic Disease Epidemiology; Direct Standardization; Indirect Standardization; Screening

Further Readings

American Cancer Society. (2002). *Cancer facts and figures 2002.* Atlanta, GA: Author.

American Cancer Society. (2006). *Cancer facts and figures 2006.* Atlanta, GA: Author.

Parkin, D. M., Bray, F. I., & Devesa, S. S. (2001). Cancer burden in the year 2000. The global picture. *European Journal of Cancer*, *37*(Suppl. 8), S4–S66.

Parkin, D. M., Bray, F. I., Ferlay, J., & Pisani, P. (2005). Global cancer statistics, 2002. *CA: A Cancer Journal for Clinicians*, *55*, 74–108.

Web Sites

American Cancer Society: http://www.cancer.org. The *Cancer Facts and Figures* and *Statistics* links on the home page provide current statistical data on cancer rates.

Descriptive Epidemiology Production Group, International Agency for Research on Cancer: http://www-dep.iarc.fr. The *GLOBOCAN 2002* link provides additional information on GLOBOCAN 2002.

Cancer Registries

A cancer registry, also referred to as a tumor registry, is a cancer surveillance system that provides continued follow-up care on all cancer patients in a given location, hospital, or state. It is the chief method in the United States by which information is systematically collected about people diagnosed with cancer. Cancer registries document and store all significant elements of a patient's history and treatment. Depending on the resources available, the information may include basic demographic data such as age, sex, ethnicity, race, residence, and place of birth; date of diagnosis; date and cause of death; the type of cancer and its anatomical location; the extent of disease at the time of diagnosis; the types of treatment received; and the outcomes of treatment and clinical management. The information collected is then used to monitor cancer trends over time; to determine cancer patterns in various populations; to guide the planning and evaluation of cancer-control programs; to help set priorities for allocating health resources; to advance clinical, epidemiological, and health services research; and to provide information for a national database for cancer incidence.

The first modern registries for epidemiological purposes were created in the state of Connecticut in 1936. In Europe, registries were started in Denmark (1942), Belgium (1943), and the United Kingdom (1945). In 1940, New York passed the first law that required reporting of cancer cases diagnosed in the state and outside New York City to the state health department; the law was amended in 1973 to include all cancer cases in the state. Currently, most state-based cancer registries require reporting of cancer cases by state law. At present, population-based cancer registries exist in 45 states, the District of Columbia, and three U.S. territories (Puerto Rico, the Republic of Palau, and the Virgin Islands). All information reported to cancer registries is considered confidential, with strict procedures in place to protect the privacy of cancer patients.

Each time a patient is diagnosed with a new tumor, a report is sent to the state health department. When a person is diagnosed with more than one type of cancer, information is obtained for each separate tumor in a case report. Most registries include reports of all malignant cancers. Some types of cancers, including all skin cancers, and certain other types of basal cell and squamous cell carcinomas may not be reported because they are rarely fatal and usually do not require hospitalization. These data are reported to a central statewide registry from various medical facilities, including hospitals, physicians' offices, therapeutic radiation facilities, freestanding surgical centers, and pathology laboratories.

Because cancer surveillance is legally required and considered a major public health priority, cancer registry data are generally of high quality, relatively complete, and representative of the state's population. Some registries, especially in states that lack the resources to maintain them, are not timely in releasing information and may be incomplete. Most registries do not track clinical outcomes in reported cases.

Within the United States, there are multiple national organizations and programs actively collecting and reporting cancer data. These include the Centers for Disease Control's (CDC's) National Program of Cancer Registries (NPCR); the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program; the Commission on Cancer of the American College of Surgeons' National Cancer Database; the Central Brain Tumor Registry of the United States; and the CDC's National Center for Health Statistics' Vital Statistics Monthly Reports and data tapes. In addition, the American Cancer Society and two professional organizations, the National Cancer Registrars Association and the North American Association of Central Cancer Registries (NAACCR), disseminate cancer-related information to the public and to professional communities.

Many cancer registries are supported through the NPCR, which was created in 1992 when Congress passed the Cancer Registries Amendment Act. NPCR has established a standard for data collection that guides reporting in states. In addition, all cancer registries are represented in the NAACCR, which also sets standards and goals for the member registries to meet and to promote the use of cancer registry data in studies of defined populations and in cancer-control programs. To facilitate the implementation of a national standard, NAACCR instituted a program in 1997 that annually reviews member registries' abilities to produce complete, accurate, and timely data. Registries that meet the highest standards receive NAACCR certification.

One of the most heavily used cancer registries is the SEER registry, which was established in 1971.

Its goals include assembling and reporting estimates of cancer incidence and mortality in the United States, monitoring annual cancer incidence trends to identify unusual changes in specific forms of cancer occurring in population subgroups, providing recent information on changes over time in diseases, and promoting studies designed to identify factors susceptible to cancer-control interventions.

SEER collects cancer data on a routine basis from designated population-based cancer registries in 11 geographic areas of the United States, including 6 states, 1 territory, and 4 metropolitan areas. The geographic areas represent an estimated 14% of the U.S. population. Trends in cancer incidence, mortality, and patient survival as well as many other studies are derived from the SEER database. The database contains information on approximately 2.3 million in situ cancer and invasive cancers diagnosed between 1973 and 1996, with approximately 125,000 new cases added each year.

Cancer registries such as SEER are used to conduct either retrospective or cross-sectional epidemiological studies. Cancer registries have been used to assess cancer trends, to monitor the impact of cancer on the general population or a subpopulation, to identify environmental carcinogens, to monitor cancer-related effects of tobacco, to identify geographic areas with higher than average cancer rates, to study patterns and outcomes of cancer care, and to identify risk groups for research and public health interventions.

—Rita Velikina and Zuo-Feng Zhang

See also Cancer; Chronic Disease Epidemiology; Governmental Role in Public Health; Public Health Surveillance

Further Readings

Birskhead, G. C., & Maylanh, C. M. (2000). State and local public health surveillance. In S. M. Teutch & R. E. Churchill (Eds.), *Principles and practice of public health surveillance* (pp. 253–286). New York: Oxford University Press.

Cancer Registries Amendment Act, Public Law 102–515, 42 U.S.C. 280e–280e-4. (1992).

Centers for Disease Control and Prevention. (2006). *Cancer registries: The Foundation for Cancer Prevention and Control*. Atlanta, GA: Author. Retrieved April 26, 2006, from http://www.cdc.gov/cancer/npcr/npcrpdfs/about2004.pdf.

Centers for Disease Control and Prevention. (2006). *National Program of Cancer Registries: Cancer Surveillance System Rationale and Approach*. Atlanta, GA: Author. Retrieved April 26, 2006, from http://www.cdc.gov/cancer/npcr/training/css.htm.

International Agency for Research on Cancer. (1985). *The role of the registry in cancer control*. New York: Oxford University Press.

Web Sites

National Cancer Institute, Surveillance, Epidemiology, and End Results (SEER): http://www.seer.cancer.gov.

Capture-Recapture Method

The capture-recapture method is a technique for estimating the size of a population that cannot be directly measured in its entirety. It is derived from ecological research methods. To take a census of a group of animals (e.g., the population of fish in a pond), researchers capture a subset of animals, mark or tag them in some way, release them, then capture another sample (recapture). Some of the animals from the first sample will reappear in the second sample as well; some animals will appear only in one of the two captures. From this information, the size of the whole population can be estimated.

Capture-recapture methods have been adapted by epidemiologists for use in the surveillance of and identification of human illnesses. Routine surveillance methods are likely to fail to identify every affected person. Capture-recapture methods can be used to more completely identify the size of a given population. In epidemiology, the ‘‘capture’’ involves identifying affected persons from lists, registers, or other sources of information about diagnosed cases of a given condition. The presence of an individual on one of these various sources is similar to the capture of an animal by an ecologist. Use of two or more overlapping but incomplete lists of cases, multiple ‘‘captures,’’ allows for estimation of the number of cases missing from all lists, and from that an estimation of the total population size.

With two different sources, each person will appear on one, both, or neither list. The status of all cases on the lists can be summarized in a 2×2 table, where cell a represents presence on both lists, cell b represents presence on List 1 but not on List 2, cell c represents presence on List 2 but not on List 1, and cell d represents absence from both lists.

Table 1 Representation in a 2 × 2 Table of the Status of All Cases Identified From Two Independent Sources

		On List 2?	
		Yes	*No*
On List 1?	*Yes*	*a*	*b*
	No	*c*	*d*

Note: The number of cases not found on either list (cell *d*) may then be estimated from this table using the property that the cross-products (*ad* and *bc*) will be the total estimated population size.

From this, we can use the property that the product of cells *a* and *d* will equal the product of cells *b* and *c* to solve for the unknown quantity in cell *d*. Specifically, $d = bc/a$. After estimating the frequency of cell *d*, the size of the entire population of interest may be estimated by summing all four cells $a + b + c + d = \textit{total estimated population size}$. However, this method assumes that Lists 1 and 2 are independent, that is, that the presence of a case on List 1 does not influence whether or not a case is on List 2. Using only two sources, this assumption cannot be tested and estimates cannot be adjusted for possible dependency between sources.

More than two sources may be used as well. With three sources, eight estimates can be produced, accounting for all possible interactions between and dependencies among sources. Methods exist for then selecting the best single estimate from the eight possible estimates. As the number of sources increases beyond three, however, the number of possible estimates quickly becomes unwieldy, though it is possible that some smaller sources could be combined to reduce the overall number of sources used in the estimation.

Capture-recapture methods, as with other methods of estimation, are subject to uncertainty and potential bias. The single most likely source of uncertainty with this method lies in the accuracy with which records for an individual can be matched across sources. In addition, a lack of complete independence between sources may create a directional bias that will affect estimates.

—Annette L. Adams

See also Bias; Field Epidemiology; Public Health Surveillance

Further Readings

Hook, E. B, & Regal, R. R. (2004). Completeness of reporting: Capture-recapture methods in public health surveillance. In R. Brookmeyer & D. F. Stroup (Eds.), *Monitoring the health of populations: Statistical principles and methods for public health surveillance* (pp. 341–359). New York: Oxford University Press.

Carcinogen

A carcinogen is an agent, mixture, or exposure that increases the occurrence of cancer. Carcinogen identification is an activity based on the evaluation of the results of scientific research. Pertinent data for carcinogen identification include human epidemiologic studies, long-term bioassays in experimental animals, and other relevant data on toxicokinetics and cancer mechanisms. Several classification systems exist to identify the degree of carcinogenicity of agents. The most widely used system is developed by the International Agency for Research on Cancer (IARC), which is part of the World Health Organization.

Most of the existing data about whether an agent might cause cancer originate from laboratory (cell culture and animal) studies. Although it is not possible to predict with certainty which substances will cause cancer in humans based on animal studies alone, virtually all known human carcinogens that have been adequately tested have been found to produce cancer in lab animals. In many cases, carcinogens are first found to cause cancer in lab animals and are later found to cause cancer in humans. For most carcinogens, it is assumed that those that cause cancer at larger doses in animals will also cause cancer in humans.

Another source of data about carcinogens comes from epidemiologic studies, which provide evidence of a carcinogenic hazard but often are not sufficiently sensitive to identify a carcinogenic hazard except when the risk is high or when it involves an unusual form of cancer. In addition, it is difficult to single out any particular exposure as having a definite link to cancer. For these reasons, laboratory studies generally provide the best means of assessing potential risks to humans.

The IARC Classification System for Carcinogens

The most widely used system for classifying carcinogens originates from IARC. This agency releases the

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, which are scientific evaluations developed by international working groups of expert scientists. Agents are selected for evaluation on the basis of evidence of human exposure and some evidence or suspicion of carcinogenicity. The *IARC Monographs* include a critical review of the pertinent peer-reviewed scientific literature as the basis for an evaluation of the weight of the evidence that an agent may be carcinogenic to humans. Published continuously since 1972, the scope of the *IARC Monographs* has expanded beyond chemicals to include complex mixtures, occupational exposures, lifestyle factors, physical and biologic agents, and other potentially carcinogenic exposures. In the past 30 years, IARC has evaluated the cancer-causing potential of about 900 likely candidates. These evaluations provide the scientific support for public health measures implemented by many national and international health agencies around the world. IARC categorizes agents into the following five potential categories.

Group 1: Carcinogenic to Humans

If a substance is classified as belonging to Group 1, it is labeled "carcinogenic to humans." This category is used when there is sufficient evidence of carcinogenicity in humans. An agent may be placed in this category when evidence of carcinogenicity in humans is less than sufficient but there is sufficient evidence of carcinogenicity in experimental animals and strong evidence in exposed humans that the agent acts through a relevant mechanism of carcinogenicity. Only 91 agents have been characterized as "carcinogenic to humans."

Group 2A: Probably Carcinogenic to Humans

This category is used when there is limited evidence of carcinogenicity in humans but sufficient evidence of carcinogenicity in experimental animals. In some cases, an agent may be classified in this category when there is inadequate evidence of carcinogenicity in humans, sufficient evidence of carcinogenicity in experimental animals, and strong evidence that the carcinogenesis is mediated by a mechanism that also operates in humans. An agent, mixture, or exposure may be classified in this category solely on the basis of limited evidence of carcinogenicity in humans. Sixty-seven agents have been classified as "probably carcinogenic to humans."

Group 2B: Possibly Carcinogenic to Humans

This category is used for agents, mixtures, and exposures for which there is limited evidence of carcinogenicity in humans and less than sufficient evidence of carcinogenicity in experimental animals. It may also be used when there is inadequate evidence of carcinogenicity in humans but there is sufficient evidence of carcinogenicity in experimental animals. In some instances, an agent, mixture, or exposure for which there is inadequate evidence of carcinogenicity in humans but limited evidence of carcinogenicity in experimental animals together with supporting evidence from other relevant data may be placed in this group. Two hundred and forty agents have been categorized as "possibly carcinogenic to humans."

Group 3: Unclassifiable as to Carcinogenicity in Humans

This category is used most commonly for agents, mixtures, and exposures for which the evidence of carcinogenicity is inadequate in humans and inadequate or limited in experimental animals. Agents for which the evidence of carcinogenicity is inadequate in humans but sufficient in experimental animals may be placed in this category when there is strong evidence that the mechanism of carcinogenicity in experimental animals does not operate in humans.

Group 4: Probably Not Carcinogenic to Humans

This category is used for agents or mixtures for which there is evidence suggesting a lack of carcinogenicity in humans and in experimental animals. In some instances, agents or mixtures for which there is inadequate evidence of carcinogenicity in humans but evidence suggesting a lack of carcinogenicity in experimental animals, consistently and strongly supported by a broad range of other relevant data, may be categorized into this group.

Other Carcinogen Classification Systems

In addition to the IARC, other programs and agencies have implemented systems for labeling the carcinogenicity of substances. For example, the National Toxicology Program, mandated in 1978 by an act of the U.S. Congress, releases its *Report of Carcinogens* approximately every 2 years and lists agents as either

"known to be a human carcinogen" or "reasonably anticipated to be a human carcinogen."

The U.S. Environmental Protection Agency (EPA) assesses the health hazards of chemical contaminants present in the environment. These assessments cover cancer and other adverse effects. The hazard assessments are coupled with dose-response assessments that the EPA uses in its regulatory and informational programs.

The California EPA maintains a list of "chemicals known to the state to cause cancer" under the mandate created by Proposition 65, a 1986 ballot initiative enacted to protect citizens from chemicals known to cause cancer, birth defects, or other reproductive harm and to inform citizens about exposures to such chemicals. A chemical is listed if an independent committee of scientists and health professionals finds that the chemical has been clearly shown to cause cancer, if an authoritative body has identified it as causing cancer, or if a California or U.S. government agency requires that it be labeled or identified as causing cancer.

—Rita Velikina and Zuo-Feng Zhang

See also Cancer; Environmental and Occupational Epidemiology

Further Readings

American Cancer Society. (2006). *Known and probable carcinogens*. Retrieved April 26, 2006, from http://www.cancer.org/docroot/PED/content/PED_1_3x_Known_and_Probable_Carcinogens.asp.

Cogliano, V. J., Baan, R. A., Straif, K., Grosse, Y., Secretan, M. B., El Ghissassi, F., et al. (2004). The science and practice of carcinogen identification and evaluation. *Environmental Health Perspectives*, *112*(13), 1269–1274.

International Agency for Research on Cancer. (2004). *Monograph: Overall evaluations of carcinogenicity to humans: An updating of monographs 1 to 42, Supplement 7*. Lyon, France: Author. Retrieved April 26, 2006, from http://monographs.iarc.fr/ENG/Monographs/suppl7/suppl7.pdf.

Millman, H. A., & Weisburger, E. K. (Eds.). (1985). *Handbook of carcinogen testing*. Park Ridge, NJ: Noyes.

Cardiovascular Disease

Cardiovascular disease (CVD) ranks as the leading cause of morbidity and mortality among developed countries and is rapidly emerging as the predominant cause of death in many developing countries. There are numerous differences in incidence across ethnic groups and geographic regions within countries. Coronary heart disease (CHD) is the predominant manifestation of CVD and responsible for the majority of cases of CVD in developed countries. Numerous longitudinal epidemiological studies conducted in the past 60 years have provided valuable insight into the natural history and risk factors associated with the development of and prognosis associated with CVD. Randomized clinical trials have demonstrated the value of treatment of several key risk factors, most notably hypertension and hypercholesterolemia, for both the primary and the secondary prevention of CVD. This entry reviews the descriptive epidemiology of CVD, its associated risk factors, assessment of CVD risk, and the evidence behind the control of CVD risk factors for the prevention of CVD.

Definitions, Incidence, and Distribution

CVD comprises many conditions, including CHD, heart failure, rheumatic fever or rheumatic heart disease, stroke, and congenital heart disease. Of these, 7 in 10 deaths from CVD are due to CHD (53%) and stroke (17%) (Figure 1). CHD, the most common CVD condition, increases in prevalence directly with age in both men and women (Figure 2).

Myocardial infarction, angina pectoris (especially if requiring hospitalization), and sudden coronary death are the major clinical manifestations of CHD. CHD initially presents as sudden coronary death in approximately one third of the cases. Other forms of documented CHD include procedures done as a result of documented significant atherosclerosis, such as coronary artery bypass grafting (CABG) or percutaneous coronary interventions (PCI), including angioplasty and stenting. Persons with documented significant disease from a coronary angiogram, echocardiogram, nuclear myocardial perfusion, magnetic resonance imaging, or computed tomography angiographic or coronary calcium scan can also confirm the presence and extent of coronary artery disease; however, because the definitions used to define significant disease vary and/or such results do not typically result in hospitalization, these cases are not normally counted as incident or prevalent CHD, nor are they considered "hard" (usually death or hospitalized diagnosed

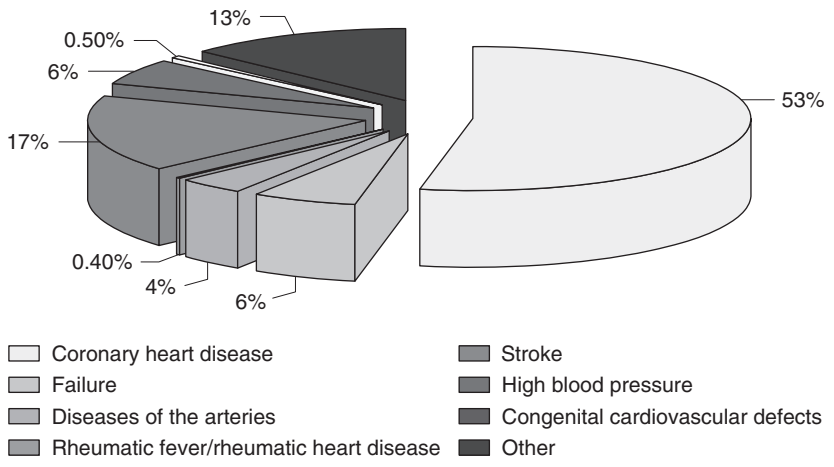


Figure 1 Percentage Breakdown of Deaths From Cardiovascular Disease

Source: Thom et al. (2006).

conditions) cardiovascular endpoints for the purposes of clinical trials. Nonfatal or fatal myocardial infarction and sudden coronary death are most typically included as "hard" CHD endpoints.

Mortality rates from CVD and CHD vary widely across different countries. From the most recent data available, among men, CVD death rates per 100,000 range from 170 in Japan to 1,167 in the Russian Federation, with the United States at 307. Corresponding CHD death rates were 53, 639, and 187, respectively. Among women, these rates were lowest in France but highest in the Russian Federation, with intermediate rates in the United States. These rates were 69, 540, and 158 for CVD, respectively, and 18, 230, and 77 for CHD, respectively. Next to the Russian Federation, Romania, Bulgaria, and Hungary had the next highest rates both for CVD and CHD. Spain, Australia, Switzerland, and France (or Japan) were the other countries with the lowest rates.

Significant variation in CVD and CHD rates also exists within the United States, with total CVD mortality highest in Mississippi at 421/100,000 and lowest in Puerto Rico at 234/100,000. CHD mortality was highest in Oklahoma at 227/100,000 and lowest in Utah at 100/100,000.

CVD and CHD prevalence rates vary dramatically by age, gender, and race. Overall CVD prevalence (including hypertension) from ages 20 to 34 to 75+ ranges from 11.2% to 77.8% in men and 6.2% to 86.4% in women, with the corresponding prevalences of CHD ranging from 0% to 16.6% in men and 0.3% to 10.3% in women. Most recent statistics show that the prevalence of CHD also varies widely by race and gender within the United States—males 8.4% overall, 8.8% in non-Hispanic white males,

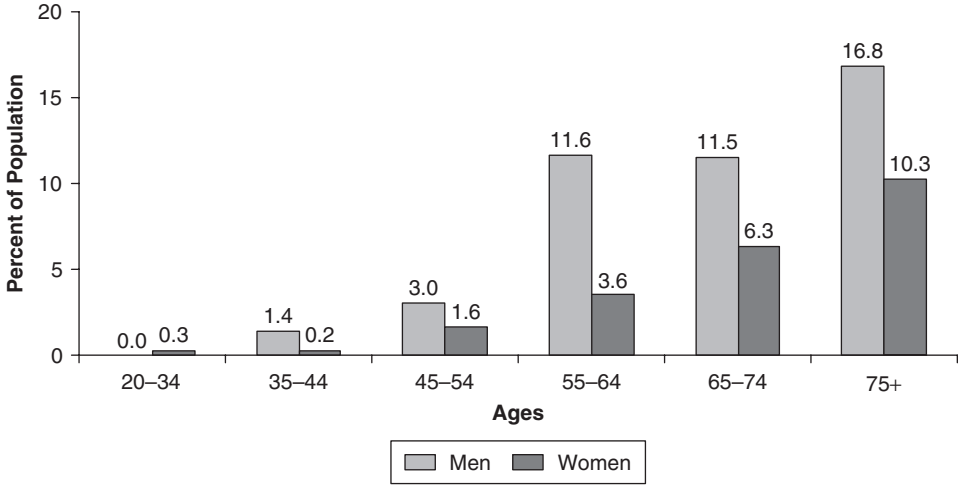


Figure 2 Prevalence of Coronary Heart Disease by Age and Sex, NHANES 1999 to 2002

Source: Thom et al. (2006).

7.4% in black males, 5.6% in Mexican American males; females 5.6% overall, 5.4% in non-Hispanic white females, 7.5% in black females, and 4.3% in Mexican American females. Prevalence rates were also noted to be 3.8% in Asians and 8.2% in American Indian and Alaska Natives. Stroke prevalence in 1999 to 2003 among U.S. adults was 3.6% in American Indians or Alaska Natives, 3.3% in blacks, 2.2% in whites, and 2.0% in Asians. For heart failure, the overall U.S. adult prevalence was 2.3% in 2003 and ranged from 1.6% in Mexican American females to 3.5% in non-Hispanic black females.

Risk Factors for Cardiovascular Disease

The Framingham Heart Study, a large longitudinal investigation of the natural history of CVD begun in 1948, originally identified many of the factors that are associated with an increased risk of CVD and originally coined the term *risk factors*. Increased awareness of major risk factors for CVD initially identified by the Framingham Heart Study and by other researchers helped lead to important public health initiatives against smoking in the 1960s, hypertension in the 1970s, and hypercholesterolemia in the 1980s. More recently, obesity and physical inactivity have also been recognized as key risk factors for CVD. Diabetes is also now widely regarded as a CHD risk equivalent, and the importance of a clustering of major cardiometabolic risk factors, commonly referred to as the metabolic syndrome, has received significant attention in the research and clinical community. Of interest is that over recent decades, while certain risk factors such as elevated cholesterol, blood pressure, and smoking have decreased in prevalence, others such as diabetes have increased (Figure 3). Risk factors often cluster together, and their co-occurrence and number of risk factors present is directly related to the incidence of CHD (Figure 4). A brief discussion of the key, traditional risk factors for CVD is provided below.

Family History

A premature family history of CHD is a well-established but unmodifiable risk factor for future CHD and can sometimes be the crucial and single most important risk factor in predisposing an individual to early CHD. A large proportion of heart attacks or strokes occurring at a young age are felt to be attributed to inherited or familial predisposition or susceptibility, and knowledge of an individual's family history can help guide preventive efforts. A premature family history is generally defined as

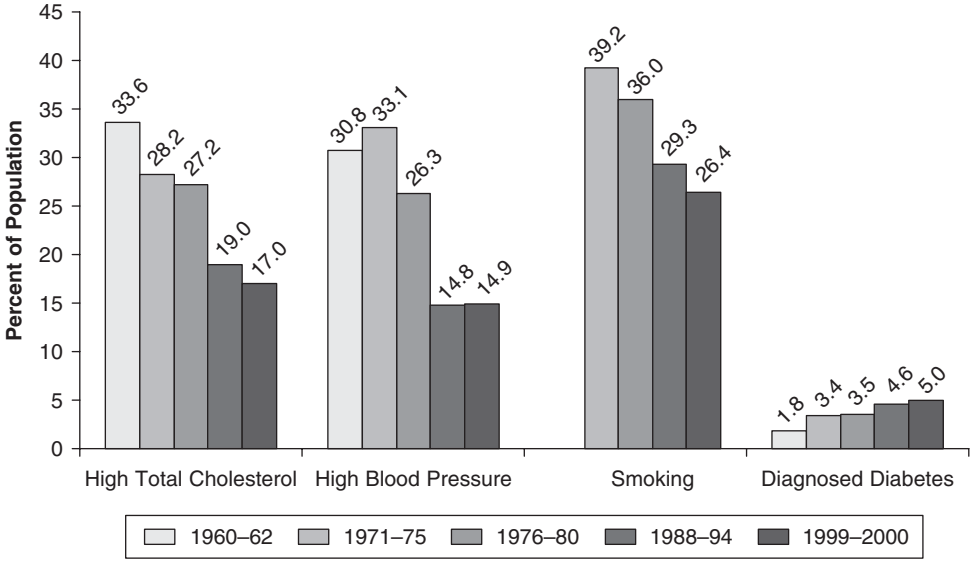
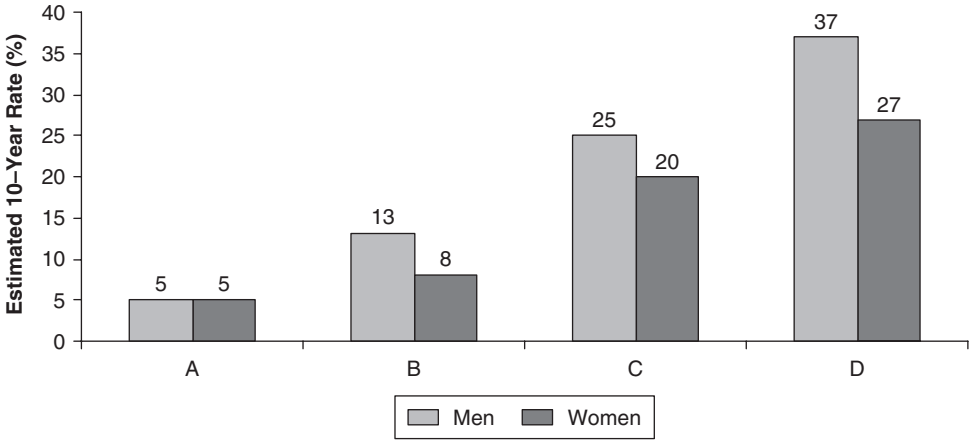


Figure 3 Trends in Cardiovascular Risk Factors in the U.S. Population, Ages 20 to 74

Source: Gregg et al. (2005).



Blood pressure (mmHg)	120/80	140/90	140/90	140/90
Total cholesterol (mg/dl)	200	240	240	240
HDL cholesterol (mg/dl)	50	50	40	40
Diabetes	No	No	Yes	Yes
Cigarettes	No	No	No	Yes

Figure 4 Estimated 10-Year CHD Risk in 55-Year-Old Adults According to Levels of Various Risk Factors: Framingham Heart Study

Source: Wilson et al. (1998).

having a male first-degree relative experiencing a first manifestation of CHD below the age of 45 or a female first-degree relative experiencing CHD below the age of 55. The number of affected relatives with premature CHD is also felt to be an important factor, whereas those with one affected relative can be shown to have a fourfold greater risk of CHD and those with two or more affected relatives may have more than a 12-fold greater risk of CHD compared with those without any affected relatives. Moreover, it has been shown that 35% of all early CHD occurs in just 3.2% of families, all of whom have a strong positive family history of CHD. Familial hypercholesterolemia, where cholesterol levels exceeding 1,000 have been commonly reported and where individuals homozygous for the defective LDL-cholesterol receptor gene have been known to have heart attacks and die by age 20, is among the most widely studied genetic conditions responsible for CHD. Other genetic defects responsible for hypertension, obesity, diabetes, and other major cardiovascular risk factors have also been identified and are the subject of major investigations.

Diabetes and Metabolic Syndrome

Diabetes mellitus is a major risk factor for CVD and is associated with a greater risk for CHD, stroke, and peripheral vascular disease. Most recent U.S. data show prevalences of approximately 10% in black and Mexican American men and as high as 12% in black women (Figure 5). More than three fourths of those with diabetes die of cardiovascular complications, most notably myocardial infarction and stroke. In appreciation of this, diabetes has been designated as a CHD risk equivalent, because the risk of CHD in those with diabetes without known heart disease has been shown to be similar to recurrent CHD events in those with CHD (but without known diabetes). Most population-based studies have shown about a twofold greater risk of CHD in men with versus without diabetes, but a three- to sevenfold greater risk of CHD in women. Diabetes is typically diagnosed by a fasting glucose > 126 mg/dl or casual glucose level of > 200 mg/dl. An abnormal glucose challenge test is also used to establish diagnosis. Clinical trial evidence to show whether controlling

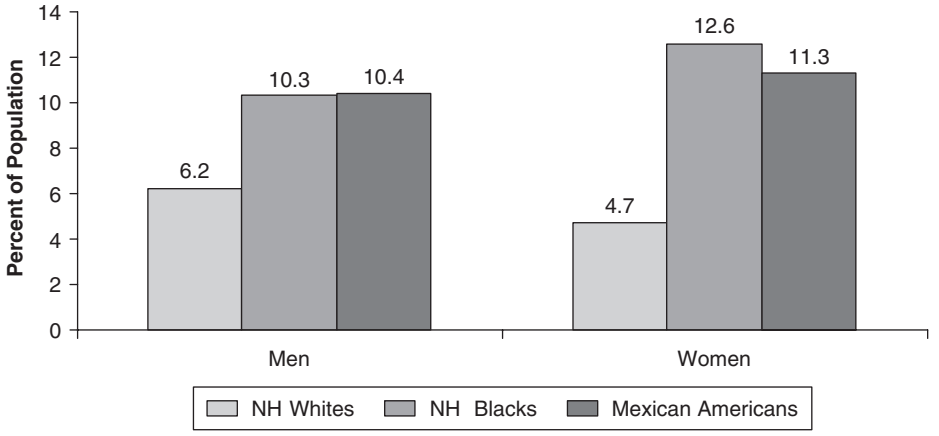


Figure 5 Age-Adjusted Prevalence of Physician-Diagnosed Diabetes in Americans Aged 20 and Older by Race/Ethnicity and Sex, NHANES: 1999 to 2002

Source: Thom et al. (2006).

diabetes lowers cardiovascular event rates is lacking. Ongoing large-scale multicenter trials are now underway to show whether aggressive glycemic control can reduce cardiovascular event risk. Most persons with diabetes have hypertension and dyslipidemia (particularly low HDL-cholesterol and elevated triglycerides). The majority of persons with diabetes are inadequately controlled for hypertension and dyslipidemia. Control of diabetes focuses on lifestyle (diet and exercise) and pharmacologic management to reduce the hemoglobin A1C levels to below 7%, blood pressure to < 130/80 mmHg, and LDL-cholesterol levels to < 100 mg/dl, which are all important to reducing the future risk of CVD events.

Of particular interest recently is the designation of metabolic syndrome, referring to a constellation of cardiometabolic risk factors that are associated with a greater risk of developing future diabetes and CVD. Certain definitions place insulin resistance or abdominal obesity as the necessary condition, for which additional conditions, including elevated blood pressure, low HDL-cholesterol, elevated triglycerides, and impaired fasting glucose, form the definition. The most recent National Cholesterol Education Program definition requires the presence of at least three of the following five criteria: abdominal obesity defined by a waist circumference > 35 in. in women or > 45 in. in men, high-density lipoprotein (HDL)-cholesterol < 40 mg/dl in men or < 50 mg/dl in women, fasting triglycerides > 150 mg/dl, elevated blood pressure > 130 mmHg systolic or > 85 mmHg diastolic or on hypertensive therapy, or impaired fasting glucose defined as > 100 mg/dl or on hypoglycemic therapy. Numerous studies have shown an increased risk of future CVD events in persons with the metabolic syndrome. Among U.S. adults, a stepwise increase in risk for CHD, CVD, and total mortality has been shown from having neither condition to metabolic syndrome without diabetes, diabetes, preexisting CVD, and CVD plus diabetes (Figure 6).

Hypertension

Blood pressure, and particularly systolic blood pressure, is strongly and positively related to the risk of future CHD and stroke. Hypertension is currently defined as a systolic blood pressure of 140 mmHg or higher, diastolic blood pressure of 90 mmHg or higher, or if the person is taking antihypertensive medication. Recent data show approximately 29% of U.S. adults to have hypertension, with a prevalence that rises dramatically with age to nearly 70% of men and more than 80% of women aged 75 years or above (Figure 7). The Seventh Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure also defines those with a level of blood pressure of 120 to 139 mmHg systolic or 80 to 89 mmHg diastolic as "prehypertensive" because of the greater future risk for these persons to develop hypertension. Among U.S. adults, approximately 29%

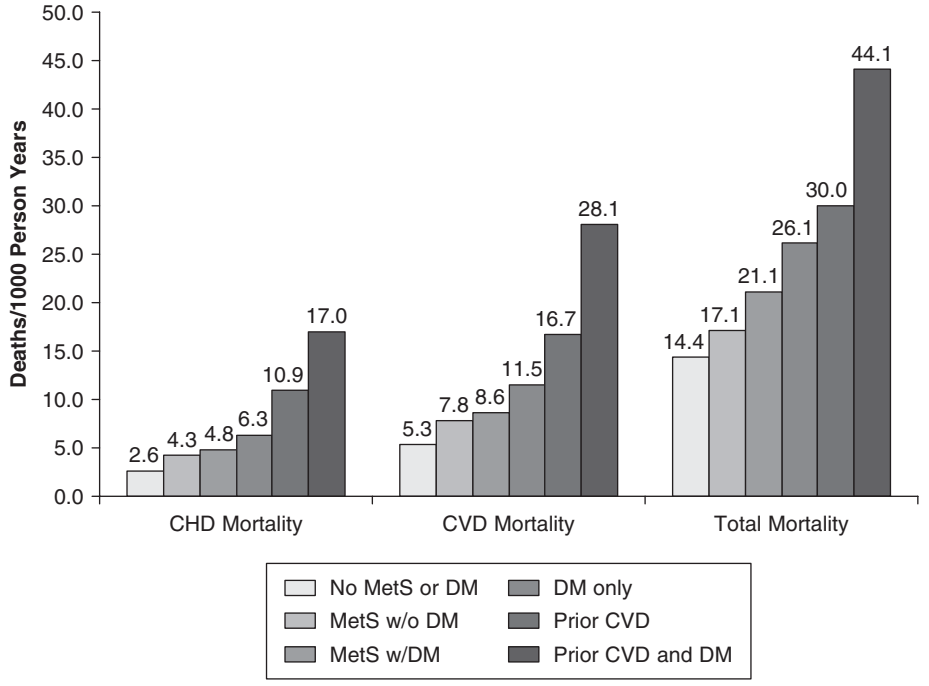


Figure 6 Mortality Rates in U.S. Adults, Ages 30 to 75, With Metabolic Syndrome (MetS), With and Without Diabetes Mellitus (DM) and Preexisting CVD NHANES II: 1976 to 1980 Follow-Up Study (Mean 13 Years Follow-Up)

Source: Malik, Wong, and Franklin (2004).

Note: Estimates are age-adjusted.

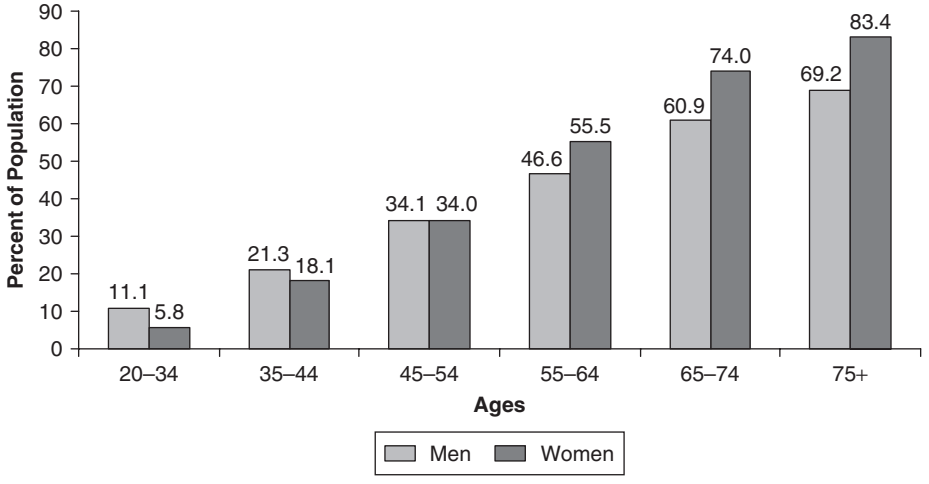


Figure 7 Prevalence of High Blood Pressure in Americans by Age and Sex, NHANES 1999 to 2002

are prehypertensive. Blood pressure is not considered normal unless < 120 mmHg systolic and < 80 mmHg diastolic.

Numerous clinical trials have shown that lowering blood pressure substantially decreases the risk of future cardiovascular events, stroke, and end-stage

renal disease. In hypertensive persons, it is recommended to achieve a blood pressure goal of < 140/90 mmHg, or < 130/80 mmHg for those with diabetes or chronic kidney disease. While there have been major improvements in the treatment of hypertension, only about one third of adult patients with hypertension are adequately controlled. This varies substantially by ethnicity (in 1999–2000, rates of control were lowest in Mexican Americans [18%], followed by non-Hispanic blacks [28%], and non-Hispanic whites [33%]).

Dyslipidemia

Increased levels of total and low-density lipoprotein (LDL)-cholesterol have long been recognized as major risk factors for CHD. A direct curvilinear relation exists between total and LDL-cholesterol with the risk of CHD; compared with those with total cholesterol levels of 200 mg/dl, those with levels of 240 mg/dl are at approximately a twofold greater risk, and those with levels of 300 mg/dl a fourfold risk. However, there is significant overlap in total cholesterol levels between those who experience CHD events versus those who do not, and approximately one third of heart attacks occur in persons with "normal" levels of total cholesterol below 200 mg/dl. Importantly, a low level of HDL-cholesterol, regardless of level of total cholesterol, is strongly associated with an increased risk of CHD (Figure 8). Other lipid abnormalities include elevated serum triglyceride levels, small particle size LDL-cholesterol (known as small dense LDL-cholesterol), and elevated lipoprotein (a) levels. While an elevated LDL-cholesterol level is the most common lipid abnormality in U.S. adults, those with diabetes often have elevated triglycerides and low HDL-cholesterol. In 2003, 40% of U.S. adults had an LDL-cholesterol > 130 mg/dl, and 23% had a low HDL-cholesterol of < 40 mg/dl.

Current National Cholesterol Education Program (NCEP) guidelines focus on treatment of LDL-cholesterol in persons with dyslipidemia. Numerous primary and secondary prevention clinical trials have documented the efficacy of lowering LDL-cholesterol, particularly by the HMG-CoA reductase inhibitor drugs ("statins"), where approximately 25% to 35% reductions in CHD and CVD incidence and, in some studies, significant reductions in mortality have been demonstrated. Furthermore, recent clinical trials also show an effect on reducing incidence of first stroke from LDL-cholesterol reduction.

Trials done specifically in those with diabetes or CHD, regardless of baseline LDL-cholesterol level and in patients with CHD who have fairly normal LDL-cholesterol levels, also show significant reductions in first or recurrent CHD events, which has led to emerging recommendations to treat all patients with CHD or diabetes with statins, even if their baseline LDL-cholesterol is < 130 mg/dl. Standard NCEP goals for

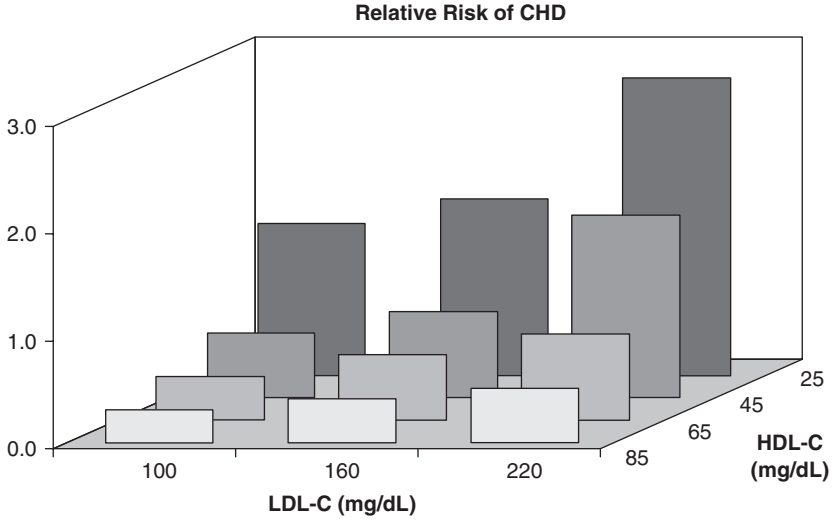


Figure 8 Low HDL-Cholesterol Is an Independent Predictor of CHD Risk Even When LDL-Cholesterol Is Low

Source: Gordon, Castelli, Hjortland, Kannel, and Dawbaer (1977).

treatment of LDL-cholesterol include achieving levels of < 100 mg/dl in those with preexisting CHD or other CHD risk equivalents such as those with diabetes or other atherosclerotic disease or a calculated global risk of > 20% for CHD in 10 years, < 130 mg/dl in those with two or more risk factors, and < 160 mg/dl in those with less than two risk factors (see the section below on risk assessment). Optional goals for lowering LDL-cholesterol to < 70 mg/dl have been recommended for those at the very highest risk (e.g., preexisting CVD plus diabetes or other uncontrolled risk factors, or with acute coronary syndromes).

Cigarette Smoking

Tobacco smoking is among the leading preventable causes of death. Numerous studies have linked tobacco use to the incidence of and mortality from CVD, with approximately a half million deaths annually in the United States being attributed to tobacco use. Furthermore, environmental tobacco smoke ("secondhand smoke") is responsible for approximately 40,000 deaths from heart disease annually in the United States. Cigarette smokers are two to four times more likely to develop CHD than nonsmokers. Also, their risk of stroke is doubled and risk for peripheral vascular disease is more than 10 times as likely as that of nonsmokers. The most recent data among U.S. adults in 2004 show an overall prevalence of smoking of 20%, with a wide variation by ethnic/gender groups, ranging from 5% in Asian females to 37% in American Indian/Alaska Native men. Among adolescents (aged 12 to 17), recent prevalence estimates also show great variation, from 9% in Asians to 30% in American Indians/Alaska Natives.

Interventions for smoking are varied. For youth, school- and community-based prevention programs, state and federal initiatives, and cessation assistance are available. For adults, behavioral treatment, self-help approaches, and pharmacologic therapy are used, with varying levels of success.

Obesity and Physical Inactivity

Recent estimates show that two thirds of U.S. adults are overweight or obese (body mass index of 25 kg/m^2 or higher), with 30% being obese (body mass index of 30 kg/m^2 or higher). Since the 1960s, there has been a dramatic, nearly threefold increase in obesity prevalence among men and more than a twofold increase in women (Figure 9).

Furthermore, recent data show one third of children and adolescents are overweight, based on being at or above the 95th percentile for sex-specific body

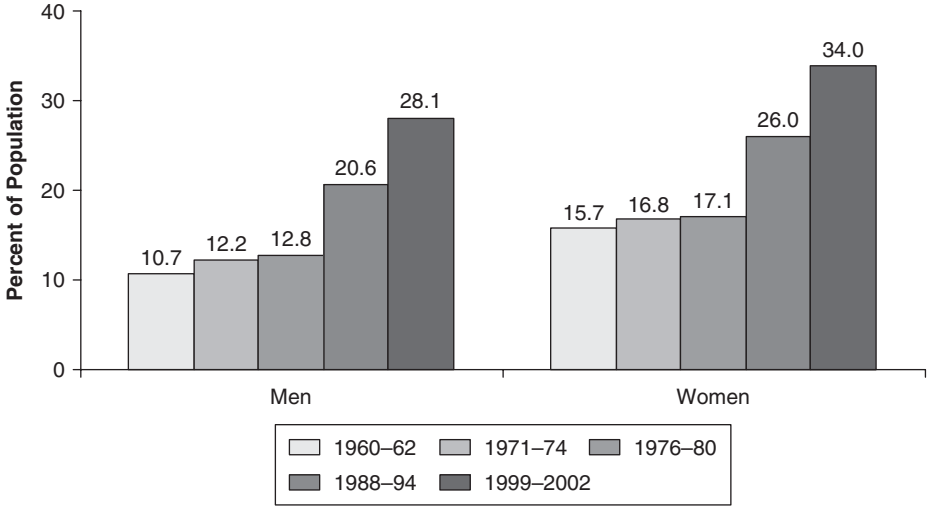


Figure 9 Age-Adjusted Prevalence of Obesity in Americans Ages 20 to 74 by Sex and Survey: NHES, 1960 to 1962; NHANES, 1971 to 1974, 1976 to 1980, 1988 to 1994, and 1999 to 2002

Source: Thom et al. (2006).

Note: Obesity is defined as a body mass index of 30 kg/m^2 or higher.

mass index for age. Further recent estimates show that only 30% of the U.S. adult population is physically active (light or moderate activity at least five times per week for 30 min or vigorous activity at least three times per week for 20 min).

Obesity has been shown by numerous studies to be associated with an approximately 1.5- to 2-fold increase in risk of death from CHD, with the increase in risk beginning below the 25 kg/m^2 cut point for overweight. Numerous studies also show approximately 20% to 40% lower risks of mortality and cardiovascular events associated with increased levels of physical activity or measured fitness. Cardiovascular risk factors are linked to obesity, including hypertension, dyslipidemia (including low HDL-cholesterol levels), type 2 diabetes, obstructive sleep apnea, and hyperinsulinemia. Increases in physical fitness have also been shown to be linked to increases in HDL-cholesterol levels and reductions in systolic and diastolic blood pressure, insulin resistance, and glucose intolerance. Abdominal obesity, most commonly indicated by a waist circumference of > 40 in. in men or > 35 in. in women, is a major component of the metabolic syndrome. Studies demonstrate that weight loss can substantially improve many cardiometabolic risk factors.

The National Institutes of Health Obesity Educational Initiative provides guidelines on the identification, evaluation, and treatment of overweight and obese adults. Assessment includes measurement of body mass index, waist circumference, as well as accompanying other risk factors. Moderate hypocaloric diets of 1,000 to 1,200 kcal per day are generally recommended to provide moderate, sustained weight loss. Incorporating diet and exercise together has been shown to result in longer-term success in weight control.

Cardiovascular Disease Risk Assessment

To understand an individual's risk of CVD or CHD, it is important to understand an individual's "global risk" of developing the condition. Most typically, this involves determination of the future risk of developing CHD in the next 10 years. This can be done by various "risk assessment" algorithms. Most commonly used for this purpose are the Framingham Risk Algorithms that are recommended to assess an individual's 10-year risk of CHD for the purposes of appropriate stratification for risk-factor management, especially lipid management. From knowing an individual's age, gender, systolic blood pressure (and treatment status), current smoking status, total cholesterol, and HDL-cholesterol, for which each factor is assigned a certain number of points according to its presence (and degree) or absence, a total score is obtained (Tables 1 and 2), which corresponds to the probability of suffering a hard CHD event in the next 10 years based on Framingham follow-up data. If the projected 10-year risk of CHD is $< 10\%$, the individual is generally considered to be at low risk of CHD, 10% to 20% intermediate risk, and if $> 20\%$ is judged to be at high risk and, in fact, a CHD risk equivalent (a condition or combination of risk factors conferring a future risk or prognosis similar to that of diagnosed CHD). Besides known CHD, other CHD risk equivalents include diabetes and other forms of atherosclerotic disease such as peripheral arterial disease, significant carotid artery disease, or abdominal aortic aneurysm. Use of these algorithms have been recommended for those with at least two major risk factors out of the following: premature family history of CHD ($<$ 45 in male or $<$ 55 in female first-degree relative), low HDL-C, hypertension, cigarette smoking, and advanced age (male $\geq$ 55 years old or female $\geq$ 65 years old). Those with fewer than two risk factors are felt to be generally at lower risk, and those with diabetes, CHD, or other atherosclerotic disease would also be designated at high risk for aggressive treatment, so such a calculation would not be warranted in these individuals.

In those where a global risk estimate is obtained, depending on the risk status, the appropriate intensity of treatment for given risk factors is considered. For instance, if an individual is a high risk or a CHD risk equivalent, treatment for dyslipidemia would be recommended to reduce the LDL-cholesterol level to < 100 mg/dl. If at intermediate risk or low risk, however, these goals would be < 130 and < 160 mg/dl, respectively.

While estimation of global risk as described above is recommended, it may be best considered a starting point in risk assessment. The presence of other risk factors not included in the risk algorithms, such as abdominal obesity, elevated triglycerides (if severe enough), impaired fasting glucose, or a strong positive premature family history, if present, however, could be used by the clinician to stratify an individual's risk or intensity of treatment upward. Others have also recommended novel risk-factor

Table 1 Estimate of 10-Year Risk of CHD for Men: Framingham Risk Assessment

Age	*Points*
20–34	–9
35–39	–4
40–44	0
45–49	3
50–54	6
55–59	8
60–64	10
65–69	11
70–74	12
75–79	13

	Points				
Total Cholesterol	*Age 20–39*	*Age 40–49*	*Age 50–59*	*Age 60–69*	*Age 70–79*
< 160	0	0	0	0	0
160–199	4	3	2	1	0
200–239	7	5	3	1	0
240–279	9	6	4	2	1
≥ 280	11	8	5	3	1

	Points				
	Age 20–39	*Age 40–49*	*Age 50–59*	*Age 60–69*	*Age 70–79*
Nonsmoker	0	0	0	0	0
Smoker	8	5	3	1	1

HDL (mg/dl)	*Points*
≥ 60	–1
50–59	0
40–49	1
< 40	2

Systolic BP (mmHg)	*If Untreated*	*If Treated*
< 120	0	0
120–129	0	1
130–139	1	2
140–159	1	2
≥ 160	2	3

(Continued)

(Continued)

Point Total	*10-Year Risk (%)*
< 0	< 1
0	1
1	1
2	1
3	1
4	1
5	2
6	2
7	3
8	4
9	5
10	6
11	8
12	10
13	12
14	16
15	20
16	25
≥17	≥ 30

Source: Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (2001).

measures, such as C-reactive protein, a measure of systemic inflammation, or the presence of subclinical atherosclerosis, such as coronary artery calcium or carotid intimal media thicknesses, to aid in risk stratification, where if present in significant amounts, could be rationale for stratifying an individual's intensity for treatment upward. High levels of high-sensitivity C-reactive protein have been shown in numerous studies to be independently related to the risk of future cardiovascular events, although controversy remains as to whether it provides incremental prediction of risk over global risk assessment. Both carotid intimal medial thickness and coronary calcium measures have been shown in numerous studies to be associated with the future risk of cardiovascular events, independent of standard risk factors. Coronary calcium measures also provide incremental predictive value over global risk assessment for the prediction of cardiovascular events.

Cardiovascular Disease Prevention

Cardiovascular epidemiology and prevention encompasses an extensive field investigating the distribution and variation of CVD conditions, most notably CHD and stroke, their risk-factor determinants, and strategies at the population and individual levels aimed to prevent the development or recurrence of CVD. Epidemiologic approaches to studying CVD provide us with the tools for preventive efforts at the individual and population levels. *Primordial prevention* is aimed at prevention of the risk factor for CVD, such as efforts aimed to prevent hypertension, obesity, or dyslipidemia. *Primary prevention* focuses on the modification

Table 2 Estimate of 10-Year Risk of CHD for Women: Framingham Risk Assessment

Age	*Points*
20–34	–7
35–39	–3
40–44	0
45–49	3
50–54	6
55–59	8
60–64	10
65–69	12
70–74	14
75–79	16

	Points				
Total Cholesterol	*Age 20–39*	*Age 40–49*	*Age 50–59*	*Age 60–69*	*Age 70–79*
< 160	0	0	1	0	0
160–199	4	3	2	1	1
200–239	8	6	4	2	1
240–279	11	8	5	3	2
≥ 280	13	10	7	4	2

	Points				
	Age 20–39	*Age 40–49*	*Age 50–59*	*Age 60–69*	*Age 70–79*
Nonsmoker	0	0	0	0	0
Smoker	9	7	4	2	1

HDL (mg/dl)	*Points*
≥ 60	–1
50–59	0
40–49	1
< 40	2

Systolic BP (mmHg)	*If Untreated*	*If Treated*
< 120	0	0
120–129	1	3
130–139	2	4
140–159	3	5
≥ 160	4	6

(Continued)

(Continued)

Point Total	*10-Year Risk (%)*
< 9	< 1
9	1
10	1
11	1
12	1
13	2
14	2
15	3
16	4
17	5
18	6
19	8
20	11
21	14
22	17
23	22
24	27
≥15	≥ 30

Source: Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (2001).

of these and other known risk factors aimed to prevent the clinical manifestations of CVD, such as myocardial infarction and stroke. *Secondary prevention* focuses on those who already have manifestations of disease, but where aggressive control of risk factors can have a major impact in preventing recurrences of disease. Concerted efforts between governmental agencies, the community, and the private sector are required to best address our continuing epidemic of CVD.

—Nathan D. Wong

See also Cholesterol; Diabetes; Framingham Heart Study; Hypertension; Obesity; Physical Activity and Health; Tobacco

Further Readings

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., Jr., et al. (2003). The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC7 report. *Journal of the American Medical Association*, *289*, 2560–2572.

Dawber, T. R. (1980). *The Framingham study: The epidemiology of atherosclerotic disease.* Cambridge, MA: Harvard University Press.

Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. (2001). Executive Summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) (Special communication). *Journal of the American Medical Association*, *285*, 2486–2497.

Gordon, T., Castelli, W. P., Hjortland, M. J., Kannel, W. B., & Dawbaer, T. R. (1977). High density lipoprotein as a protective factor against CHD. The Framingham study. *American Journal of Medicine*, *62*, 707–714.

Gregg, E. W., Cheng, Y. J., Cadwell, B. L., Imperatore, G., Williams, D. E., Flegal, K. M., et al. (2005). Secular

trends in cardiovascular disease risk factors according to body mass index in U.S. adults. *Journal of the American Medical Association*, *293*, 1868–1874.

Grundy, S. M., Cleeman, J. I., Daniels, S. R., Donato, K. A., Eckel, R. H., Franklin, B. A., et al. (2005). Diagnosis and management of the metabolic syndrome: An American Heart Association/National Heart Lung and Blood Institute Scientific Statement. *Circulation*, *112*, e2735–e2752.

Grundy, S. M., Cleeman, J. I., Merz, C. N. B., Brewer, H. B., Jr., Clark, L. T., Hunninghake, D. B., et al. (2004). Implications of recent clinical trials for the National Cholesterol Educational Program Adult Treatment Panel III Guidelines. *Circulation*, *110*, 227–239.

Hopkins, P. N., Hut, S. C., & Wu, L. L. (2000). Family history and genetic factors. In N. D. Wong, H. R. Black, & J. M. Gardin (Eds.), *Preventive cardiology: A practical approach* (chap. 6). New York: McGraw-Hill.

Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N., & Stokes, J., III. (1961). Factors of risk in development of coronary heart disease—six year follow-up experience: The Framingham Study. *Annals of Internal Medicine*, *55*, 33–50.

Kaseta, J., & Sowers, J. R. (2005). Diabetes and the metabolic syndrome. In N. D. Wong, H. R. Black, & J. M. Gardin (Eds.), *Preventive cardiology: A practical approach* (chap. 9). New York: McGraw-Hill.

Luepker, R. V., & Lando, H. A. (2005). Tobacco use, passive smoking, and smoking cessation interventions. In N. D. Wong, H. R. Black, & J. M. Gardin (Eds.), *Preventive cardiology: A practical approach* (chap. 5). New York: McGraw-Hill.

Malik, S., Wong, N. D., & Franklin, S. S. (2004). Impact of the metabolic syndrome on mortality from coronary heart disease, cardiovascular disease, and all causes in United States adults. *Circulation*, *110*, 1239–1244.

McCowen, K. C., & Blackburn, G. L. (2005). Obesity and weight control. In N. D. Wong, H. R. Black, & J. M. Gardin (Eds.), *Preventive cardiology: A practical approach* (chap. 10). New York: McGraw-Hill.

National Institutes of Health, Obesity Education Initiative. (1998). *Clinical guidelines on the identification, evaluation, and treatment of overweight and obese adults* (Executive summary). Bethesda, MD: Author.

Pearson, T. A., Mensah, G. A., Alexander, R. W., Anderson, J. L., Cannon, R. O., III, Criqui, M., et al. (2003). Markers of inflammation and cardiovascular disease: Application to clinical and public health practice. A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation*, *107*(3), 499–511.

Report of Intersociety Commission for Heart Disease. (1984). Optimal resources for primary prevention of atherosclerotic disease. *Circulation*, *70*(Suppl A), 155A–205A.

Thom, T., Haase, N., Rosamond, W., Howard, V. J., Rumsfeld, J., Manolio, T., et al. (2006). Heart disease and stroke statistics—2006 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*, *113*(6), e85–e151.

Williams, R. R., Hunt, S. C., Heiss, G., Province, M. A., Bensen, J., Higgins, M., et al. (2001). Usefulness of cardiovascular family history data for population-based preventive medicine and medical research (The Health Family Tree Study and the NHLBI Family Heart Study). *American Journal of Cardiology*, *87*, 129–135.

Wilson, P. W. F., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., & Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation*, *97*, 1837–1847.

Wilson, P. W. F., Smith, S. C., Blumenthal, R. S., Burke, G. L., & Wong, N. D. (2003). Task Force #4: How do we select patients for atherosclerosis imaging? In 34th Bethesda Conference: Can atherosclerosis imaging techniques improve the detection of patients at risk for ischemic heart disease. *Journal of the American College of Cardiology*, *41*, 1898–1906.

Wong, N. D., Bassin, S. L., & Haskell, W. (2005). Physical activity. In N. D. Wong, H. R. Black, & J. M. Gardin (Eds.), *Preventive cardiology: A practical approach* (chap. 12). New York: McGraw-Hill.

Case-Cohort Studies

The case-cohort design was proposed by Ross Prentice as an alternative design in epidemiologic follow-up studies that is less expensive than a full-scale cohort study. The case-cohort design involves collecting covariate data only for cases, that is, subjects who experience the event of interest in a cohort, and for members of a relatively small randomly selected subcohort. The subcohort may serve as a comparison group for several different types of disease outcomes.

The case-cohort design can substantially reduce cost and effort of exposure assessment by limiting exposure to a small fraction of the cohort with only a small loss of efficiency compared with a full cohort design. The case-cohort design is most beneficial when the most expensive part of the study is not in ascertaining subjects but in measuring their exposures. If the main cost is in ascertainment, a full cohort analysis might be a more sensible approach to analysis.

The design is particularly suited to settings such as molecular epidemiologic studies, where raw materials for covariate information, for example, biospecimens, can be collected and stored. For the cases, these specimens can then be analyzed after the failure, that is, event of interest, has occurred, to determine what an

individual's levels of exposure were at the times before failure. The case-cohort design has been applied in cancer research, cardiovascular disease, and HIV research and has become increasingly popular in genetic epidemiologic studies.

Most methods used to study relative risks in failure time models are based on the Cox proportional hazards model, which assumes a multiplicative form for the model of disease occurrence

$$\lambda(t,x(t),\beta) = \lambda_0(t)rr(x(t),\beta),$$

where $x(t)$ denotes covariates at time t, $\lambda_0(t)$ stands for the baseline hazard rate for subjects with $x = 0$, and rr denotes the relative risk part, with $rr(x(t),\ 0) = 1$. Under the case-cohort design, most existing relative risk estimators are based on modifications of the full cohort partial likelihood score functions, by weighing the contributions from cases and subcohort members by the inverse of their true or estimated sampling probabilities. In the pseudo-likelihood approach proposed by Ross Prentice, for each failure, a sampled risk set is formed by the case and the controls who are in the subcohort. Subcohort members contribute to the analysis over the entire time on study, but the failures outside the subcohort contribute only at their failure times. The pseudo-likelihood contribution is based on the conditional probability that the case fails given that someone fails among those sampled into the risk set. The pseudo likelihood is then the product of the conditional probabilities over failure times. Letting Y_i denote the "at-risk" indicator of the ith subject at the failure time and rr the relative risk part, the pseudo likelihood is given by

$$\prod_{\text{failure times}} \frac{rr_{\text{case}}(x(t),\beta)}{\sum_{\text{case and subcohort}} Y_i rr_i(x(t),\beta)}$$

This pseudo likelihood differs from the partial likelihood for a full cohort study in that the denominator is summing over subjects at risk in the subcohort rather than subjects at risk in the entire cohort. In addition, as cases are added at the time of event, the risk sets are not nested.

The score of the pseudo likelihood has expected value of 0 at the true value of β, but the inverse information does not estimate the variance of the estimator due to sampling-induced covariances between the score terms. The estimator has an asymptotic normal distribution and several approaches to variance estimation are available. Standard Cox regression software can be used to estimate parameters for case-cohort samples; however, variance computations need to be adapted to accommodate the design, which can be done using the *delta beta* option available in many statistics software packages.

When a correlate of the exposure is available for all cohort members, an exposure-stratified subcohort may substantially improve efficiency over a randomly chosen subcohort.

Absolute risk can be estimated from the subcohort using, for example, the Nelson-Aalen estimator for the cumulative baseline hazard, $\int_0^t \lambda_0(u)du$, by summing up the contributions for failure times up to time t:

$$\frac{1}{1/f \sum_{\text{subcohort}} rr_i(x(t),\hat{\beta})},$$

where f denotes the proportion of the cohort sampled into the subcohort.

A second approach to sampling from assembled cohort studies is the nested case-control design. In the nested case-control design, cases of a disease that occur in a cohort are identified, and for each a specified number of matched controls is selected from among those cohort members who have not developed the disease by the time of disease occurrence in the case. Efficiency comparisons between the nested case-control design and the case-cohort design indicate that the former can be somewhat more efficient, due to the sampling-induced positive correlation in score terms in the case-cohort design.

An attractive feature and an advantage of the case-cohort design over the nested case-control design is that the subcohort may serve as the comparison group for multiple disease outcomes. Methods for variance adjustments and hypothesis testing when using a common control group for a range of disease outcomes in case-cohort studies have been proposed, and cause-specific baseline hazard estimates to obtain cause specific absolute risk estimates are available.

—Ruth Pfeiffer

See also Cox Model; Genetic Epidemiology; Study Design

Further Readings

Borgan, O., Langholz, B., Samuelsen, S. O., Goldstein, L., & Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis*, *6*, 39–58.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, *73*, 1–11.

Self, S. G., & Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Annals of Statistics*, *16*, 64–81.

Therneau, T. M., & Li, H. Z. (1998). Computing the Cox model for case cohort designs. *Lifetime Data Analysis*, *5*, 99–112.

Case-Control Study

See Study Design

Case Definition

A case definition is a set of criteria used in making a decision as to whether an individual has a disease or health event of interest. Given that one of the goals of epidemiology is to quantify the magnitude of disease in a population, a case definition is an imperative step in this process because it establishes what criteria constitute a case of the disease. Case definitions are used both in ongoing public health surveillance and during outbreak investigations in field epidemiology.

A case definition should include several key characteristics. It should be clear, simple, and concise so that it can be easily applied to the population of interest. A case definition should be applied equally to all individuals being investigated. By applying the case definition in such a standardized way, the possibility of misclassification bias is minimized. Typically, a case definition includes both clinical and laboratory characteristics. These are ascertained by one or many methods that might include diagnosis by a physician, completion of a survey, and/or routine population screening methods. Individuals meeting a case definition can be categorized as ''confirmed,'' ''probable,'' or ''suspected.''

Case definitions are used in ongoing public health surveillance to track the occurrence and distribution of disease within a given jurisdiction. In the United States, the Centers for Disease Control and Prevention have published a list of uniform case definitions for the mandatory reporting of several diseases called ''Case Definitions for Infectious Conditions Under Public Health Surveillance.'' This list provides explicit case definitions for diseases of interest so that they can be reported by clinicians to public health authorities in a standard and uniform way across geographic locations. This is particularly useful for studies that compare the prevalence of disease across regions as they can use the same case definitions and, therefore, obtain a more accurate picture of disease.

The establishment and application of case definitions are also critical components of outbreak investigations in field epidemiology. A case definition is developed at an early stage of the outbreak investigation so that individual cases can then be identified. While the same criteria apply for developing a case definition in routine public health surveillance, in an outbreak investigation a case definition may also include information regarding person, place, and time, in addition to clinical and laboratory characteristics. For example, a case definition developed for a foodborne outbreak may include only those individuals who ate at a certain restaurant during a specified period of time. Furthermore, a case definition may initially be more broadly defined in an outbreak investigation scenario. This is done to increase sensitivity and therefore capture as many true cases as possible and minimize missing true cases. As the investigation continues and more knowledge is gained about the nature of the cases, the definition may be narrowed and therefore more specific. This is particularly important for a newly emerging disease where a standard case definition does not yet exist. An example of this was the outbreak of Severe Acute Respiratory Syndrome in Toronto in 2003 where a new case definition had to be developed as this was a new disease.

—Kate Bassil

See also Notifiable Disease; Outbreak Investigation; Public Health Surveillance; Sensitivity and Specificity

Further Readings

Centers for Disease Control and Prevention. (1997). Case definitions for infectious conditions under public health surveillance. *Morbidity and Mortality Weekly Report*, *46*(RR-10).

Case-Fatality Rate

Case-fatality rate (CFR) is a clinical measure that describes a person's likelihood of dying from a disease

once he or she is diagnosed with it. It is calculated by dividing the number of deaths from a specified disease during a specified time by the number of individuals who have that disease during that time and multiplying this ratio by 100. It is a measure of disease severity and is often used for prognosis. It can also be used to evaluate the effect of new treatments—with improved treatments, the CFR is expected to decrease. The CFR is not a constant; it can vary between populations and over time depending on the interplay between agent, host, and environment, as well as available treatments and quality of care. The higher the CFR, the higher the likelihood of dying from the disease.

Although CFR has a simple formula, getting an accurate estimate of CFR is not simple. One of the difficulties in estimating CFR is ensuring the accuracy of the numerator. This becomes harder as duration of the disease of interest lengthens because a person becomes more likely to die of another cause prior to death from the disease itself. As a result, the CFR may be underestimated because people who die from another cause will not be counted in the numerator although they may have died from the disease at a later date had they not succumbed to something else first. For an accurate estimate of CFR, one also must be certain that those included in the numerator actually died from the disease in question. If this number includes people who died from other causes, the CFR will be overestimated. These difficulties explain why CFR tends to be a measure used for acute infectious diseases or diseases with short duration rather than for chronic diseases or diseases with long duration.

The denominator used to calculate CFR can also pose a challenge for an accurate estimate. If less severe cases are missed, and therefore not included in the denominator, CFR will be overestimated. For example, underestimating CFR may occur because of an inaccurate denominator that is determined during the early stages of an investigation when people who have the disease in question are missed because they died of the disease prior to the investigation starting.

It is important to point out the difference between CFR and mortality rate. Although number of deaths is the numerator for both, mortality rate is calculated by dividing the number of deaths by the population at risk during a certain time frame, and as a true rate, it estimates the risk of dying of a certain disease. As an example, let us consider two populations. One of them has 1,000 people; of these, 300 people have the disease and 100 people die of it. In this case, the mortality rate for the disease is $100/1{,}000 = 0.1$, or 10%, and the CFR is $100/300 = 0.33$, or 33%. The other population also has 1,000 people but 50 people have the disease and 40 die from it. Here, the mortality rate is $40/1{,}000 = 0.04$, or 4%, and the CFR is $40/50 = 0.8$, or 80%. The incidence of death from the disease is higher in the first population, but the severity is greater in the second. This points to the fact that the two measures provide us with different information.

In summary, the CFR is a proportion that depicts the percentage of people diagnosed with a certain disease who die from the disease within a certain period of time after diagnosis and provides us with information about disease severity.

—Rebecca Harrington

See also Mortality Rates; Proportion

Further Readings

Gordis, L. (2000). Measuring the occurrence of disease. In *Epidemiology* (2nd ed., pp. 31–59). Philadelphia: W. B. Saunders.

Case Reports and Case Series

Case reports and case series in epidemiology and medicine are a type of descriptive study based on singular or grouped uncontrolled observations of patients. Case reports are defined here as singular reports on one individual patient, while case series are collections of information on more than one patient. The subjects in a series usually share one or more common characteristics, such as disease, treatment, or side effect. Usually subjects are followed for a period of time to observe a particular outcome; however, for reasons often to do with convenience or practicability, series may be retrospective collections of patient histories. A series may also be purely descriptive of a point in time—for example, describing variations of clinical presentations. While some series are purely observational, case series can also be uncontrolled clinical trials of interventions, the most sophisticated of which are Phase I or II studies conducted by the pharmaceutical industry. Case

reports and series are an important part of medical publishing due to their perceived methodological ease and observational character.

The number of patients in series is not limited, and large observational studies may be considered case series. This definition by logical extension would also include any series of clinical or physiological experiments in healthy or diseased subjects provided it is uncontrolled; however, in practice these investigations are usually not considered case reports or series.

While case series have no explicit comparison group, most have implicit comparisons, namely, a "normal" clinical course or population. This comparison group may be mentioned or even quantified based on available clinical data.

The most undisputed and common usage of case reports is in clinical education. There they are used as a practical and easily available method to illustrate typical or unusual presentations or clinical courses. Most clinical teaching in medicine relies on cases discussed either informally during ward rounds, during more formalized conferences, or in publications.

Case reports and series are particularly valuable for the reporting of adverse events, where they are used to quickly alert the public to potential side effects of new interventions. Due to the limited number of patients exposed during the development of any clinical interventions, rare side effects usually cannot be detected; thus case reporting is essential in this area. Numerous guidelines have been developed to establish causal relationships in adverse event reporting. However, due to their anecdotal character, case reports and series of adverse events can lead to false-positive reporting. One example is Debendox (Bendectin), a drug that was licensed for pregnancy-related morning sickness. After unsubstantiated reports of fetal malformations, the drug failed on the market, even though large studies refused a causal relationship between the drug and the birth defects.

Case series have been and are extensively used to inform the practitioner about the clinical presentations of diseases or syndromes. They inform about the frequency of signs and symptoms and are thus indispensable for clinical science. Case series may also be used to identify a new set of signs, symptoms, or clinical course as a new pathological entity or separate a clinical entity as a different disease or syndrome. The latter may be comparative and then does not fulfill our criteria. Exceptional case reports may also illuminate the pathophysiology of a particular condition and may overlap with basic science reporting.

A particular strength of case reports is their use in genetics. A well-documented family history can identify a hereditary trait as well as its mode of inheritance and thus document genetic origin; however, it cannot illustrate the relevance of spontaneous mutations in clinical practice.

Case reports and case series have a long history of being used to establish therapeutic effect and to promote treatments that were later discarded because their beneficial effect could not be confirmed by more stringent research such as controlled clinical trials. Therefore, case reports and case series are now considered to be unsuitable for establishing the superiority of one treatment over another, and they have been shown to be particularly prone to reporting bias. However, case reports and series do influence clinical practice, and they remain the only source of information about new therapies under circumstances in which clinical trials are not feasible. To ensure that the information provided by case reports or series is optimized, reporting needs to be complete and transparent for the reader. The authors should give information about diagnosis and diagnostic criteria and outcome measurement, describe the treatment in detail, and use consecutive patients to avoid selection bias as much as possible. Exclusion and inclusion criteria should be explained, the patients must give informed consent, and reporting should be of intention to treat; that is, all patients who start a treatment must be followed and reported whether they finish the treatment or give it up. All issues pertinent to patient safety must be reported. Finally, the authors should describe the expected outcome under standard therapy or the natural course of the disease to give an indication of the comparative advantage or disadvantage of their therapy, although case series can never quantify this difference due to lack of controls. To improve the inference from case reports, the *n* of 1 randomized trial has been suggested, which is a crossover design used mainly to inform individual treatment decisions. In situations where the general outcome of a disease is death and patients on a new treatment survive, case reports and case series can reach the highest grade of evidence, and controlled trials are not needed to establish causal relation. Effects such as these are rare in medicine; an example is insulin treatment of type 1 diabetes mellitus. In observational case reports or

series, the distinction between innovative, often compassionate treatment and experimental clinical trials may be blurry. Phase I and II trials, however, are sophisticated clinical studies and cornerstones of drug development. Most of them are essentially case series, although some studies use placebo controls in Phase I to evaluate side effects, and in the later phases of Phase II drug development controls may be used; these trials may also be randomized.

—Joerg Albrecht and Michael Bigby

See also Bias; Clinical Trials; Study Design

Further Readings

Abel, U. (1999). Erkenntnisgewinn mittels nichtrandomisierter Therapiestudien (Gaining understanding through nonrandomised therapeutic studies). *Ellipse*, *15*, 48–58.

Albrecht, J., Meves, A., & Bigby, M. (2005). Case reports and case series from *Lancet* had significant impact on medical literature. *Journal of Clinical Epidemiology*, *58*, 1227–1232.

Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal*, *312*, 1215–1218.

Jenicek, M. (2001). *Clinical case reporting in evidence-based medicine* (2nd ed.). London: Arnold.

Phillips, B., Ball, C., Sackett, D., Badenoch, D., Straus, S., Haynes, B., et al. (2001). Centre for Evidence-Based Medicine. *Levels of evidence and grades of recommendation*. Retrieved July 27, 2006, from http://www.cebm.net/levels_of_evidence.asp.

Stevens, M. D. B., Talbot, J. C. C., & Routledge, P. A. (Eds.). (1998). *Detection of new adverse drug reactions*. London: Macmillan.

Vandenbroucke, J. P. (2004). When are observational studies as credible as randomised trials? *Lancet*, *363*, 1728–1731.

Categorical Data, Analysis of

While many health-related outcomes are measured on a continuous basis (age, blood pressure, virus particles per milliliter of blood), many common types of epidemiologic data analysis are performed using categorical data. Categorical data are data that exist in discrete groupings. Categories may be predefined, such as gender (male or female), while others may be defined by cutpoints in the data, such as hypertensive (blood pressure greater than or equal to 120/80) or normotensive (blood pressure lower than 120/80). This type of classification results in having only two categories, as opposed to the full range of blood pressure measurements possible. Because these types of data are so common, a specialized field of statistical techniques has been developed. This section documents the key approaches for categorical data analysis used in epidemiology.

Simple Analysis

The 2×2 (“two-by-two”) table is frequently analyzed in very basic epidemiologic studies. This approach is suitable for a study with two levels of exposure (e.g., exposed, not exposed), two categories of outcome (e.g., disease, no disease), and no other factors that might influence the association between exposure and outcome (i.e., no confounding or effect modification). Table 1 displays the common layout of a 2×2 table for an epidemiologic study.

The type of study design will determine the methods used to compute measures of interest, such as the frequency of the outcome for each exposure group, or an estimate of the association between exposure and outcome. This entry will first focus on the simple 2×2 table, and then cover stratified analysis of 2×2 tables to assess the influence of confounding and effect modification.

Simple (Crude) Analysis of Cohort Studies and Randomized Trials

In cohort studies, the investigator wants to determine if an exposure causes new cases of disease (or other outcome). While no single cohort study can definitively prove such a relationship, determining causation is the underlying goal of the research. Cohort studies provide at least two pieces of information

Table 1 Classic 2×2 Table for Epidemiologic Study Data Analysis

	Outcome	No Outcome	Total
Exposed	a	b	$a+b$
Not exposed	c	d	$c+d$
Total	$a+c$	$b+d$	$T=a+b+c+d$

that are useful in understanding causal relationships: (1) determining if an association between exposure and outcome exists (including strength of association) and (2) temporal association. At the beginning of the study period, individuals are classified as exposed or not exposed, and no participants have the outcome of interest. Participants are followed forward to determine who develops the outcome.

From a study design perspective, prospective cohort studies and randomized trials are very similar, primarily differing in how exposure status is assigned. In randomized trials, exposure is assigned by the researchers based on a random assignment protocol, while in cohort studies, exposure is not under the control of the experimenter. The same analytic methods are appropriate for both cohort studies and randomized trials.

A Simple Example With Equal Follow-Up Time

A classic example of a retrospective cohort study is a food poisoning outbreak at a wedding reception. The day after attending an outdoor reception catered by a local company, 65 of 130 guests had symptoms of gastroenteritis (i.e., vomiting, diarrhea, stomach cramps) that resulted in an emergency department visit for medical care. The epidemiologist interviewed all 130 individuals who attended the reception and determined what they ate, if they had symptoms of gastroenteritis, and when those symptoms began. All individuals were asked about symptoms on the wedding day and for the week after the wedding. The investigation identified another 7 individuals with symptoms meeting the case definition for the disease, bringing the total to 72 cases. Of these, 4 people had these symptoms the morning of the wedding. Because of the warm afternoon sun at the reception, consumption of potato salad was suspected as the source of infection (the exposure). Table 2 was designed for the data analysis:

Table 2 Example: Hypothetical Outbreak of Gastroenteritis Among Wedding Guests

	Disease	*No Disease*	*Total*
Potato salad	66	34	100
No potato salad	2	24	26
Total			126

Note that the total for the analysis is 126, not 130. This is because in cohort studies, all individuals must be outcome-free at the beginning of the study. Since 4 people reported symptoms starting before the wedding, this illness could not have been caused by food consumed at the reception. Incidence of disease, meaning the amount of new disease, for exposed and nonexposed individuals is computed using *incidence proportions* (also called *cumulative frequencies*).

$$\begin{aligned}&\text{Incidence proportion (ate potato salad)}\\&= a/(a+b)\\&= 66/100\\&= 0.660\\&= 66\%.\end{aligned}$$

$$\begin{aligned}&\text{Incidence proportion (did not eat potato salad)}\\&= c/(c+d)\\&= 2/26\\&= 0.077\\&= 7.7.\end{aligned}$$

Thus, 66% of individuals who ate potato salad at the wedding developed gastroenteritis during the following week; 7.7% of those who did not eat potato salad became ill with gastroenteritis. Two measures help clarify the association between exposure and disease; that is, are people at more risk of developing gastroenteritis if they ate potato salad than if they did not? The *relative risk* is the ratio of the two incidence proportions:

$$\text{Relative risk} = 0.660/0.077 = 8.6.$$

Those who ate potato salad were about eight times more likely to develop gastroenteritis in the following week compared with those who did not. This is far in excess of a relative risk of one, which would indicate no association between exposure and outcome; that is, the frequency of disease is the same regardless of exposure.

The *risk difference* is the difference of the two incidence proportions.

$$\text{Risk difference} = 0.660 - 0.077 = 0.583 = 58.3/100.$$

The risk difference is helpful to understand how many people developed the outcome because they

were exposed (if the exposure was causal). Another term for risk difference is *attributable risk*, because if the exposure is causal, this provides an estimate of the number of people who developed disease attributable to the exposure. In this case, the risk difference is 0.583, or 58.3/100. That is, of the 100 people exposed to potato salad, it is estimated that 58 people developed gastroenteritis because they ate it. What about the other 8 people who ate potato salad and developed gastroenteritis? Based on the proportion of people who developed gastroenteritis in the nonexposed group, there is at least a potential that about 8 people in the exposed (potato salad) group developed gastroenteritis from other causes. There are several possible explanations, including a different source of gastroenteritis causing disease in the community.

In this example, everyone was followed for 1 week (or until they became ill). So everyone had an equal chance to be discovered ill by the epidemiologist conducting the investigation. The measures of incidence and related measures of association described in this section are appropriate only if everyone in the study is followed for the same length of time or until developing the outcome.

For *cross-sectional* studies, the analysis methods are the same as presented in this section for cohort studies; however, the difference in interpretation is critical. In cross-sectional studies, prevalence is measured, not incidence. Thus, the measure of association focuses on factors associated with existing disease, not new disease or outcome.

Unequal Follow-Up Time for Cohort Studies and Randomized Trials

When a study is conducted over a longer period of time, it is unlikely that everyone will be followed for the entire study period. In a hypothetical 20-year follow-up study, if one person is followed for 1 year and another individual is followed for 15 years, there is more opportunity to identify disease (or another outcome) in the second person compared with the first. Thus, when unequal lengths of follow-up in cohort studies or randomized trials occur, incidence is measured as a rate and risk can be estimated from the rate if desired.

Consider a hypothetical study of a toxic exposure in a small community of 2,000 people. Public health epidemiologists registered people in the community and determined their exposure based on laboratory tests and the person's location during the brief exposure period. Then, the individuals in the registry were monitored to determine what health effects, if any, were related to exposure. The exposures were grouped into three categories: (1) low ($n = 1{,}200$), (2) moderate ($n = 500$), and (3) high ($n = 300$). Over time, some people were lost to follow-up. Every 5 years, the epidemiologists analyzed the data to determine if there was any identifiable disease related to exposure. Because the toxins were known to cause respiratory tract diseases, the focus of investigation was specific to lung diseases. After 10 years of follow-up, the incidence rates were as calculated as shown in Table 3:

Incidence rate (low exposure)
$= 7/10{,}987$ person-years at risk
$= 6.4$ per 10,000 person-years at risk.
Incidence rate (moderate exposure)
$= 15/4837$ person-years at risk
$= 31.0$ per 10,000 person-years at risk.
Incidence rate (high exposure)
$= 12/2878$ person-years at risk
$= 41.7$ per 10,000 person-years at risk.

The reason that the person-years at risk do not add up to 12,000, 5,000, and 3,000, respectively, is because each person was not followed disease-free for 10 years. A contribution of less than 10 years occurs if (1) a person develops lung disease (no longer at

Table 3 Example: Toxic Exposure in Small Community

	Low Exposure	*Moderate Exposure*	*High Exposure*	*Total*
Number of cases of lung disease	7	15	12	34
Person-years of risk (py)	10,987	4,837	2,878	18,702
Rates per 10,000 py	6.4	31.0	41.7	18.2

risk) or (2) a person is lost to follow-up and the outcome of interest (lung disease) is unknown.

$$\begin{aligned}&\text{Rate ratio (comparing moderate with low exposure)}\\&\quad = 31/10{,}000 \text{ person-years}\\&\quad = 4.84 = 6.4/10{,}000 \text{ person-years.}\\&\text{Rate ratio (comparing high with low exposure)}\\&\quad = (41.7/10{,}000 \text{ person-years})/\\&\qquad (6.4/10{,}000 \text{ person-years}) = 6.52.\\&\text{Rate difference (comparing moderate with}\\&\quad \text{low exposure}) = (31/10{,}000 \text{ person-years})\\&\qquad\qquad - (6.4/10{,}000 \text{ person-years})\\&\qquad\qquad = 24.6/10{,}000 \text{ person-years.}\\&\text{Rate difference (comparing high}\\&\quad \text{with low exposure}) = (41.7/10{,}000 \text{ person-years})\\&\qquad\qquad - (6.4/10{,}000 \text{ person-years})\\&\qquad\qquad = 35.3/10{,}000 \text{ person-years.}\end{aligned}$$

Thus, the rate at which the groups develop disease increases as the level of exposure increases. For those with low exposure, about 6 persons develop disease for each 10,000 person-years of risk that occurs. The rate of disease is about 4.8 and 6.5 times faster for those who had a moderate or high exposure compared with low exposure.

The rate difference provides insight into the estimated number of people who may have become sick due to the exposure, if it was truly responsible for the disease differentials. Using these calculations, for every 10,000 person-years of moderate exposure, about 24 cases of lung disease potentially could be attributed to exposure. Given that there were about 5,000 person-years at risk that experienced moderate levels of toxic exposure in the 10-year period of the study, there may be about 12 cases of lung disease that occurred related to moderate exposure. It is important to emphasize that this example proposes a situation where the "toxic exposure" is causing the increased risk of disease seen.

Case-Control Study and Odds Ratio

In contrast to the cohort studies discussed above, the analysis of case-control studies is not straightforward because of the way study participants are selected. In a case-control study, participants are selected on outcome status and then their exposure is determined.

Table 4 Example: Consumption of Tea and Oral Cancer

	Oral Cancer	*No Oral Cancer*
At least 1 cup of tea/day	46	199
Less than 1 cup of tea/day	28	205
Total	74	404

This presents the problem of not being able to look at incidence proportions for the different exposure groups, as was done for the wedding outbreak above. The basic statistic of interest in a case-control study is the relative odds (also called the odds ratio [*OR*]) of exposure between the cases and controls.

Consider a hypothetical study designed to assess if there was a difference in development in oral cancer based on tea consumption. Cases were identified as patients in a specific region with newly diagnosed oral cancer, and controls were people randomly selected from the same region. Note that in modern epidemiology, controls are selected to represent the population from which the cases arose. Thus, it is possible that a person with oral cancer can be in the control group; however, this is very unlikely due to the rarity of oral cancer. In this example, no one in the control group has oral cancer. Epidemiologists interviewed the cases and controls to determine if they drank an average of at least one cup of tea per day or if they drank less than one cup of tea per day, on average. The results are shown in Table 4.

Odds are calculated as the number of exposed cases (or controls) divided by the number of unexposed cases (or controls).

$$\begin{aligned}&\text{Odds of exposure among cases}\\&\quad = a/c = 46/28 = 1.64.\\&\text{Odds of exposure among controls}\\&\quad = b/d = 199/205 = 0.97.\end{aligned}$$

The relative odds or odds ratio is calculated by dividing the odds of exposure among cases by the odds of exposure among controls.

$$\begin{aligned}\text{Odds ratio} &= \text{Relative odds of exposure}\\&\qquad (\text{cases compared with controls})\\&= (a/c)/(b/d) = (ad)/(bc) = 1.69.\end{aligned}$$

Note that the relative odds of exposure are mathematically equivalent to the relative odds of having the disease for exposed compared with nonexposed.

$$\text{Relative odds of disease (exposed compared with nonexposed)} = (a/b)/(c/d) = (ad)/(bc).$$

Thus, in this hypothetical example, the odds of drinking at least one cup of tea per day are about 1.7 times as high in individuals with oral cancer compared with individuals without oral cancer. Because of the study design, no direct estimates of risk can be calculated for the given exposures, and therefore, relative risk and risk difference measures cannot be calculated from the analysis of a case-control study. However, if the disease being studied is rare, commonly defined as less than 10% in all exposure categories, then the odds ratio estimates the relative risk. Thus, the interpretation of this study is that the risk of developing oral cancer among daily tea drinkers is about 1.7 times that of those who drink less or no tea. In epidemiology, if the odds ratio does not estimate the relative risk, it is rarely of interest.

Confidence Intervals and Test Statistics

When reporting estimates, whether incidence, prevalence, proportion exposed, or measures of association, it is important to provide some estimate of precision. The most common form is the confidence interval. Confidence intervals are based on classical statistics theory that requires some form of random selection or random assignment (conditions rarely met in epidemiologic studies). However, confidence intervals are often used for epidemiologic studies that do not involve randomization and are more loosely interpreted as information on precision of the estimate. Confidence intervals for ratio measures (e.g., relative risk, odds ratio) are highly skewed, and thus their confidence intervals are not symmetrical. Fortunately, as long as the sample size of the study is reasonably large, the natural logarithm of the ratio measure (e.g., relative risk, odds ratio, rate ratio) has an approximately normal distribution. Thus, to compute a confidence interval for ratio measures, we first take the logarithm of the measure, compute the confidence interval for the logarithm of the measure (e.g., ln(odds ratio)), then transform the confidence interval by exponentiating each value to obtain the confidence interval in the original units.

For example, using the data from the case-control study, the $OR = 1.69$. Thus,

$$\ln(OR) = 0.52.$$

The variance of the $\ln(OR) = 1/a + 1/b + 1/c + 1/d$.

$$\text{Variance}[\ln(OR)] = 1/46 + 1/199 + 1/28 + 1/205 = 0.067$$

$$95\%\text{confidence interval for } \ln(OR): \ln(OR) + z_{1-\alpha/2} \times \sqrt{\text{variance}[\ln(OR)]}$$

$$0.52 + 1.96 \times (0.259).$$

Exponentiating, the 95% confidence interval for *OR* is (1.01, 2.80).

The Chi-Square Test

The presence of an association between two categorical variables, such as exposure and outcome, can also be tested using the chi-square (χ^2) test statistic. The χ^2 test is used to determine if the observed data in a 2×2 table are statistically significantly different from what would be expected, given the row and column totals for the table. Chi-square testing can be performed for a single 2×2 table, or it can be used to test for overall association when performing stratified analysis, which is discussed in detail below. Chi-square testing is not recommended when the cell counts are low, often defined as having one or more cells with expected value below five, or when there are empty cells in the table.

For a simple table, the χ^2 test statistic is calculated as

$$\chi^2 = \Sigma[(\text{observed data} - \text{expected data})^2/(\text{expected data})],$$

where observed data are the frequencies for the individual cells in the table as observed in the study, and expected data are the frequencies that would be expected, given the distribution of the data in the margins, if there was no association between exposure and outcome. Critical values for the χ^2 distribution are found using a standard χ^2 table, with knowledge of the significance level (α) and the degrees of

freedom. Degrees of freedom for a table are calculated as follows:

$$\text{Degrees of freedom} = (\text{No. of rows} - 1) \times (\text{No. of columns} - 1).$$

Thus, for a 2×2 table, there will be 1 *df*.

The calculation of the χ^2 test statistic will be illustrated using the case-control study example from above. Observed and expected data are shown in Table 5.

Based on these data,

$$\chi^2 = (46-38)^2/38 + (199-207)^2/207 + (28-36)^2/36 + (205-197)^2/197 = 4.10.$$

The critical value for the χ^2 distribution for significance at the $\alpha = .05$ level with 1 *df* is 3.84. Since the computed χ^2 test statistic is greater than the critical value, the null hypothesis of no association between exposure and outcome can be rejected, indicating that there is a statistically significant association. This corresponds to what was found in the calculation of the confidence interval for the odds ratio, where the null value (i.e., 1) was not contained in the calculated confidence interval. The χ^2 test statistic is based on classical statistical theory that requires random sampling, a requirement seldom met in epidemiology studies. In general, epidemiologists focus primarily on estimates and compute a confidence interval to provide some indication of precision rather than relying on test statistics and *p* values because the latter provide limited information given that assumptions for the test typically are not met.

Stratified Analysis: Considering Potential Confounders and Effect Modifiers

Rarely is the association between exposure and outcome unaffected by other factors. Issues of confounding, effect modification, and mediation are particularly important considerations in epidemiologic studies.

Effect Modification

Effect modifiers are factors that modify the relationship between exposures and outcomes. For example, boys have three times the asthma incidence as girls until puberty; after puberty, the asthma incidence is about the same for both boys and girls. For example, analyzing the association between gender and asthma, without considering age, using a simple 2×2 table would not reveal this complexity because all ages would be combined in the table. Stratified analysis is a simple approach that provides an opportunity to examine the potential role of effect modifiers.

In its simplest form, stratified analysis uses 2×2 table data analysis techniques, creating a separate table for each subgroup. For the example of gender and asthma incidence, age is an effect modifier, thus a separate table needs to be made for each age group, as shown in Table 6.

Because the association between gender and asthma incidence is different according to age group, there is no value in combining the information.

Table 5 Observed Data and Expected Data

	Oral Cancer	*No Oral Cancer*	*Total*
Observed data			
≥ 1 cup of tea/day	46	199	245
< 1 cup of tea/day	28	205	233
Total	74 (15.5%)	404 (84.5%)	
Expected data			
≥ 1 cup of tea/day	$245 \times 0.155 = 38$	$245 \times 0.845 = 207$	245
< 1 cup of tea/day	$233 \times 0.155 = 36$	$233 \times 0.845 = 197$	233
Total	74 (15.5% = 0.155)	404 (84.5% = 0.845)	

Table 6 Example: Stratified Analysis of Gender and Asthma

Age: 13 Years or Older		
	Asthma	*No Asthma*
Boys	**Relative Risk = 1**	
Girls		

Age: Less Than 13 Years		
	Asthma	*No Asthma*
Boys	**Relative Risk = 3**	
Girls		

Rather, the answer to the question "What is the association between gender and asthma incidence among children?" must be "It depends on the age of the children." When effect modification occurs, an overall estimate of the association might be misleading. In fact, an overall relative risk that combines all ages likely would partially hide the strong association between gender and asthma among young children and erroneously suggest that there is an association between gender and asthma among older children.

Clinical medicine incorporates effect modifiers frequently for medication decisions. For example, it is known that oral contraceptives are associated with an increased risk of stroke. However, clinical researchers found that the association between oral contraceptives and stroke varies based on a woman's age, smoking status, and presence of hypertension. The risk of stroke is low for young, nonsmoking women without hypertension but increases for older women, smokers, and those with hypertension. These relationships were identified, in part, through a stratified analysis where the association between oral contraceptives and stroke was studied for each combination of age, smoking, and hypertension.

Confounding

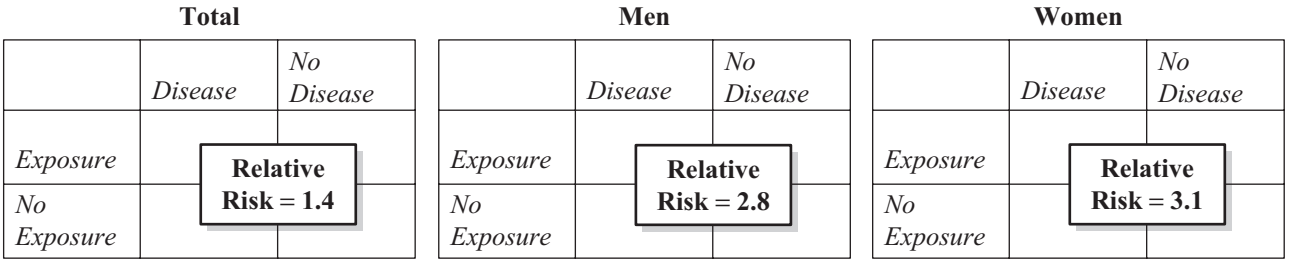
Three key criteria define a confounder: (1) risk factor for the outcome being studied, (2) associated with the exposure being studied independent of the outcome, and (3) not in the causal pathway between the exposure and outcome. Confounders are nuisance factors for the research at hand; if ignored, they can distort the true association between exposure and outcome, resulting in incorrect (biased) estimates. However, if information on confounders is collected, they can be neutralized in the study design or statistical analysis. That is, their effect can be controlled in the analysis so that these factors do not bias the measure of association between exposure and outcome. We focus on confounding in the absence of effect modification. If effect modification by a factor exists, whether the factor is also a confounder is not important from a decision-making perspective in the analysis process.

Confounding can be assessed using stratified analysis. When the data are stratified on a potential confounder, confounding exists if the stratum-specific estimates are (1) approximately the same and (2) different from the overall (also called simple or crude) analysis. Consider the example of a 2-year study between an exposure and disease where gender is thought to be a potential confounder. The stratified analysis results are in Table 7.

For men, exposure is associated with about three times more disease than nonexposure. The same is true for women. However, when all data are combined, the relative risk is 1.4—very different from the stratified results. The difference in overall relative risk is due to confounding. Simply put, the overall estimate of the relative risk is *wrong*. The best estimate of the association between exposure and disease is about 3.

When confounding occurs, the stratum-specific estimates tend to be similar but not exactly the same. Unlike effect modification, where the stratum-specific rates enhance the explanation of the relationship between exposure and outcome, confounding simply distorts the estimate of association between exposure and outcome. Thus, it makes sense to combine the strata into one estimate, as described in the next section.

Table 7 Example of Confounding



Adjusted Estimates of Measures of Association

Epidemiologists have several techniques to remove the confounding effect in an estimate of the association between an exposure and outcome. This process is known as adjusting for confounders. Two methods of estimation that are particularly popular use a stratified analysis approach; both take weighted averages of the stratum-specific estimates.

Mantel-Haenszel Stratified Analysis

The Mantel-Haenszel method is the best-known method for estimating an adjusted measure of association to control for confounding. This method takes a weighted average of the stratum-specific estimates making two key assumptions: (1) The sample size overall is large (note that the sample sizes in the strata may be small but the total sample size must be large), and (2) there is no association between the exposure and outcome. This last assumption appears to be unreasonable for most studies; however, if an association exists and the ratio measure is between 0.4 and 2.5, then the assumption of no association produces only a minor bias toward the null value (i.e., the estimate is slightly biased underestimating the strength of association). Most studies of risk factors for chronic diseases meet this strength criterion. The first assumption is not needed to compute an adjusted estimate but is important to compute confidence intervals or test statistics.

The Mantel-Haenszel method was developed in 1959 when computers were not commonly available. From a computational perspective, its attraction was substantially due to its simplicity. Because of the assumption of no association, the formula for a weighted average of stratum-specific estimates becomes very easy to compute by hand. This particular advantage is no longer important given the availability of computers.

Mantel-Haenszel estimates have another important advantage when dealing with data that include empty cells. When taking weighted averages of strata-specific estimates, in most situations, if a zero occurs in a stratum, even if there are 100 people in the stratum, no data from that stratum will be incorporated into the overall adjusted estimate of the measure of association. The reason is because the measure of association (e.g., relative risk, odds ratio) would be either zero or not defined for the stratum. However, the Mantel-Haenszel method will typically incorporate about half of the information from that stratum into the overall adjusted estimate. Thus, if a zero cell exists, more information will be used by the Mantel-Haenszel method than by other methods also designed to calculate a weighted average of the strata-specific estimates.

The confidence interval for a Mantel-Haenszel adjusted estimate is called a *test-based confidence interval.* The reason it is called test-based is because it is developed from the Mantel-Haenszel test statistic, based on the assumption that no association between exposure and outcome exists.

Precision-Based (Taylor Series) Stratified Analysis

The precision-based, or Taylor series, approach to stratified analysis also takes a weighted average of the stratum-specific estimates (e.g., relative risks). The weight for each stratum is the inverse of the variance

for that stratum. This selection of weights is very common in statistics. The estimate from the stratum with the most data, and thus the most information, typically has the largest weight; similarly, the estimate from the stratum with the least data typically has the smallest weight. Therefore, the overall estimate of association, controlling for the confounder (or confounders), is pulled toward the stratum with the most information. The main disadvantage of this method is that if a stratum has a zero cell, then none of the data in that stratum may be included in the analysis. If no zero cells exist, then this approach is mathematically stronger than the Mantel-Haenszel method because no assumption about the strength of association is needed.

Formulas for computing Mantel-Haenszel and precision-based point estimates, confidence intervals, and overall tests of association are displayed in Table 8.

Stratified Analysis Problem

Most real studies provide data that are not straightforward as in the examples above. The purpose of presenting this problem is to provide an example of stratified analysis using real data and provide some insight into how decisions are made by epidemiologists during the analytic process. This study concerns the number of cesarean sections performed in two regions of a state. The researchers took a quick look at the association between region (exposure) and type of delivery (outcome) for women hospitalized in these regions with their first singleton delivery. This simple analysis is presented in Table 9.

$$\text{Relative risk} = \frac{28.1\%}{17.1\%} = 1.64.$$

Before concluding that the proportion of cesarean sections is higher in Region A due to physician practice (the leading hypothesis), it is important to consider potential confounders and effect modifiers. For instance, the reason for the difference could be due to differences in age or comorbidities (i.e., differences in guideline-appropriate cesarean sections are due to underlying patient characteristics among women giving birth in the two regions). There are several methods to evaluate if ''other factors'' are the reason for an association.

Table 10 presents the data stratified in three ways: on age, receipt of Medicaid benefits (a proxy measure for poverty), and comorbidities. Looking at the age stratification, for girls 16 to 17 years of age, Region A is associated with about 2.2 times the number of cesarean sections compared with Region B. However, for women 18 to 49 years, Region A is associated with about 1.5 times the cesarean sections compared with Region B. There are relatively fewer girls delivering their first baby compared with women 18 years and older, so the confidence interval is larger for the stratum of girls than the stratum of women. Also, while there are fairly large differences in percentages of cesarean sections for women 18 to 34 years and 35 to 49 years within each region, the relative risks for each age group are only moderately different. This distinction is important: Careful analysis requires looking at both the percentages and the measures of association.

Examination of the other stratification variables indicates that women with comorbidities were more likely to have a cesarean section and those with Medicaid were slightly less likely to have a cesarean section. These results are expected, but it is always good to assess if the study's data are consistent with the literature.

Stratification allows for the evaluation of confounding and effect modification simultaneously, using the general criteria for confounders and effect modifiers discussed above. Considering the association between

Table 8 Formulas for Precision-Based and Mantel-Haenszel Point Estimates and Confidence Intervals for Cumulative Incidence Ratio, Incidence Density Ratio, and Odds Ratio

Based on a series of stratified tables laid out as

	Outcome	*No Outcome*	
Exposure	a_i	b_i	N_{1i}
No exposure	c_i	d_i	N_{01}
	M_{1i}	M_{0i}	T_i

The following formulas can be used:

	Cumulative Incidence Ratio	*Incidence Density Ratio*
Precision-based point estimate	$aCIR = \hat{e}\left[\dfrac{\sum_{i=1}^{I}\left(\dfrac{a_i c_i N_{1i} N_{0i}}{a_i d_i N_{1i} + b_i c_i N_{0i}}\right)\left(\ln \hat{CIR}_i\right)}{\sum_{i=1}^{I}\left(\dfrac{a_i c_i N_{1i} N_{0i}}{a_i d_i N_{1i} + b_i c_i N_{0i}}\right)}\right]$	$aIDR = \hat{e}\left[\dfrac{\sum_{i-1}^{I}\left(\dfrac{a_i b_i}{a_i + b_i}\right)\left(\ln \hat{IDR}_i\right)}{\sum_{i-1}^{I}\left(\dfrac{a_i b_i}{a_i + b_i}\right)}\right]$
Precision-based confidence interval	$\hat{e}\left[\ln\left(\hat{aCIR}\right) \pm z_{1-\alpha/2}\left(\dfrac{1}{\sqrt{\sum_{i=1}^{I}\left(\dfrac{a_i c_i N_{1i} N_{0i}}{a_i c_i N_{1i} + b_i c_i N_{0i}}\right)}}\right)\right]$	$\hat{e}\left[\ln\left(\hat{aIDR}\right) \pm z_{1-\alpha/2}\left(\dfrac{1}{\sqrt{\sum_{i=1}^{I}\left(\dfrac{a_i b_i}{a_i + b_i}\right)}}\right)\right]$
Mantel-Haenszel point estimate	$mCIR = \dfrac{\sum_{i=1}^{I}\dfrac{a_i N_{0i}}{T_i}}{\sum_{i=1}^{I}\dfrac{c_i N_{1i}}{T_i}}$	$mIDR = \dfrac{\sum_{i=1}^{I}\dfrac{a_i N_{0i}}{T_i}}{\sum_{i=1}^{I}\dfrac{b_i N_{1i}}{T_i}}$
Mantel-Haenszel confidence interval	$\hat{e}\left[\ln \hat{mCIR} \pm z_{1-\alpha/2}\sqrt{\dfrac{\sum_{i=1}^{I}\dfrac{(M_{1i}N_{1i}N_{0i} - a_i c_i T_i)}{T_i^2}}{\left(\sum_{i=1}^{I}\dfrac{a_i N_{0i}}{T_i}\right)\left(\sum_{i=1}^{I}\dfrac{c_i N_{1i}}{T_i}\right)}}\right]$	$\hat{e}\left[\ln \hat{mIDR} \pm z_{1-\alpha/2}\sqrt{\dfrac{\sum_{i=1}^{I}\dfrac{(M_{1i}N_{1i}N_{0i})}{T_i^2}}{\left(\sum_{i=1}^{I}\dfrac{a_i N_{0i}}{T_i}\right)\left(\sum_{i=1}^{I}\dfrac{b_i N_{1i}}{T_i}\right)}}\right]$
Mantel-Haenszel test statistic for overall association	$\chi^2_{MH} = \dfrac{\left(\sum_{i=1}^{I}\left(\dfrac{a_i d_i - b_i c_i}{T_i}\right)\right)^2}{\sum_{i=1}^{I}\dfrac{N_{1i}N_{0i}M_{1i}M_{0i}}{T_i^2(T_i - 1)}}$	$\chi^2_{MH} = \dfrac{\left(\sum_{i=1}^{I} a_i - \sum_{i=1}^{I}\dfrac{M_{1i}N_{1i}}{T_i}\right)^2}{\sum_{i=1}^{I}\left(\dfrac{M_{1i}N_{1i}N_{0i}}{T_i^2}\right)}$

(Continued)

(Continued)	*Odds Ratio*
Precision-based point estimate	$aOR = \hat{e}\left[\dfrac{\sum_{i=1}^{I}\left(\frac{1}{\frac{1}{a_i}+\frac{1}{b_i}+\frac{1}{c_i}+\frac{1}{d_i}}\right)\left(\ln\ \hat{OR}_i\right)}{\sum_{i=1}^{I}\left(\frac{1}{\frac{1}{a_i}+\frac{1}{b_i}+\frac{1}{c_i}+\frac{1}{d_i}}\right)}\right]$
Precision-based confidence interval	$\mathrm{e}\left[\ln\left(a\hat{CIR}\right) \pm z_{1-\alpha/2}\left(\dfrac{1}{\sqrt{\sum_{i=1}^{I}\left(\frac{1}{\frac{1}{a_i}+\frac{1}{b_i}+\frac{1}{c_i}+\frac{1}{d_i}}\right)}}\right)\right]$
Mantel-Haenszel point estimate	$mOR = \dfrac{\sum_{i=1}^{I}\frac{a_i d_i}{T_i}}{\sum_{i=1}^{I}\frac{b_i c_i}{T_i}}$
Mantel-Haenszel confidence interval	$\hat{e}\left[\ln\ \hat{MOR} \pm z_{1-\alpha/2}\sqrt{\dfrac{\sum_{i=1}^{I}\left(\frac{a_i+d_i}{T_i}\right)\left(\frac{a_i d_i}{T_i}\right)}{2\left(\sum_{i=1}^{I}\frac{a_i d_i}{T_i}\right)^2} + \dfrac{\sum_{i=1}^{I}\left[\left(\frac{a_i+d_i}{T_i}\right)\left(\frac{b_i c_i}{T_i}\right)+\left(\frac{b_i+c_i}{T_i}\right)\left(\frac{a_i d_i}{T_i}\right)\right]}{2\sum_{i=1}^{I}\left(\frac{a_i d_i}{T_i}\right)\sum\left(\frac{b_i c_i}{T_i}\right)} + \dfrac{\sum_{i=1}^{I}\left(\frac{b_i+c_i}{T_i}\right)\left(\frac{b_i c_i}{T_i}\right)}{2\left(\sum_{i=1}^{I}\frac{b_i c_i}{T_i}\right)^2}}\right]$
Mantel-Haenszel test statistic for overall association	$\chi^2_{\mathrm{MH}} = \dfrac{\left(\sum_{i=1}^{I}\left(\frac{a_i d_i - b_i c_i}{T_i}\right)\right)^2}{\sum_{i=1}^{I}\frac{N_{1i}N_{0i}M_{1i}M_{0i}}{T_i^2(T_i-1)}}$

Table 9 Example: Region of State and Type of Delivery

	Cesarean		*Vaginal*		
	N	%	*N*	%	*Total*
Region A	6918	28.1	17,740	71.9	24,658
Region B	2717	17.1	13,147	82.9	15,864

Table 10 Crude and Stratified Tabular Results for Analysis of Association Between Hospital Location in Two Regions of a State (Region A and Region B) and Singleton Deliveries by Cesarean Section

	Region A				*Region B*				
	Cesarean		*Vaginal*		*Cesarean*		*Vaginal*		
Delivery Type	*N*	%	*N*	%	*N*	%	*N*	%	*RR (95% CI)*
Crude	6,918	28.1	17,740	71.9	2,717	17.1	13,147	82.9	1.64 (1.57, 1.70)
Age									
16–17	64	15.5	349	84.5	32	7.0	424	93.0	2.21 (1.48, 3.30)
18–34	4,856	26.3	13,631	73.7	2,037	16.1	10,598	83.9	1.63 (1.56, 1.71)
35–49	1,998	34.7	3,760	65.3	648	23.4	2,125	76.6	1.48 (1.38, 1.60)
Medicaid benefits									
Yes	1,058	20.5	4,104	79.5	1,836	15.9	9,703	84.1	1.29 (1.20, 1.38)
No	5,860	30.0	13,636	70.0	881	20.4	3,444	79.6	1.48 (1.39, 1.57)
Comorbidities									
2 or more	1,976	27.5	5,222	72.6	1,005	19.5	414	80.5	1.41 (1.31, 1.50)
0 or 1	4,942	28.3	12,518	71.7	1,712	16.0	9,007	84.0	1.77 (1.69, 1.86)

region and delivery type, stratified separately by each factor, it appears that

- Age is an effect modifier.
- Medicaid benefits is a confounder (both strata estimates are lower than the crude, thus the crude overall estimate would overestimate the true association).
- Comorbidity is an effect modifier.

The following thought process will help develop our conclusions. The first question to consider is, ''Does the crude result portray the overall association between the exposure and outcome?''

- If the answer is ''*yes*,'' then no adjusted or stratified estimates are needed. The analysis is done and the crude table with its estimates can be used to present the information.
- If the answer is ''*no*,'' then the next question is, ''Do we need to provide each stratum-specific estimate individually or can we summarize some or all of the estimates into one?'' The key point is to focus on what is the best way to explain the association between exposure (region) and outcome (type of delivery). In this example, $RR = 1.64$ does not provide a good estimate overall.

The literature also indicates that the more risk factors pointing to the benefit of cesarean sections, the more likely they will be done regardless of where a woman lives. This is called the *ceiling effect.* In this case, both older age and multiple comorbidities are

Table 11 Multiply Stratified Tabular Results for Analysis of Association Between Hospital Location in Two Regions of a State (Region A and Region B) and Deliveries by Cesarean Section

			Region A				*Region B*				
	Delivery Type		*Cesarean*		*Vaginal*		*Cesarean*		*Vaginal*		
Medicaid Benefits	*Comorbidities*	*Age*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*RR (95% CI)*
No	0 or 1	16–17	24	16.8	119	83.2	3	6.8	41	93.2	2.46 (0.78, 7.79)
		18–34	2,950	28.7	7,329	71.3	401	18.7	1,748	81.3	1.54 (1.40, 1.69)
		35–49	1,306	34.5	2,475	65.5	193	21.6	699	78.4	1.60 (1.40, 1.82)
No	2 or more	16–17	13	22.8	44	77.2	2	7.7	24	92.3	2.96 (0.72, 12.2)
		18–34	1,021	27.2	2,737	72.8	166	19.1	705	80.9	1.43 (1.23, 1.65)
		35–49	546	36.9	932	63.1	116	33.8	227	66.2	1.09 (0.93, 1.28)
Yes	0 or 1	16–17	20	15.2	112	84.9	10	4.1	233	95.9	3.68 (1.78, 7.63)
		18–34	560	19.8	2,266	80.2	916	14.3	5,505	85.7	1.39 (1.26, 1.53)
		35–49	82	27.4	217	72.6	189	19.5	781	80.5	1.41 (1.12, 1.76)
Yes	2 or more	16–17	7	8.6	74	91.3	17	11.9	126	88.1	0.73 (0.31, 1.68)
		18–34	325	20.0	1,299	80.0	554	17.4	2,640	82.7	1.15 (1.02, 1.31)
		35–49	64	32.0	136	68.0	150	26.4	418	73.6	1.21 (0.95, 1.55)

Table 12 Summary Table for Analysis Stratified by Comorbidity Status and Age Group

		Region A				*Region B*				
		Cesarean		*Vaginal*		*Cesarean*		*Vaginal*		
Delivery Type	*Age*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*aRR (95% CI)*
Crude		6,918	28.1	17,740	71.9	2,717	17.1	13,147	82.9	1.64 (1.57, 1.70)
Comorbidities										
0 or 1	16–17	44	16.0	231	84.0	13	4.5	274	95.5	3.53 (1.95, 6.41)
	18–49	4,898	28.5	12,287	71.5	1,699	16.3	8,733	83.7	1.49 (1.41, 1.58)
2 or more	16–17	20	14.5	118	85.5	19	11.2	150	88.8	1.29 (0.72, 2.32)
	18–49	1,956	27.7	5,104	72.3	986	19.8	3,990	80.2	1.23 (1.13, 1.33)

Note: Adjusted estimates for 18- to 49-year-olds are adjusted for age groups. All estimates adjusted for use of Medicaid insurance.

factors that increase the likelihood that a cesarean section would be beneficial, so it is likely that these factors would increase the proportion of cesarean sections being done anywhere. Thus, note that women 35 to 49 years with two comorbidities had the largest proportion of cesarean sections, but the relative risk portraying the association between regions and delivery type were some of the smallest in magnitude (i.e., 1.09 and 1.21), indicating a large proportion of births by cesarean section in this group, regardless of region.

Also note on the other extreme that for girls 16 to 17 who are healthy and unlikely to have a cesarean

section, a regional difference in practice patterns in this group will be most noticeable. In this example, it is clear that such an effect is evident. For the study at hand, there is no "correct" choice. It may be a study where Table 11 is provided. A smaller summary may be prepared as in Table 12. With only one anomaly in combining ages 18 to 49 years within the other two factors, the overall story appears reasonable to combine these ages. Collapsing the estimate by insurance type may also be useful.

To summarize, many decisions need to be made during the statistical analysis of epidemiologic data. Maintaining a focus on the research question of interest is important and more difficult for those with little experience. Understanding the nature of the association, including how the exposure may be related to the outcome, even if only theoretically, can aid the decision-making process.

—Robert Bednarczyk and Louise-Anne McNutt

See also Causal Diagrams; Causation and Causal Inference; Confounding; Effect Modification and Interaction; Study Design

Further Readings

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Daniel, W. W. (1999). *Biostatistics: A foundation for analysis in the health sciences*. Hoboken, NJ: Wiley.

Goetz, M. A. (2000). *Think outside the box: Analysis of categorical data*. Presented at 2000 Northeast SAS Users Group, Inc. Meeting. Retrieved June 28, 2007, from http://nesug.org/proceedings/nesug00/ps/ps7008.pdf.

Gordis, L. (2004). *Epidemiology* (3rd ed.). Philadelphia: Elsevier Saunders.

Hennekens, C. H., Buring, J. E., & Mayrent, S. L. (1987). *Epidemiology in medicine*. Boston: Little, Brown.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*(4), 719–748.

Rothman, K. J. (2002). *Epidemiology: An introduction.* Oxford, UK: Oxford University Press.

Rothman, K. J, & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical data analysis using the SAS system* (2nd ed.). Cary, NC: SAS Institute.

Szklo, M., & Nieto, F. J. (2004). *Epidemiology: Beyond the basics*. Sudbury, MA: Jones & Bartlett.

Causal Diagrams

From their inception in the early 20th century, causal systems models (more commonly known as structural-equations models) were accompanied by graphical representations or path diagrams that provided compact summaries of qualitative assumptions made by the models. Figure 1 provides a graph that would correspond to any system of five equations encoding these assumptions:

1. Independence of *A* and *B*
2. Direct dependence of *C* on *A* and *B*
3. Direct dependence of *E* on *A* and *C*
4. Direct dependence of *F* on *C*
5. Direct dependence of *D* on *B*, *C*, and *E*

The interpretation of "direct dependence" was kept rather informal and usually conveyed by causal intuition, for example, that the entire influence of *A* on *F* is "mediated" by *C*.

By the 1980s, it was recognized that these diagrams could be reinterpreted formally as probability models, which opened the visual power of graph theory for use in probabilistic inference and allowed easy deduction of other independence conditions implied by the assumptions. By the 1990s, it was further recognized that these diagrams could also be

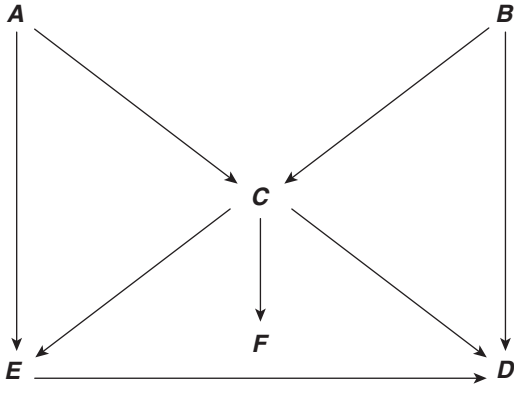


Figure 1 Example of a Directed Acrylic Graph

used as a formal tool for causal inference, such as predicting the effects of external interventions. Given that the graph is correct, one can see whether the causal effects of interest (target effects, or causal estimands) can be estimated from available data, or what additional observations are needed to validly estimate those effects. One can also see how to represent the effects as familiar standardized effect measures.

This entry gives an overview of (1) components of causal graph theory, (2) probability interpretations of graphical models, and (3) the methodological implications of the causal and probability structures encoded in the graph.

Basics of Graph Theory

As befitting a well-developed mathematical topic, graph theory has an extensive terminology that, once mastered, provides access to a number of elegant results that may be used to model any system of relations. The term *dependence* in a graph, usually represented by connectivity, may refer to mathematical, causal, or statistical dependencies. The connectives joining variables in the graph are called *arcs*, *edge*, or *links*, and the variables are also called *nodes* or *vertices.* Two variables connected by an arc are *adjacent* or *neighbors,* and arcs that meet at a variable are also adjacent. If the arc is an arrow, the tail (starting) variable is the *parent* and the head (ending) variable is the *child.* In causal diagrams, an arrow represents a "direct effect" of the parent on the child, although this effect is direct only relative to a certain level of abstraction, in that the graph omits any variables that might mediate the effect.

A variable that has no parent (such as A and B in Figure 1) is *exogenous* or *external*, or a *root* or *source* node, and is determined only by forces outside the graph; otherwise it is *endogenous* or *internal.* A variable with no children (such as D in Figure 1) is a *sink* or *terminal node.* The set of all parents of a variable X (all variables at the tail of an arrow pointing into X) is denoted by $\text{pa}[X]$; in Figure 1, $\text{pa}[D] = \{B, C, E\}$.

A *path* or *chain* is a sequence of adjacent arcs. A *directed path* is a path traced out entirely along arrows tail-to-head. If there is a directed path from X to Y, X is an *ancestor* of Y and Y is a *descendant* of X. In causal diagrams, directed paths represent causal pathways from the starting variable to the ending variable; a variable is thus often called a cause of its descendants and an effect of its ancestors. In a *directed* graph, the only arcs are arrows, and in an *acyclic* graph there is no feedback loop (directed path from a variable back to itself). Therefore, a directed acyclic graph (DAG) is a graph with only arrows for edges and no feedback loops (i.e., no variable is its own ancestor or its own descendant). A causal DAG represents a complete causal structure in that all sources of dependence are explained by causal links; in particular, all common (shared) causes of variables in the graph are also in the graph.

A variable *intercepts* or *mediates* a path if it is in the path (but not at the ends); similarly, a set of variables S intercepts a path if it contains any variable intercepting the path. Variables that intercept directed paths are *intermediates* on the pathway. A variable is a *collider* on the path if the path enters and leaves the variable via arrowheads (a term suggested by the collision of causal forces at the variable). Note that being a collider is relative to a path; for example, in Figure 1, C is a collider on the path $A \rightarrow C \leftarrow B \rightarrow D$ and a noncollider on the path $A \rightarrow C \rightarrow D$. Nonetheless, it is common to refer to a variable as a collider if it is a collider along any path (i.e., if it has more than one parent). A path is *open* or *unblocked* at noncolliders and *closed* or *blocked* at colliders; hence, a path with no collider (such as $E \leftarrow C \leftarrow B \leftarrow D$) is *open* or *active*, while a path with a collider (such as $E \leftarrow A \leftarrow B \rightarrow D$) is *closed* or *inactive.*

Two variables (or sets of variables) in the graph are *d-separated* (or just separated) if there is no open path between them. Some of the most important constraints imposed by a graphical model correspond to independencies arising from separation; for example, absence of an open path from A to B in Figure 1 constrains A and B to be marginally independent (i.e., independent if no stratification is done). Nonetheless, the converse does not hold; that is, presence of an open path allows but does not imply dependency. Independence may arise through cancellation of dependencies; as a consequence, even adjacent variables may be marginally independent; for example, in Figure 1, A and E could be marginally independent if the dependencies through paths $A \rightarrow E$ and $A \rightarrow C \rightarrow E$ canceled each other. The assumption of faithfulness, discussed below, is designed to exclude such possibilities.

Some authors use a bidirectional arc (two-headed arrow, $\leftrightarrow$) to represent the assumption that two

variables share ancestors that are not shown in the graph; $A \leftrightarrow B$ then means that there is an unspecified variable U with directed paths to both A and B (e.g., $A \leftarrow U \rightarrow B$).

Control: Manipulation Versus Conditioning

The word "control" is used throughout science, but with a variety of meanings that are important to distinguish. In experimental research, to control a variable C usually means to manipulate or set its value. In observational studies, however, to control C (or more precisely, to control for C) more often means to condition on C, usually by stratifying on C or by entering C in a regression model. The two processes are very different physically and have very different representations and implications.

If a variable X is influenced by a researcher, the DAG would need an ancestor R of X to represent this influence. In the classical experimental case in which the researcher alone determines X, R and X would be identical. In human trials, however, R more often represents just an *intention* to treat (with the assigned level of X), leaving X to be influenced by other factors that affect compliance with the assigned treatment R. In either case, R might be affected by other variables in the graph. For example, if the researcher uses age to determine assignments (an age-biased allocation), age would be a parent of R. Ordinarily, however, R would be exogenous, as when R represents a randomized allocation.

In contrast, by definition, in an observational study there is no such variable R representing the researcher influence on X, and conditioning is substituted for experimental control. Conditioning on a variable C in a DAG can be represented by creating a new graph from the original graph to represent constraints on relations within levels (strata) of C implied by the constraints imposed by the original graph. This conditional graph can be found by the following sequence of operations:

1. If C is a collider, join ("marry") all pairs of parents of C by undirected arcs; here dashed lines without arrowheads will be used (some authors use solid lines without arrowheads).
2. Similarly, if A is an ancestor of C and a collider, join all pairs of parents of A by undirected arcs.
3. Erase C and all arcs connecting C to other variables.

Figure 2 shows the graph derived from conditioning on C in Figure 1: The parents A and B of C are joined by an undirected arc, while C and all its arcs are gone. Figure 3 shows the result of conditioning on F: C is an ancestral collider of F and so again its parents A and B are joined, but only F and its single arc are erased. Note that, because of the undirected arcs, neither figure is a DAG.

Operations 1 and 2 reflect that if C depends on A and B through distinct pathways, the marginal dependence of A on B will *not* equal the dependence of A on B stratified on C (apart from special cases). To illustrate, suppose A and B are binary indicators (i.e., equal to 1 or 0), marginally independent, and $C = A + B$. Then among persons with $C = 1$, some will have $A = 1$, $B = 0$ and some will have $A = 0$, $B = 1$ (because other combinations produce $C \neq 1$). Thus, when $C = 1$, A and B will exhibit perfect negative dependence: $A = 1 - B$ for all persons with $C = 1$.

Conditioning on a variable C reverses the status of C on paths that pass through it: Paths that were open at C are closed by conditioning on C, while paths that were closed at C become open at C (although they may remain closed elsewhere). Similarly, conditioning on a descendant of C partially reverses the status of C: Typically, paths that were open at C remain open, but with attenuated association across the path; while paths that were closed at C become open at C, although not as open as when conditioning on C itself. In other words, conditioning on a variable tends to partially reverse the status of ancestors on paths passing through the ancestors. In particular, conditioning on a variable may open a path even if it is not on the path, as with F in Figure 1.

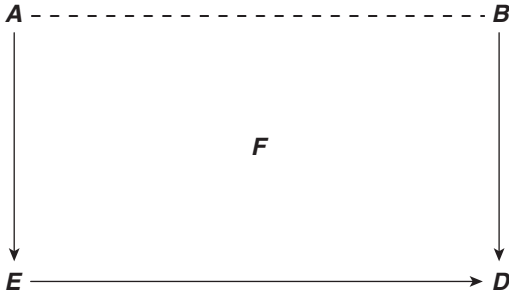


Figure 2 Graph Resulting From Figure 1 After Conditioning on C

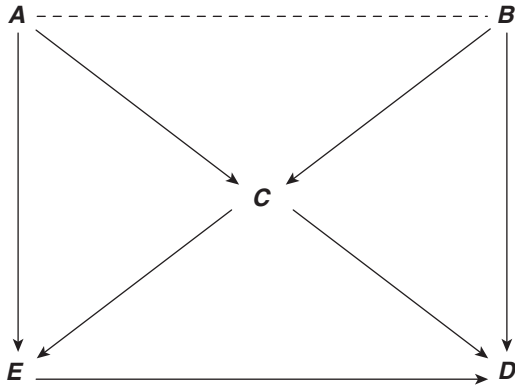


Figure 3 Graph Resulting From Figure 1 After Conditioning on F

A path is closed after conditioning on a set of variables S if S contains a noncollider along the path, or if the conditioning leaves the path closed at a collider; in either case, S is said to block the path. Thus, conditioning on S closes an open path if and only if S intercepts path and opens a closed path if S contains no noncolliders on the path and every collider on the path is either in S or has a descendant in S. In Figure 1, the closed path $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$ will remain closed after conditioning on S if S contains A or B or if S does not contain C, but will be opened if S contains only C, F, or both.

Two variables (or sets of variables) in the graph are *d-separated* (or just separated) by a set S if, after conditioning on S, there is no open path between them. Thus, in Figure 1, $\{A, C\}$ separates E from B, but $\{C\}$ does not (because conditioning on C alone results in Figure 2, in which E and B are connected via the open path A). In a DAG, pa[X] separates X from every variable that is not affected by X (i.e., not a descendant of X). This feature of DAGs is sometimes called the "Markov condition," expressed by saying the parents of a variable "screen off" the variable from everything but its effects. Thus, in Figure 1, pa[E] = $\{A, C\}$, which separates E from B but not from D.

Dependencies induced by conditioning on a set S can be read directly from the original graph using the criterion of *d-separation*, by tracing the original paths in the graph while testing whether colliders are, or have, descendants in S. The conditional dependencies are then illustrated in the original graph by drawing a circle around each C in S to denote the conditioning, then defining a path blocked by S if C is a noncollider on the path, or by a circle-free collider that does not have a circled descendant. Thus, if we circle C in Figure 1, it will completely block the $E - D$ paths $E \leftarrow C \leftarrow B \rightarrow D$ and $E \leftarrow A \rightarrow D$ but unblock the path $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$ via the circled collider C, which is equivalent to having a dashed arc as in Figure 2. Were we to circle F but not C, no open path would be completely blocked, but the collider C would again be opened by virtue of its circled descendant F, which is equivalent to having a dashed arc as in Figure 3.

Selection Bias and Confounding

There is considerable variation in the literature in the usage of terms such as *bias*, *confounding*, and related concepts that refer to dependencies that reflect more than just the effect under study. To capture these notions in a causal graph, we say that an open path between X and Y is a *biasing path* if it is not a directed path. The association of X with Y is then *unbiased* for the effect of X on Y if the only open paths from X to Y are the directed paths. Next, consider a set of variables S that contains no effect (descendant) of X (including those descended through Y). The dependence of Y on X is *unbiased given* S if, after conditioning on S, the open paths between X and Y are exactly (only and all) the directed paths in the starting graph. In such a case, we say S is *sufficient* to block bias in the $X - Y$ dependence and is *minimally sufficient* if no proper subset of S is sufficient.

The exclusion from S of descendants of X in these definitions arises first, because conditioning on X-descendants Z can partially block directed (causal) paths that are part of the effect of interest (if those descendants are intermediates or descendants of intermediates); and second, because conditioning on X descendants can unblock or create paths that are not part of the $X - Y$ effect, and thus create new bias. For example, biasing paths can be created when one conditions on a descendant Z of both X and Y. The resulting bias is called *Berksonian bias*, after its discoverer, Joseph Berkson.

Informally, confounding is a source of bias arising from causes of Y that are associated with but not affected by X. Thus, we say an open nondirected path from X to Y is a *confounding path* if it ends with an arrow into Y. Variables that intercept confounding paths between X and Y are *confounders.* If a confounding path is present, we say *confounding* is present and

that the dependence of Y on X is *confounded*. If no confounding path is present, we say the dependence is *unconfounded*, in which case the only open paths from X to Y through a parent of Y are directed paths. Note that an unconfounded dependency may still be biased due to nondirected open paths that do not end in an arrow into Y (e.g., if Berksonian bias is present).

The dependence of Y on X is *unconfounded given* S if, after conditioning on S, the only open paths between X and Y through a parent of Y are the directed paths. Consider again a set of variables S that contains no descendant of X. S is *sufficient* to block confounding if the dependence of Y on X is unconfounded given S. "No confounding" thus corresponds to sufficiency of the empty set. A sufficient S is called *minimally sufficient* to block confounding if no proper subset of S is sufficient.

A *backdoor path* from X to Y is a path that begins with a parent of X (i.e., leaves X from a "backdoor") and ends at Y. A set S then satisfies the *backdoor criterion* with respect to X and Y if S contains no descendant of X and there are no open backdoor paths from X to Y after conditioning on S. In a DAG, the following simplifications occur:

1. All biasing paths are backdoor paths; hence, the dependence of Y on X is unbiased whenever there is no open backdoor path from X to Y.
2. If X is exogenous, the dependence of any Y on X is unbiased.
3. All confounders are ancestors of either X or of Y.
4. A backdoor path is open if and only if it contains a common ancestor of X and Y.
5. If S satisfies the backdoor criterion, then S is sufficient to block $X - Y$ confounding.

These conditions do not extend to non-DAGs such as Figure 2. Also, although pa[X] always satisfies the backdoor criterion and hence is sufficient in a DAG, it may be far from minimal sufficient. For example, in a DAG there is no confounding and hence no need for conditioning whenever X separates pa[X] from Y (i.e., whenever the only open paths from pa[X] to Y are through X).

The terms *confounding* and *selection bias* have somewhat varying and overlapping usage. Epidemiologists typically refer to Berksonian bias as selection bias, and some call any bias created by conditioning selection bias. Nonetheless, some writers (especially in econometrics) use selection bias to refer to what epidemiologists call confounding. Indeed, Figures 1 and 3 show how selection on a nonconfounder (F) can generate confounding. As a final caution, we note that the biases dealt with by the above concepts are only confounding and selection biases. Biases due to measurement error and model-form misspecification require further structure to describe.

Statistical Interpretations

A joint probability distribution for the variables in a graph is *compatible* with the graph if two sets of variables are independent given S whenever S separates them. For such distributions, two sets of variables will be statistically unassociated if there is no open path between them. Many special results follow for distributions compatible with a DAG. For example, if in a DAG, X is not an ancestor of any variable in a set T, then T and X will be independent given pa[X]. A distribution compatible with a DAG thus can be reduced to a product of factors $Pr(x|\text{pa}[X])$ with one factor for each variable X in the DAG; this is sometimes called the "Markov factorization" for the DAG. When X is a treatment, this condition implies the probability of treatment is fully determined by the parents of X, pa[X].

Suppose now we are interested in the effect of X on Y in a DAG, and we assume a probability model compatible with the DAG. Then, given a sufficient conditioning set S, the only source of association between X and Y within strata of S will be the directed paths from X to Y. Hence the *net effect* of $X = x_1$ versus $X = x_0$ on Y when $S = s$ is defined as $Pr(y|x_1, s) - Pr(y|x_0, s)$, the difference in risks of $Y = y$ at $X = x_1$ and $X = x_0$. Alternatively, one may use another effect measure such as the risk ratio $Pr(y|x_1, s)/Pr(y|x_0, s)$. A *standardized effect* is a difference or ratio of weighted averages of these stratum-specific $Pr(y|x, s)$ over S, using a common weighting distribution. The latter definition can be generalized to include intermediate variables in S by allowing the weighting distribution to causally depend on X. Furthermore, given a set Z of intermediates along all directed paths from X to Y with $X - Z$ and $Z - Y$ unbiased, one can produce formulas for the $X - Y$ effect as a function of the $X - Z$ and $Z - Y$ effects ("front-door adjustment").

The above form of standardized effect is identical to the forms derived under other causal models.

When S is sufficient, some authors go so far as to identify the $Pr(y|x, s)$ with the distribution of potential outcomes given S. There have been objections to this identification on the grounds that not all variables in the graph can be manipulated and that potential-outcome models do not apply to nonmanipulable variables. The objection loses force when X is an intervention variable, however. In that case, sufficiency of a set S implies that the potential-outcome distribution equals $\sum_s Pr(y|x, s)Pr(s)$, the risk of $Y = y$ given $X = x$ standardized to the S distribution.

Some Epidemiologic Applications

To check sufficiency and identify minimally sufficient sets of variables given a graph of the causal structure, one need to only see whether the open paths from X to Y after conditioning are exactly the directed paths from X to Y in the starting graph. Mental effort may then be shifted to evaluating the reasonableness of the causal independencies encoded by the graph, some of which are reflected in conditional independence relations. This property of graphical analysis facilitates the articulation of necessary background knowledge and eases teaching nonstatisticians algebraically difficult concepts.

As an example, spurious sample associations may arise if each variable affects selection into the study, even if those selection effects are independent. This phenomenon is a special case of the collider-stratification effect illustrated earlier. Its presence is easily seen by starting with a DAG that includes a selection indicator $F = 1$ for those selected, 0 otherwise, as well as the study variables, then noting that we are always forced to examine associations within the $F = 1$ stratum (i.e., by definition, our observations stratify on selection). Thus, if selection (F) is affected by multiple causal pathways, we should expect selection to create or alter associations among the variables.

Figure 4 displays a situation common in randomized trials, in which the net effect of E on D is unconfounded, despite the presence of an unmeasured cause U of D. Unfortunately, a common practice in health and social sciences is to stratify on (or otherwise adjust for) an intermediate variable F between a cause E and an effect D, and then claim that the estimated (F residual) association represents that portion of the effect of E on D not mediated through F. In Figure 4, this would be a claim that on

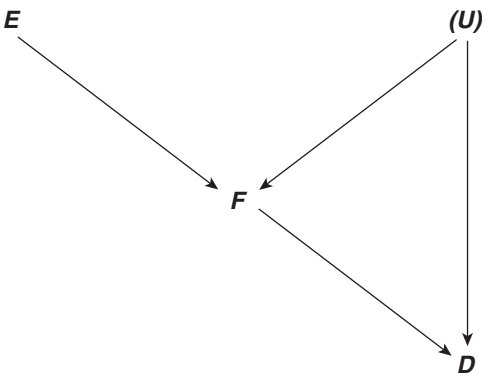


Figure 4 Graph in Which Net (Total) Effect of E on D Is Unconfounded but the Direct Effect Is Confounded by U

stratifying on F, the $E - D$ association represents the direct effect of E on D. Figure 5, however, shows the graph conditional on F, in which we see that there is now an open path from E to D through U, and hence the residual $E - D$ association is confounded for the direct effect of E on D.

The $E - D$ confounding by U in Figure 5 can be seen as arising from the confounding of the $F - D$ association by U in Figure 4. In a similar fashion, conditioning on C in Figure 1 opens the confounding path through A and B in Figure 2; this path can be seen as arising from the confounding of the $C - E$ association by A and the $C - D$ association by B in Figure 1. In both examples, further stratification on either A or B blocks the created path and thus removes the new confounding.

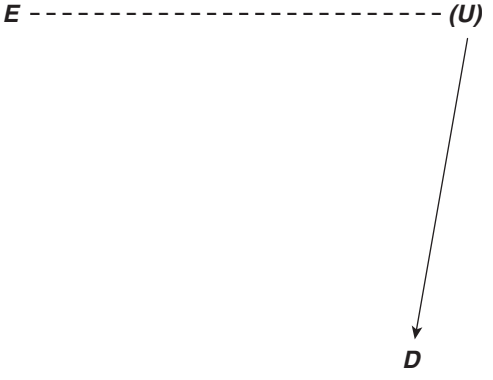


Figure 5 Graph Resulting From Figure 2 After Conditioning on F

The generation of biasing paths by conditioning on a collider or its descendant has been called ''collider bias.'' Starting from a DAG, there are two distinct forms of this bias: confounding induced in the conditional graph (Figures 2, 3, and 5) and Berksonian bias from conditioning on an effect of X and Y. Both biases can in principle be removed by further conditioning on variables along the biasing paths from X to Y in the conditional graph. Nonetheless, the starting DAG will always display ancestors of X or Y that, if known, could be used remove confounding; in contrast, no variable need appear that could be used to remove Berksonian bias.

Figure 4 also provides a schematic for estimating the $F - D$ effect, as in randomized trials in which E represents assignment to or encouragement toward treatment F. Subject to additional assumptions, one can put bounds on confounding of the $F - D$ association (and with more assumptions remove it entirely) through use of E as an *instrumental variable* (a variable associated with X and separated from Y by X).

Questions of Discovery

While deriving statistical implications of graphical models is uncontroversial, algorithms that claim to discover causal (graphical) structures from observational data have been subject to strong criticism. A key assumption in certain ''discovery'' algorithms is a converse of compatibility called *faithfulness.*

A compatible distribution is *faithful to* or *perfectly compatible with* a given graph if for all X, Y, and S, X and Y are independent given S only when S separates X and Y (i.e., the distribution contains no independencies other than those implied by graphical separation). A distribution is *stable* if there is a DAG to which it is faithful. Methods exist for constructing a distribution that is faithful to a given DAG. Methods also exist for constructing a minimal DAG compatible with a given distribution (minimal in that no arrow can be removed from the DAG without violating compatibility). Faithfulness implies that minimal sufficient sets in the graph will also be minimal for consistent estimation of effects. Nonetheless, there are real examples of near cancellation (e.g., when confounding obscures a real effect), which make faithfulness questionable as a routine assumption. Fortunately, faithfulness is not needed for the uses of graphical models discussed here.

Whether or not one assumes faithfulness, the generality of graphical models is purchased with limitations on their informativeness. The nonparametric nature of the graphs implies that parametric concepts such as effect modification cannot be displayed by the graphs (although the graphs still show whether the effects and hence their modification can be estimated from the given information). Similarly, the graphs may imply that several distinct conditionings are minimal sufficient (e.g., both $\{A, C\}$ and $\{B, C\}$ are sufficient for the $E - D$ effect in Figure 1), but offer no further guidance on which to use. Open paths may suggest the presence of an association, but that association may be negligible even if nonzero. For example, bounds on the size of direct effects imply more severe bounds on the size of effects mediated in multiple steps (indirect effects), with the bounds becoming more severe with each step. As a consequence, there is often good reason to expect certain phenomena (such as the conditional $E - D$ confounding shown in Figures 2, 3, and 5) to be small in epidemiologic examples. Thus, when quantitative information is used, graphical modeling becomes more a schematic adjunct than an alternative to causal modeling.

—Sander Greenland and Judea Pearl

Authors' Note: Full technical details of causal diagrams and their relation to causal inference can be found in Pearl (2000) and Spirtes, Glymour, and Scheines (2001). Less technical reviews geared toward health scientists include Greenland, Pearl, and Robins (1999), Greenland and Brumback (2002), Jewell (2004), and Glymour and Greenland (2008).

See also Bias; Causation and Causal Inference; Confounding

Further Readings

Glymour, M. M., & Greenland, S. (2008). Causal diagrams. In K. J. Rothman, S. Greenland, & T. L. Lash (Eds.). *Modern epidemiology* (3rd ed., chap. 12). Philadelphia: Lippincott.

Greenland, S., & Brumback, B. A. (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology*, *31*, 1030–1037.

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, *10*, 37–48.

Jewell, N. P. (2004). *Statistics for epidemiology.* Boca Raton, FL: Chapman & Hall.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, CA: Morgan Kaufmann.

Pearl, J. (1995). Causal diagrams for empirical research (with discussion). *Biometrika*, *82*, 669–710.

Pearl, J. (2000). *Causality.* New York: Cambridge University Press.
Pearl, J. (2001). Causal inference in the health sciences: A conceptual introduction. *Health Services and Outcomes Research Methodology*, *2*, 189–220.
Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (2nd ed.). Cambridge: MIT Press.

Methodologic Applications and Issues

Cole, S., & Hernán, M. A. (2002). Fallibility in estimating direct effects. *International Journal of Epidemiology*, *31*, 163–165.
Freedman, D. A., & Humphreys, P. (1999). Are there algorithms that discover causal structure? *Synthese*, *121*, 29–54.
Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, *29*, 722–729. (Erratum: 2000, *29*, 1102)
Greenland, S. (2003). Quantifying biases in causal models: Classical confounding versus collider-stratification bias. *Epidemiology*, *14*, 300–306.
Hernán, M. A., Hernandez-Diaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology*, *15*, 615–625.
Hernán, M. A., Hernandez-Diaz, S., Werler, M. M., & Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation. *American Journal of Epidemiology*, *155*, 176–184.
Robins, J. M. (2001). Data, design, and background knowledge in etiologic inference. *Epidemiology*, *12*, 313–320.
Robins, J. M., & Wasserman, L. (1999). On the impossibility of inferring causation from association without background knowledge. In C. Glymour & G. Cooper (Eds.), *Computation, causation, and discovery* (pp. 305–321). Menlo Park, CA: AAAI Press. Retrieved February 6, 2007, from http://www.biostat.harvard.edu/%7Erobins/impossibility.pdf.

Causation and Causal Inference

In the health sciences, definitions of cause and effect have not been tightly bound with methods for studying causation. Indeed, many approaches to causal inference provide no definition, leaving users to imagine causality however they prefer. Without a formal definition of causation, an association is distinguished as causal only by having been identified as such based on external and largely contextual considerations. Because they have historical precedence and are still widely used, this entry first reviews such methods. It then discusses definitions and methods based on formal models of causation, especially those based on counterfactuals or potential outcomes.

Canonical Inference

The oldest and most common systematic approach to causal inference in epidemiology was the comparison of observations to characteristics expected of causal relations. The characteristics might derive from subject-matter judgments or from consideration of causal models, and the comparisons might employ formal statistical methods to estimate and test those characteristics. Perhaps the most widely cited of such an approach is based on the considerations of Sir Austin Bradford Hill, which are discussed critically in numerous sources as well as by Hill himself.

The canonical approach usually leaves terms such as *cause* and *effect* as undefined concepts around which the self-evident canons are built, much like axioms are built around concepts such as *set* and *is an element of* in mathematics. In his famous 1965 article on association and causation, Hill noted that he did not want to undertake a philosophical discussion of causation. Only proper temporal sequence (cause must precede effect) is a necessary condition for a cause-effect relation to hold. The remaining considerations are more akin to diagnostic symptoms or signs of causation—that is, they are properties an association is assumed more likely to exhibit if it is causal than if it is not. Furthermore, some of these properties (such as specificity and dose response) apply only under specific causal models. Thus, the canonical approach makes causal inference most closely resemble clinical judgment than experimental science, although experimental evidence is listed among the considerations. Some of the considerations (such as temporal sequence, association, dose-response or predicted gradient, and specificity) are empirical signs and thus subject to conventional statistical analysis. Others (such as plausibility) refer to prior belief, and thus (as with disease symptoms) require elicitation from experts, the same process used to construct prior distributions for Bayesian analysis.

The canonical approach is widely accepted in epidemiology, subject to many variations in detail. Nonetheless, it has been criticized for its incompleteness and informality, and the consequent poor fit it affords to the deductive or mathematical approaches familiar to classic science and statistics.

Although there have been some interesting attempts to reinforce or reinterpret certain canons as empirical predictions of causal hypotheses, there is no generally accepted mapping of the entire canonical approach into a single analytic methodology. One simply uses standard statistical techniques to test whether empirical canons are violated. For example, if the causal hypothesis linking X to Y predicts a strictly increasing trend in Y with X, a test of this statistical prediction may serve as a statistical criterion for determining whether the hypothesis fails the dose-response canon. Such usage falls squarely in the falsificationist/frequentist tradition of 20th-century statistics, but it leaves unanswered most of the policy questions that drive causal research; this gap led to the development of methodologic modeling.

Methodologic Modeling

In the second half of the 20th century, a more rigorous approach to observational studies emerged in the wake of major policy controversies, such as those concerning cigarette smoking and lung cancer. This approach begins with the idea that within strata of some sufficient set of confounders Z, there is a population association or relation between X and Y that is the target of inference. In other words, the Z-stratified associations are presumed to accurately reflect the effect of X on Y in that population stratum, however effect may be defined. Estimates of this presumably causal association are then the effect estimates.

Observational and analytic shortcomings bias or distort these estimates. Units may be selected for observation in a nonrandom fashion; stratifying on additional unmeasured covariates U may be essential for the $X-Y$ association to approximate a causal effect; inappropriate covariates may be entered into the analysis; components of X, Y, or Z may not be adequately measured; and so on. In methodologic or bias modeling, one models these shortcomings. In effect, one attempts to model the design and execution of the study, including features (such as selection biases and measurement errors) beyond investigator control. The process is thus a natural extension to observational studies of the design-based paradigm in experimental and survey statistics.

Nonetheless, many of the parameters in realistic bias models will not be estimable from the data, necessitating inferential approaches well beyond those of conventional statistics. The simplest approach is to fix these parameters at specific values, estimate effects assuming these values are correct, and see how effect estimates change as these values are varied. This process is called sensitivity analysis. One can also assign the parameters prior probability distributions based on background information and summarize the effect estimates over these distributions or over the resulting posterior distribution.

These ideas are well established in engineering and policy research, albeit in a wide variety of forms and specialized applications. Models for specific biases have a long if scattered history in epidemiology; nonetheless, methods for statistical inference from bias models have only recently begun to appear in epidemiologic research.

Statistical Formulation

Consider the problem of estimating the effect of X on Y, given a collection of antecedent covariates Z. Standard approaches estimate the regression of Y on X and Z, $E(Y|x, z)$, and then taking the fitted (partial) regression of Y on X given Z as the effect of X on Y. Usually, a parametric model $r(x, z; \beta)$ for $E(Y|x, z)$ is fit and the coefficient for X is taken as the effect (this approach is reflected in common terminology that refers to such coefficients as ''main effects''); the logistic model for a binary Y is the most common epidemiologic example. Model fitting is almost always done as if

1. within levels of X and Z, the data are a simple random sample and any missingness is completely random;
2. the causal effect of X on Y is accurately reflected by the association of X and Y given Z (i.e., there is no residual confounding—as might be reasonable to assume if X were randomized within levels of Z), and
3. X, Y, and Z are measured without error.

In reality, it is more frequently the case that (1) sampling and missing-data probabilities may jointly depend on X, Y, and Z in an unknown fashion; (2) stratifying or adjusting for certain unmeasured (and possibly unknown) covariates U might be essential for the association of X and Y to correspond to a causal effect of X on Y; and (3) some of the X, Y, and Z components are mismeasured. Several approaches have been developed to deal with these problems.

Selection Biases

Let $V = (X, Y, Z)$. One approach to sampling (selection) biases posits a model $s(v; \sigma)$ for the probability of selection given v, then uses this model in the analysis along with $r(x, z; \beta)$, for example, by incorporating $s(v; \sigma)$ into the likelihood function or by using $1/s(v; \sigma)$ as a weighting factor. The parameters β and σ usually cannot be completely estimated from the data under analysis, so one must either posit various fixed values for σ and estimate β for each chosen σ (sensitivity analysis), or else give β and σ a prior distribution and conduct a Bayesian analysis. A third approach, Monte Carlo risk analysis or Monte Carlo sensitivity analysis (MCSA), repeatedly samples σ from its prior distribution, resamples (bootstraps) the data, and re-estimates β using the sampled σ and data; it then outputs the distribution of results obtained from this repeated sampling-estimation cycle. MCSA can closely approximate Bayesian results under certain (though not all) conditions. These selection-modeling methods can be generalized (with many technical considerations) to handle arbitrary missing data.

Unmeasured Confounders

Suppose U is a collection of unmeasured (latent) covariates required for identification of the effect of X on Y. One approach to problem (2) is to model the distribution of U and V with a probability model $p(u, v; \beta, \gamma) = p(y|u, x, z; \beta)p(u, x, z; \gamma)$. Again, one can estimate β by likelihood-based or by weighting methods, but because U is unmeasured, the parameter (β, γ) will not be fully estimable from the data and so some sort of sensitivity analysis or prior distribution will be needed. Results will depend heavily on the prior specification given U. For example, U may be a specific unmeasured covariate (e.g., smoking status) with well-studied relations to X, Y, and Z, which affords straightforward Bayesian and MCSA analyses. On the other hand, U may represent an unspecified aggregation of latent confounders, in which case the priors and hence inferences are more uncertain.

Measurement Error and Misclassification

Suppose that the collection of "true" values $V = (X, Y, Z)$ has a corresponding collection of measurements or surrogates W (which might include multiple surrogates for X, Y, or Z). The measurement-error problem (Problem 3) can then be expressed as follows: For some or all units, at least one of the V components is missing, but the measurements in W that correspond to the missing V components are present. If enough units are observed with both V and W complete, the problem can be handled by standard missing-data methods. For example, given a model for the distribution of V and W, one can use likelihood-based methods, or impute V components where absent and then fit the model $r(x, z; \beta)$ for $E(Y|x, z)$ to the completed data or fit the model to the complete records using weights derived from all records using a model for missing-data patterns. Many direct Bayesian approaches to measurement error are also available.

Alternatively, there are many measurement-error correction procedures that adjust the "naïve" β estimates obtained by fitting the regression using W as if it were V. This adjustment is usually accomplished with a model relating V to W fitted to the complete records, as in instrumental-variable (regression calibration) corrections and their extensions. Many recent methods based on assuming various subsamples with information on multiple surrogates are available (so W may be of much higher dimension than V and may have complex missing-data patterns).

All methods assume that missingness in V and W components is random, which is often quite implausible because noncooperation increases with demands on subjects, collection of some components may be demanding (e.g., as when W includes diet diaries or biomarkers), and cooperation may be related to unobserved true values or confounders. Thus, selection modeling will be needed along with measurement modeling to account for this nonrandom ("nonignorable") missingness. Further nonidentified modeling becomes a necessity if a component of V is never observed on any unit (or, more practically, if there are too few complete records to support large-sample missing-data or measurement-error procedures). Latent-variable methods are natural for this situation. For example, one could model the distribution of (V, W) or a sufficient factor from that distribution by a parametric model; the unobserved components of V are then the latent variables in the model. As before, the parameters will not be fully identified, making Bayesian methods a natural choice for summary inferences.

Realistic specification for nonidentified measurement-error models can become quite complex, with inferences

displaying extreme sensitivity to parameter constraints or prior distributions. Nonetheless, methodologic modeling helps provide an honest accounting for the large uncertainty that can be generated by even modest measurement error.

Counterfactuals and Potential Outcomes

Skeptical that induction in general and causal inference in particular could be given a sound logical basis, the 18th-century Scottish philosopher David Hume nonetheless captured the foundation of the potential-outcome approach when, in his *Enquiry Concerning Human Understanding*, he defined "a cause to be an object, followed by another, ... where, if the first object had not been, the second had never existed." A key aspect of this view of causation is its *counterfactual* element: It refers to how a certain outcome event (the "second object," or effect) would not have occurred if, *contrary to fact*, an earlier event (the "first object," or cause) had not occurred. In the 20th century, this counterfactual view of causation was adopted by numerous philosophers and scientists. In parallel, there appeared statistical theories of causal inference that incorporated this view into their foundation, which is today widely recognized under the heading of *potential-outcome models* of causation.

To describe these models, suppose we wish to study the effect of an intervention variable X on a subsequent outcome variable Y defined on an observational unit or a population; for example, X could be the daily dose regimen for a drug in a clinical trial, and Y could be survival time. Given X has potential values $x_1, \ldots, x_J$ (e.g., drug doses), we suppose that there is a list of *potential outcomes* $\boldsymbol{y} = (y(x_1), \ldots, y(x_J))'$, such that if $X = x_j$, then $Y = y(x_j)$. The list $\boldsymbol{y}$ thus exhibits the correspondence between treatments, interventions, or actions (the X values) and outcomes or responses (the Y values) for the unit, and so is sometimes called a *response schedule*.

Under this model, assignment of a unit to a treatment level x_j is a choice of which potential outcome $y(x_j)$ from the list $\boldsymbol{y}$ attempts to observe. It is ordinarily assumed that the assignments made for other units do not affect the outcomes of another unit, although there are extensions of the model to include between-unit interactions, as in contagious outcomes. Regardless of the X assignment, the remaining potential outcomes are treated as existing pretreatment covariates on which data are missing. Because at most one of the J potential outcomes is observed per unit, the remaining potential outcomes can be viewed as missing data, and causal inference can thus be seen as a special case of inference with missing data.

To say that intervention x_i causally affects Y relative to intervention x_j means that $y(x_i) \neq y(x_j)$, that is, X "matters" for Y for the unit. The *sharp* (or strong) null hypothesis is that $y(x)$ is constant over x within units; this means that changing X would not affect the Y of any unit, that is, $y(x_i) = y(x_j)$ for every unit and every x_i and x_j; this hypothesis forms the basis of exact permutation tests such as Fisher's exact test. The effect of intervention x_i relative to x_j on a unit may be measured by the difference in potential outcomes $y(x_i) - y(x_j)$. If the outcome is strictly positive (such as life expectancy or mortality risk), it could instead be measured by the ratio $y(x_i)/y(x_j)$.

Because we never observe two potential outcomes on a unit, we can only estimate population averages of the potential outcomes. We do this by observing average outcomes in differently exposed groups and substituting these observations for the average potential outcomes—a perilous process whenever the observed exposure groups are atypical of the population of interest with respect to other risk factors for the outcome. A more subtle problem is that only for difference measures will the population effect (the difference of average potential outcomes) equal the population average effect (the average difference of potential outcomes). Hence, the average of the differences $y(x_i) - y(x_j)$ in the population is often called the *average causal effect*. For some popular measures of effect, such as rate ratios and odds ratios, the population effect may not even equal any average of individual effects.

The theory extends to probabilistic outcomes by replacing the $y(x_j)$ by probability functions $p_j(y)$. The theory also extends to continuous X by allowing the potential-outcome list $\boldsymbol{y}$ to contain the potential outcome $y(x)$ or $p_x(y)$ for every possible value x of X. Both extensions are embodied in graphical probability models for intervention effects. Finally, the theory extends to complex longitudinal data structures by allowing the treatments to be different event histories or processes.

From Randomized to Observational Inference

Potential outcomes were developed as part of a design-based strategy for causal inference in which randomization provided the foundation for inference. Indeed, before the 1980s, the model was often referred to as "the randomization model," although the causal concepts within it do not hinge on randomization. It thus seems that the early strong linkage of potential outcomes to randomized designs deflected consideration of the model for observational research. By the 1970s, however, their extension to observational studies was well under way. In this context, the models made clear the distinction between causal and statistical relations: Causal relations refer to relations of treatments to potential outcomes *within* treated units, whereas statistical relations refer to associations of treatments with actual outcomes *across* units. Consequently, the models have aided in distinguishing confounding from collapsibility, synergy from statistical interaction, and causation probabilities from attributable fractions.

The conceptual clarification also stimulated development of statistical methods for observational studies, leading, for example, to propensity-scoring and inverse-probability-of-treatment methods for confounder adjustment, as well as new insights into analysis of trials with noncompliance. In many cases, such insights have led to methodologic refinements and better-informed choices among existing methods. In the longitudinal data setting, however, potential-outcome modeling has led to entirely new methodologies for analysis of time-varying covariates and outcomes, including *g*-estimation and marginal structural modeling.

A serious caution arises, however, when it is not clear that the counterfactual values for X (treatments other than the actual one) represent physical possibilities or even unambiguous states of nature. A classic example is gender (biological sex). Although people speak freely of gender (male vs. female) as cause of heart disease, given a particular man it is not clear what it would mean for that man to have been a woman instead. Do we mean that the man cross-dressed and lived with a female identity his entire life? Or that he received a sex-change operation after birth? Or that the zygote from which he developed had its male chromosome replaced by a female chromosome?

Potential-outcome models bring to light such ambiguities in everyday causal language, but do not resolve them. Some authors appear to insist that use of the models be restricted to situations in which ambiguities are resolved, so that X must represent an intervention variable, that is, a precise choice among treatment actions or decisions. Many applications do not meet this restriction, however, and some go so far as to confuse outcomes (Y) with treatments (X), which can lead to nonsensical results. Examples include estimates of mortality after "cause removal," for example, removal of all lung cancer deaths. Sensible interpretation of any effect estimate requires asking what intervention a unit could have given the unit a value of X (here, lung cancer death) other than the one that was observed, and what side effects that intervention would have. One cannot remove all lung cancer deaths by smoking cessation. A treatment with a 100% cure rate might do so, but need not guarantee the same subsequent life span as if the cancer never occurred. If such questions cannot be given at least a speculative answer, the estimates of the impact of cause removal cannot be expected to provide valid information for intervention and policy purposes.

Structural Equations and Causal Diagrams

Paralleling the development of potential-outcome models in the 20th century, an entirely different approach to causal analysis arose in observational research in economics and related fields. Like methodologic modeling, this *structural-equations* approach does not begin with a formal definition of cause and effect, but instead develops models to reflect assumed causal associations, from which empirical (and hence testable) associations may be derived. These models may, however, be given a potential-outcomes formulation.

Like most of statistics, before the 1980s, structural-equations methods were largely limited to normal linear models to derive statistical inferences. Because these models bear no resemblance to typical epidemiologic data, this limitation may in part explain the near absence of structural equations from epidemiology, despite their ubiquity in social science methodology. From their inception, however, the models have been accompanied by graphical representations or path diagrams that provide compact summaries of qualitative assumptions made by the structural model.

Such diagrams also provide visual explanations of insights and problems difficult to see with other methods, and so provide an invaluable teaching tool.

Conclusion

Different approaches to causal inference represent separate historical streams rather than distinct methodologies and can be blended in various ways. The result of any modeling exercise is simply one more input to informal judgments about causal relations, which may be guided by canonical considerations. Insights and innovations in any approach can thus benefit the entire process of causal inference, especially when that process is seen as part of a larger context. Other traditions or approaches (some perhaps yet to be imagined) may contribute to the process. It thus seems safe to say that no one approach or blend is a complete solution to the problem of causal inference, and that the topic remains one rich with open problems and opportunities for innovation.

—*Sander Greenland*

Author's Note: Portions of this entry are adapted from Greenland (2004, chap. 1).

See also Causal Diagrams; Confounding; Counterfactual Models; Effect Modification and Interaction; Hill's Considerations for Causal Inference

Further Readings

General

Greenland, S. (2004). An overview of methods for causal inference from observational studies. In A. Gelman & X. L. Meng (Eds.), *Applied Bayesian modeling and causal inference from an incomplete-data perspective* (pp. 3–13). New York: Wiley.

Pearl, J. (2000). *Causality.* New York: Cambridge University Press.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed., chaps. 2, 4). Philadelphia: Lippincott.

Canonical Inference

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, *58*, 295–300.

Hume, D. (1998). *An enquiry concerning human understanding* [Reprint]. Chicago: Open Court Press. (Original work published 1748)

Susser, M. (1991). What is a cause and how do we know one? A grammar for pragmatic epidemiology. *American Journal of Epidemiology*, *133*, 635–648.

Weed, D. L. (1997). On the use of causal criteria. *International Journal of Epidemiology*, *26*, 1137–1141.

Weiss, N. S. (2002). Can ''specificity'' of an association be rehabilitated as a basis for supporting a causal hypothesis? *Epidemiology*, *13*, 6–8.

Methodologic Modeling

Eddy, D. M., Hasselblad, V., & Schachter, R. (1992). *Meta-analysis by the confidence profile method.* New York: Academic Press.

Greenland, S. (2005). Multiple-bias modeling for observational studies (with discussion). *Journal of the Royal Statistical Society*, *Series A*, *168*, 267–308.

Greenland, S., & Lash, T. L. (2008). Bias analysis. In K. J. Rothman, S. Greenland, & T. L. Lash (Eds.). *Modern epidemiology* (3rd ed., chap. 19). Philadelphia: Lippincott.

Gustafson, P. (2003). *Measurement error and misclassification in statistics and epidemiology.* New York: Chapman & Hall.

Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, *47*, 1213–1234.

Counterfactuals, Potential Outcomes, and Causal Models

Berk, R. (2003). *Regression analysis: A constructive critique*. Thousand Oaks, CA: Sage.

Greenland, S. (2005). Epidemiologic measures and policy formulation: Lessons from potential outcomes (with discussion). *Emerging Themes in Epidemiology*, *2*, 1–4.

Greenland, S. (1999). The relation of the probability of causation to the relative risk and the doubling dose: A methodologic error that has become a social problem. *American Journal of Public Health*, *89*, 1166–1169.

Greenland, S., & Brumback, B. A. (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology*, *31*, 1030–1037.

Hernán, M. A. (2005). Hypothetical interventions to define causal effects: Afterthought or prerequisite? *American Journal of Epidemiology*, *162*, 618–620.

Lewis, D. K. (1973). Causation. *Journal of Philosophy*, *70*, 556–567.

MacMahon, B., & Pugh, T. F. (1967). Causes and entities of disease. In D. W. Clark & B. MacMahon (Eds.), *Preventive medicine* (pp. 11–18). Boston: Little, Brown.

Robins, J. M. (1999). Marginal structural models versus structural nested models as tools for causal inference. In M. E. Halloran & D. A. Berry (Eds.), *Statistical models in epidemiology* (pp. 95–134). New York: Springer-Verlag.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *56*, 688–701.

Simon, H. A., & Rescher, N. (1966). Cause and counterfactual. *Philosophy of Science*, *33*, 323–340.

Censored Data

Samples are often collected in such a way that the exact value of one or more cases is unknown. Such missing information is referred to as *censored data.* In one source of such censored data, values are known to exceed or fall below some limit. Often, for example, in a study based on the survival times of laboratory animals, a protocol requires that all data be collected within a specified period of time. For an appreciably sized subset of subjects, an exact survival time may not be known, simply because the study ends before the event of interest (such as the animal's death) could be observed. Because the exact survival times for those who lived longer than the length of the study are not known, this illustrates the generation of right-censored data. Another common example occurs when the age of study participants is recorded in exact years for most subjects, but the ages of those younger than 18 are all recorded as less than 18 years. Because the exact ages of persons younger than 18 years are not known, this is an example of left-censored data. In general, when some, but not all, values are recorded on a continuous scale of measurement, special techniques are required that differ from the many of the common procedures based on what is called *maximum likelihood* (ML) estimation.

Work with censored data is facilitated by a notational convention for observed values called *order statistics.* The sample median is an example of an order statistic. Unlike the mean of a size n sample, which is represented by placing a bar above a letter, as in $\overline{X}$, the subscripted symbol $X_{([n+1]/2)}$, when n is an odd number, or $[X_{(n/2)} + X_{([n/2]+1)}]/2$, when n is an even number, designates the median. More generally, by calling on parenthesized subscripts to denote ascending variate value magnitudes, the order of variate values can be designated. For example, the smallest and largest values are denoted respectively as $X_{(1)}$ and $X_{(n)}$. (In case of a tied value, the counterpart of a coin toss distinguishes between variate designations, say, between $X_{(i)}$ and $X_{(i+1)}$.)

Censored and Incomplete Ordered Measurements

The value placed within some ith order statistic's parenthesized subscript denotes a quantity called a *rank.* Nonparametric inferential methods that are based on ranks often help trade off statistical power and/or efficiency, on the one hand, to gain robustness, in other words, insensitivity to departures from assumptions, on the other. In the context of the many applications that make use of censored samples, order statistic-based methods have been developed that are both robust and efficient. This is accomplished by statistical approaches that differ from ML procedures, such as the well-known Kaplan-Meier (KM) estimator.

The title of Kaplan and Meier's classic paper, "Nonparametric Estimation From Incomplete Observations," refers to incomplete observations, not censored observations. This distinction is central to an understanding of the epidemiological roles that order statistics often play and can be illustrated by a common laboratory experiment. In toxicological studies, measurements are often entered in a three-column spreadsheet as shown in Figure 1. All but the last entry (across from Day 734 in Column A) record the dates when one or more animals die. The last entries record when animals are sacrificed to determine if the targeted tumor is present at the end of the study's data-gathering stage, X_s. In Column B, across from days-survived column entries, the daily number of natural or sacrifice deaths where, at autopsy, the targeted tumor is found is recorded, here as 7. Column C records the number of deaths that day deemed to be non-toxic-substance attributable (because at autopsy there is no trace left of the targeted tumor), here as 14.

Suppose, on the one hand, an animal dies prior to X_s and, at autopsy, no tumor is found. Then, because some nontargeted factor has limited this animal's period of observation, information concerning this animal is said to be *incomplete.* On the other hand, when at or soon after sacrifice it is detected that had the experiment been continued beyond X_s, an animal would have had a good chance of dying due to the targeted tumor, this animal contributes a *censored* measurement.

For example, were all measurements complete and the experiment long enough to assure that more than $n/2$ animals die, then a sample median can provide an estimate of expected survival time. Yet even if all animals die prior to X_s, were there even one

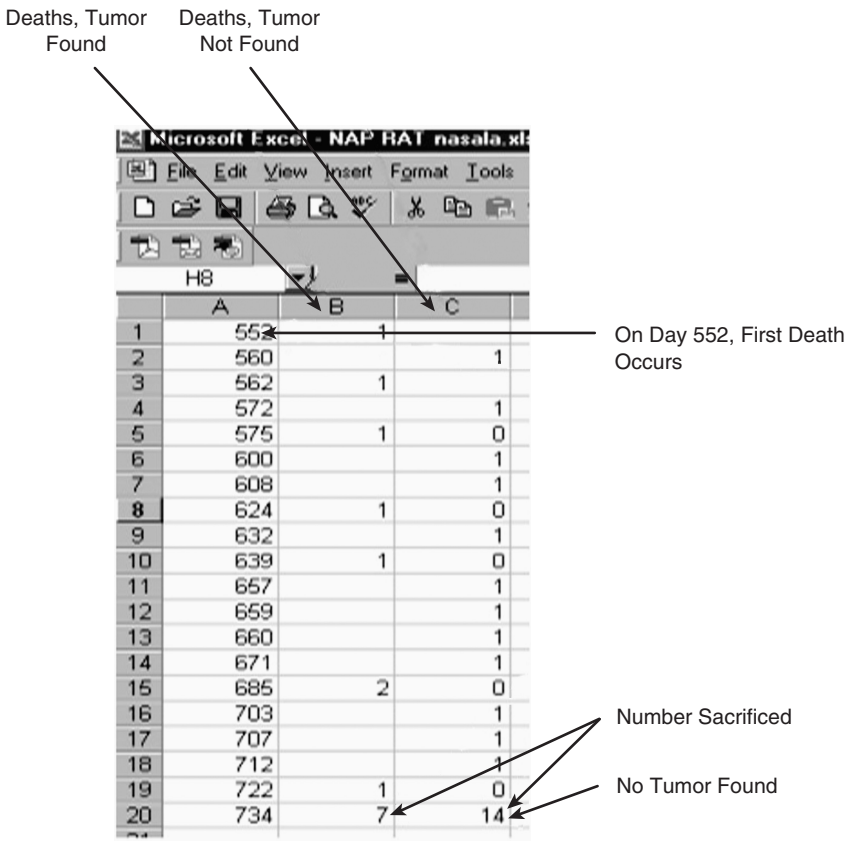


Figure 1 Example of Survival Spreadsheet That Illustrates the Distinction Between Censored and Incomplete Measurements

incomplete measurement, it would follow that KM, life-table, or other specialized methods, not order statistics-based approaches such as the sample median, must be used. Among these approaches, one general class of method has been found to be both computationally simple and, given several reasonable assumptions, efficient in the sense of making optimal use of available data.

Computational and Data Usage Efficiency

Although it plays many other important statistical roles, the median itself is rarely used today to study censored data from animal experiments and other investigations. Unlike the median as a way of working with censored measurements, among their many advantages best linear unbiased (BLU) techniques can help study data from a wide range of experiments and observational investigations. They can be used in any study based on six or more subjects among which, prior to date X_s, a death or some other targeted outcome occurs on 2 or more days. Besides being applicable to a wide range of problems, once one of the many published BLU tables is located, computational procedures require little more than hand calculation.

For statistical models such as the normal, logistic, and exponential, simplicity is attributable to the *L*, linear, part of the BLU designation. Tables are readily available that can be used to calculate BLU estimates. For a user-specified n and a specified degree of censoring, these tables list linear model weights. For example, the sample mean $\overline{X}$ of an uncensored independent and identically distributed normal sample serves a dual role as an ML and a BLU estimator. Since $\overline{X}$ can be computed by multiplying each respective order statistic by the same weight, $1/n$, when $n = 5$, the first tabled sequence of weights lists the value 0.2000 five times.

Any second set of weights listed in a BLU estimator table defines the BLU generalization of an *s* or weighted sample range estimator of the scale parameter σ. In the $n = 5$ example, the weights -0.3724, -0.1352, 0.0000, 0.1352, and 0.3724 are listed below each 0.2000 entry. Correspondingly, for $n = 5$, the estimator $-0.3724\, X_{(1)} - 0.1352\, X_{(2)} + 0.1352\, X_{(4)} + 0.3724\, X_{(5)}$ is the BLU counterpart of the sample standard deviation *s*. Below these uncensored-case BLU estimator weights, censored sample counterpart weights are listed.

Tables are provided by Dixon and Massey that correct for *s*'s bias when the sample standard deviation is calculated from normal data. The *B* of BLU stands for *best* among all linear *unbiased* (LU) estimators. In other words, whether or not it is based on censored data, among LU estimators it makes the best use of data, as judged by the usual mean squared error, efficiency, criterion. The sample standard deviation, *s*, although a biased estimator of the

normal model's scale parameter σ, is slightly more efficient than its BLU counterpart.

Since its computation requires root finding, s is not a linear estimator. Even when it is calculated using uncensored data, any BLU-based estimator of σ is almost as efficient in the normal case as either s, or s's ML counterpart $\sqrt{[(n-1)/n]}s$. The determination of BLU-based estimators of μ and σ that are based on censored data requires even fewer calculations than are required to compute their size n sample uncensored counterparts. Hence, it is surprising that no ML or other statistical technique provides much competition for BLU approaches as a way to study censored samples.

—Michael E. Tarter

See also Bias; Cox Model; Interquartile Range; Kaplan-Meier Method; Life Tables; Nonparametric Statistics; Sample Size Calculations and Statistical Power; Survival Analysis

Further Readings

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A first course in order statistics.* New York: Wiley.

Balakrishnan, N., & Rao, C. R. (1998). *Order statistics: Theory and methods.* Amsterdam: Wiley.

David, H. A. (1970). *Order statistics.* New York: Wiley.

Dixon, W. J., & Massey, F. J. (1983). *Introduction to statistical analysis* (4th ed.). New York: McGraw-Hill.

Greenwood, J. A., & Hartley, H. O. (1962). *Guide to tables in mathematical statistics.* Princeton, NJ: Princeton University Press.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457–481.

Pearson, E. S., & Hartley, H. O. (1972). *Biometrika tables for Statisticians* (Vol. 2). Cambridge, UK: Cambridge University Press.

Sarhan, A. E., & Greenberg, B. G. (1962). *Contributions to order statistics.* New York: Wiley.

Tarter, M. E. (2000). *Statistical curves and parameters: Choosing an appropriate approach.* Natick, MA: A K Peters.

Centers for Disease Control and Prevention

The Centers for Disease Control and Prevention (CDC), a component of the U.S. Department of Health and Human Services, has two primary purposes: to improve people's health in their daily lives and to respond to health emergencies. CDC is the principal agency responsible for improving public health in the United States, and it conducts research and public health interventions both in the United States and globally. CDC headquarters are located in Atlanta, Georgia, with a workforce of more than 8,000 employees in various locations throughout the world. As the name suggests, CDC consists of a number of centers that focus on particular aspects of public health. These include the National Center for Environmental Health/Agency for Toxic Substances and Disease Registry, the National Center for Injury Prevention and Control, the National Center for Health Statistics, the National Immunization Program, and the National Institute for Occupational Safety and Health.

The roots of the CDC lie in the Malaria Control in War Areas (MCWA) program, whose mission was to control or prevent malaria and murine typhus fever in the southern United States during World War II. The Communicable Disease Center, the direct descendant of the MCWA, was organized in Atlanta, Georgia, on July 1, 1946. The role of the new Communicable Disease Center was much expanded from that of the MCWA; it was responsible for researching and controlling all communicable diseases except tuberculosis and venereal disease, which at that time were handled through separate offices located in Washington, D.C. Over the years, CDC has expanded to include many agencies addressing specific diseases and health issues, including venereal disease (1957), tuberculosis (1960), immunization (1960), quarantines (1967), and smoking (1986). CDC's purview has expanded to include all diseases and conditions that affect human health, including chronic diseases and related risk factors such as obesity, tobacco use, and exposure to environmental toxins. The name *Center for Disease Control* was adopted in 1970; in 1981, this became *Centers for Disease Control* and in 1992, *Centers for Disease Control and Prevention*; however, the acronym CDC is still used for the organization.

CDC has played an important role in several of the major infectious disease issues of the 20th century. For instance, when some children inoculated with the Salk vaccine were infected with polio, the national inoculation program was halted; however, CDC was able to trace the cases to contaminated vaccine from a laboratory in California, and the inoculation program was resumed. Guidelines for national influenza vaccination were developed after CDC used surveillance procedures to trace the course of the 1957 influenza

epidemic. CDC established a smallpox surveillance unit in 1962, developed improved vaccine and vaccination techniques, and established surveillance procedures adopted by the World Health Organization in its campaign for the global eradication of smallpox. The first diagnosis of AIDS was described in the June 15 issue of *Morbidity and Mortality Weekly Report* (*MMWR*), published by CDC. CDC investigators traced toxic shock syndrome to a particular brand of tampon, which was subsequently removed from the marketplace.

The Epidemic Intelligence Service (EIS), headquartered at the CDC Atlanta offices, consists of physicians, scientists, and other public health professionals who have been trained at the CDC in applied epidemiology. EIS was founded in 1951 after the outbreak of the Korean War, in response to a perceived threat of biological warfare; its role today is to provide experts in surveillance and response to epidemics, and its purview includes chronic diseases and injuries as well as infectious diseases. Today, EIS officers serve in a variety of locations, including CDC offices and state and local health departments, and are available to respond to requests for epidemiological assistance, including investigation of disease outbreaks, all over the world.

MMWR, established in 1961, is a weekly online and print journal published by the CDC, which contains primarily current reports of disease occurrence and risk factors. Each week includes a section of notifiable diseases, and the remainder of the articles are on various topics, including reports of disease outbreaks and focused analyses of publicly available data sets such as the National Health Interview Survey (NHIS). Electronic copies of *MMWR* dating back to 1982 are available through the CDC Web site or by e-mail subscription without charge, and paper copies may be purchased through the same Web site.

See also Behavioral Risk Factor Surveillance System; Field Epidemiology; Governmental Role in Public Health; National Center for Health Statistics; National Immunization Survey; Outbreak Investigation

—Sarah Boslaugh

Further Readings

Etheridge, E. W. (1992). *Sentinel for health: A history of the Centers for Disease Control.* Berkeley: University of California Press.

Office of the Directory, Epidemiology Program Office, Centers for Disease Control and Prevention. (1996). Historical perspectives: History of CDC. *Morbidity and Mortality Weekly Reports*, *45*(25), 526–530. Retrieved October 4, 2006, from www.cdc.gov/mmwr/preview/mmwrhtml/00042732.htm.

Web Sites

Centers for Disease Control and Prevention: http://www.cdc.gov.

Central Limit Theorem

A series of theorems in mathematical statistics called the central limit theorems provide theoretical justification for approximating the true sampling distribution of many sample statistics with the normal distribution. This entry discusses one such theorem for the sample mean. Similar theorems exist for sample median, sample standard deviation, and sample proportion. The word *central* in the name of the theorem means ''fundamental.'' The central limit theorem for the sample mean states that for a large sample size, the sampling distribution of the sample mean $\overline{X}$ is approximately normal, no matter what the population distribution looks like. The approximation becomes better with increasing sample size. This surprising fact was proved in fairly general form in 1810 by Pierre-Simon Laplace.

The graphs in Figures 1 through 5 show the idea of the Central Limit Theorem.

Figure 1 is a histogram of a random sample data from a normally distributed population with a mean 50 and a standard deviation 2; Figure 2 is a histogram for a sampling mean $\overline{X}$ with a sample size $n = 2$ from a normal distribution with a mean 50 and a standard deviation 2. Even the sample size is small; in this case, $\overline{X}$ is still normal with a mean 50 and a standard deviation $2/\sqrt{2} = 1.4142$. In fact, when the population is normal, the sampling distribution of $\overline{X}$ is exactly normal for any sample size.

Figure 3 is a histogram of a random sample data from a chi-square distribution with 1 *df*. This is a right-skewed distribution; Figure 4 is a histogram for a sampling mean $\overline{X}$ for a sample size 2 from a chi-square distribution with 1 *df*. When sample size is small, the sampling distribution of $\overline{X}$ is still right-skewed; Figure 5 is a histogram for a sampling mean

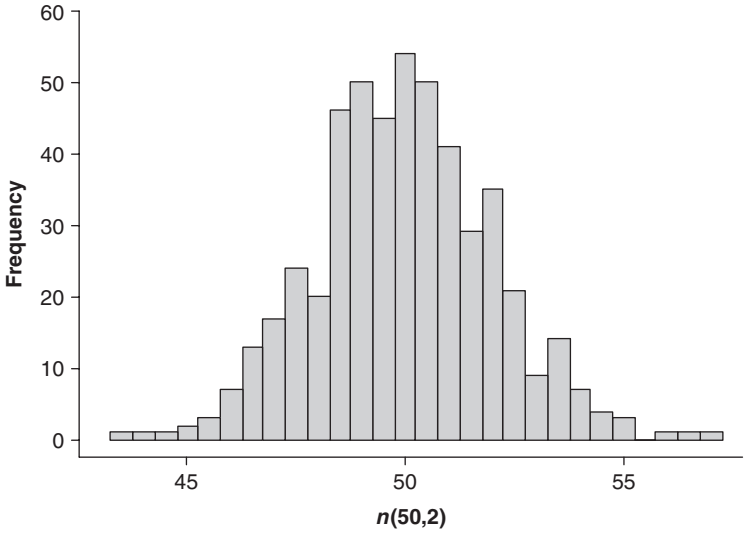


Figure 1 A Histogram of a Random Sample Data From a Normally Distributed Population With a Mean 50 and a Standard Deviation of 2

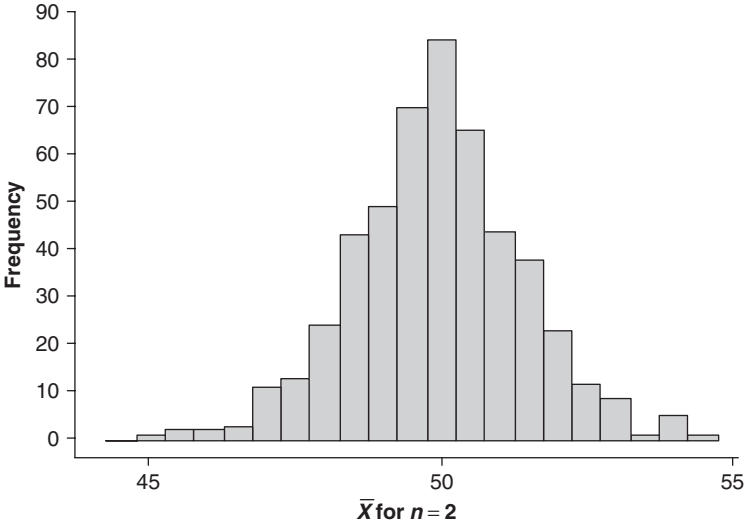


Figure 2 A Histogram for a Sampling Mean $\overline{X}$ With a Sample Size $n = 2$ From the Same Normal Distribution With a Mean 50 and a Standard Deviation 2

$\overline{X}$ for a sample size 30 from the same chi-square distribution with 1 *df*. In this case, the sampling distribution of $\overline{X}$ is approximately normal.

In symbol, let X be a random variable with mean μ and standard deviation σ, and $\overline{X}$ be the sample mean; when the sample size n is large, the standardized variable Z is approximately the standard normal variable:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx N(0, 1).$$

This result has enabled statisticians to develop some large-sample procedures for making inferences about a population mean μ even when the shape of the population distribution is unknown. Application of the

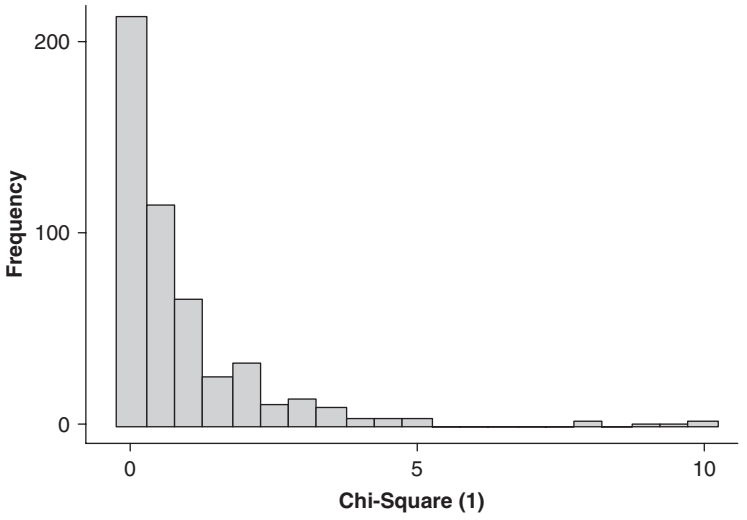


Figure 3 A Histogram for a Random Sample Data From a Chi-Square Distribution With 1 *df*

Note: This is a right-skewed distribution, that is, values near 0 are most common, and the probability of a given value decreases as the value increases.

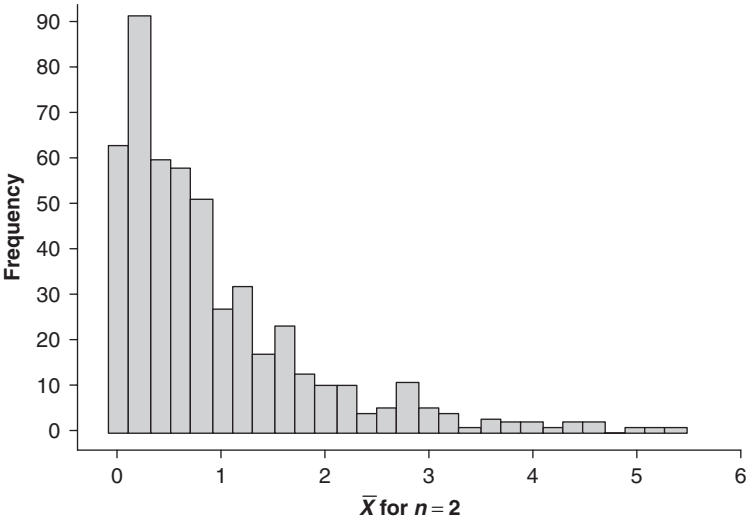


Figure 4 A Histogram for a Sampling Mean $\overline{X}$ With a Sample Size 2 From a Chi-Square Distribution With 1 *df*

central limit theorem requires a rule of thumb for deciding whether n is indeed sufficiently large. When the population distribution is quite skewed, only the ones for $n = 30$, 40, or more, the sampling distribution of $\overline{X}$ may have reasonably normal shapes. If the population distribution is somewhat skewed, then $n = 10$ or 15, and the sampling distribution of $\overline{X}$ may have reasonably normal shapes. The rule that many statisticians recommend is $n \geq 30$.

Example

Suppose that a random sample of size 64 is to be selected from a population with mean 40 and standard

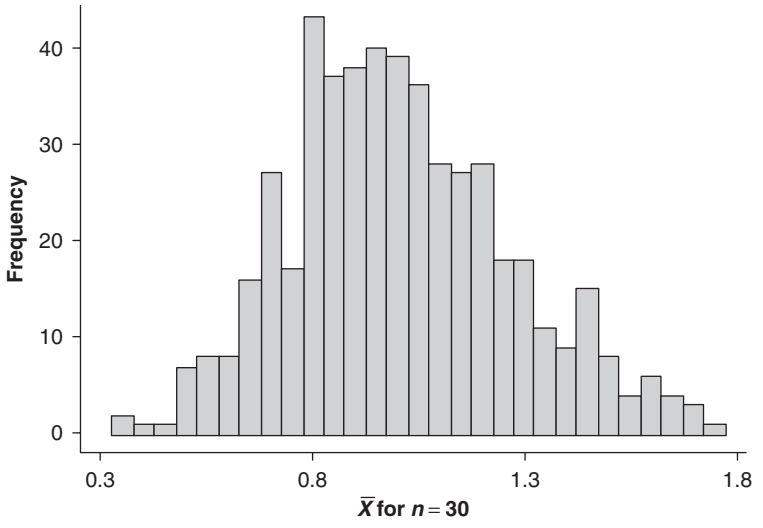


Figure 5 A Histogram for a Sampling Mean $\overline{X}$ With a Sample Size 30 From a Chi-Square Distribution With 1 *df*

deviation 5. What is the approximate probability that $\overline{X}$ will be within 0.5 of the population mean μ?

Solution: Since $\mu = 40$, and $\sigma = 5$, so that

$$\frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{64}} = 0.625$$

$$\begin{aligned} &P(-0.5 < \bar{x} - \mu < 0.5) \\ &= p\left(\frac{-0.5}{0.625} < \frac{\bar{x} - \mu}{0.625} < \frac{0.5}{0.625}\right) \\ &\approx p(-0.8 < z < 0.8) \\ &= 0.7881 - 0.2119 = 0.5762. \end{aligned}$$

The approximate probability that the sample mean $\overline{X}$ will be within 0.5 of the population mean μ is about 57.62%.

—Renjin Tu

See also Measures of Central Tendency; Nonparametric Statistics; Normal Distribution; Sampling Distribution; Study Design; *Z* Score

Further Readings

De Veaux, R. D., Velleman, P. F., & Bock, D. E. (2005). *Stats data and models.* Boston: Addison-Wesley.

Johnson, R. A., & Bhattacharyya, G. K. (2006). *Statistics principles and methods* (5th ed.). Danvers, MA: Wiley.

Child Abuse

Child abuse and neglect remains a significant public health and social problem in the United States. Epidemiologic methods provide a systematic approach for surveillance, determination of risk factors, and estimates of service delivery needs on the basis of prevalence and incidence estimates. During the past decade, the information provided by many multidisciplinary studies has increased awareness about child maltreatment, improved the treatment of families and offenders, and promoted stricter law enforcement. This work is at the base of the current trend that shows a decline in child abuse and neglect, with 38,000 less children abused in 2004 compared with the previous year, according to "Child Maltreatment 2004," a report from the Children's Bureau, U.S. Department of Health and Human Services (2006). This entry reviews the definitions of physical abuse, sexual abuse, and neglect; discusses prevalence (number of victims in a population at any point in time with a maltreatment experience that occurred recently or a long time ago) and incidence data (number of new cases occurring in the population during a given

period of time) for child abuse in the United States; and identifies risk factors for child abuse.

Definitions of Child Abuse

In the United States, federal legislation establishes the minimum standards for the definition of child abuse and neglect that states must incorporate as part of their definitions. The Federal Child Abuse Prevention and Treatment Act (CAPTA) (42 U.S.C.A. §5106g) defines child abuse and neglect as

> a recent act or failure to act on the part of a parent or caretaker which results in death, serious physical or emotional harm, sexual abuse or exploitation; or an act or failure to act which presents an imminent risk of serious harm. (Child Welfare Information Gateway, 2005, p. 1)

Some states define child abuse and neglect as a single concept, while other states include definitions for several categories of abuse and neglect. Physical abuse, sexual abuse, and neglect are included in definitions of child abuse in all states. Physical abuse is generally defined as a nonaccidental physical injury to the child or any action that results in a physical impairment. Sexual abuse includes

> the employment, use, persuasion, inducement, enticement, or coercion of any child to engage in, or assist any other person to engage in, any sexually explicit conduct or simulation of such conduct for the purpose of producing a visual depiction of such conduct; or the rape, and in cases of caretaker or interfamilial relationships, statutory rape, molestation, prostitution, or other form of sexual exploitation of children, or incest with children. (Child Welfare Information Gateway, 2005, p. 1)

Neglect generally encompasses failure to provide for no apparent financial reason (deprivation of adequate food, clothing, shelter, medical care) and lack of supervision. The majority of states (except Georgia and Washington) include emotional abuse in their definitions, generally defined as damage to the psychological capacity or emotional stability of the child (evidenced by symptoms such as anxiety, depression, aggressive behavior, withdrawal, and other substantial changes in behavior, cognition, or emotional response). Abandonment is included by many states either as part of their definition of neglect (18 states) or as a separate definition (13 states), and it includes situations

> when the parent's identity or whereabouts are unknown, the child has been left by the parent in circumstances where the child suffers serious harm, or the parent has failed to maintain contact with the child or to provide reasonable support for a specified period of time. (Child Welfare Information Gateway, 2005, p. 2)

Some states include in their definition of child abuse or neglect the abuse of substances by a parent, generally defined as the manufacturing of drugs in the presence of a child; using drugs in the presence of a child; selling, distributing, or giving drugs or alcohol to a child; and use of drugs that impairs the caregiver's ability to adequately care for the child (Child Welfare Information Gateway, 2005).

In the definition of child abuse and neglect, a "child" means a person less than 18 years old. The persons responsible for the child and reportable under the civil child abuse reporting laws to child protective services (CPS) include parents, guardians, foster parents, relatives, or other caretakers responsible for the child.

Prevalence

According to "Child Maltreatment 2004," a national report based on data submitted annually by the states to the federal government, an estimated 872,000 children in the United States were determined to be victims of child abuse or neglect in 2004, with a rate of victimization of 11.9 per 1,000 children. Girls were slightly more likely to be abused, while children below 3 years old had the highest rate of victimization (16.1 per 1,000 children of the same age group). The majority of children were victims of neglect (more than 60%), almost 18% suffered physical abuse, 10% were sexually abused, and 7% were emotionally abused. An estimated 1,490 children died in 2004 due to abuse or neglect, and more than 80% of them were younger than 4 years old. Approximately 80% of the persons responsible for the abuse or neglect were parents; of those, almost 60% were female (mostly the mother) (Children's Bureau, 2006).

A new source of prevalence data is the National Survey of Child and Adolescent Well-Being (NSCAW), the first national probability study of children investigated

for child abuse and neglect, which included a sample of 5,501 children (ages 0 to 14) who were randomly selected from the families who entered the U.S. child welfare system between October 1999 and December 2000. NSCAW baseline's report showed a higher prevalence of physical abuse (27%) than the prevalence reported in "Child Maltreatment 2004" for the same period (for the year 2000 the prevalence was 19%, almost the same as 2004). NSCAW reported that for almost 47% of children, the most serious type of abuse was neglect (27% failure to supervise, 20% failure to provide), 11% were victims of sexual abuse, 7% of emotional maltreatment, and 1.6% of abandonment (NSCAW, 2005).

Incidence

The National Incidence Study (NIS) of child maltreatment is a national survey of sentinel reporters from randomly selected counties across the nation. These sentinels are professionals from different fields, including education, health, social work and law, who submit data forms on any children who were maltreated during the study period, regardless of whether the case has been reported to CPS. NIS includes more than 5,600 professionals serving 42 counties. The last NIS (NIS-3) reported that in 1993, birth parents accounted for 78% of the maltreatment under the NIS-3 harm standard (that requires that an act or omission results in demonstrable harm to be classified as abuse or neglect). Of the children maltreated by birth parents, 75% were maltreated by their mothers and 46% by their fathers. The pattern of abuse differs by the gender of the alleged perpetrator; in general, children tended to suffer more neglect from women and more abuse from men. Mothers were responsible for 60% of the cases of physical abuse, 28% of the cases of sexual abuse, 55% of the cases of emotional abuse, 93% of the cases of physical neglect, 78% of the cases of emotional neglect, and 86% of the cases of educational neglect (Sedlack & Broadhurst, 1996).

A second source of child maltreatment incidence data is the Gallup nationwide survey of parents conducted in 1995. In this survey, 1,000 parents self-reported discipline methods used with their children during the past 12 months using the Parent-Child Conflict Tactic Scales (CTSPC). The CTSPC measures whether a parent has carried out specific acts of physical and psychological aggression regardless of whether the child was injured. Approximately half of the parents reported using corporal punishment including spanking on bottoms with bare hands (47%) and slapping on the hand, arms, or leg (37%). A smaller proportion reported severe physical assault such as shaking a child below 2 years (4%), beating them up (0.2%), and throwing or knocking them down (0.2%). Neglect was more prevalent than severe physical assault. Parents reported that they left the child alone even when an adult should be with him or her (20%), were not able to make sure that the child received the food he or she needed (11%), and were so drunk that they had problems taking care of the child (2.3%). The Gallup survey found that mothers have a higher rate of corporal punishment when children are younger. The rate of severe physical assault for mothers was more than double that of fathers (mothers 5.91 per 1,000, fathers 2.58 per 1,000), but the difference was not statistically significant (Straus, Hamby, Finkelhor, Moore, & Runyan, 1998).

Risk Factors

A number of risk factors have been associated with child maltreatment and its recurrence. In terms of the family characteristics, risk factors include family composition (single parents are more likely to abuse or neglect their child), caregiver substance abuse, domestic violence, mental health problems of caregivers, young maternal age, and maternal history of abuse. In terms of the maltreatment characteristics, cases involving neglect are most likely to recur, followed by cases involving physical abuse; cases of sexual maltreatment are the least likely to recur. Regarding children's ages, general reports of child abuse and neglect recurrence have shown that younger children are more likely to experience recurrence compared with older children. One hypothesis for this difference is that older children (12 to 17 years old) are more likely to age out of the system before a subsequent event of maltreatment. Children's health status has also been associated with the risk of abuse and neglect. High-risk infants, children with developmental delays, and children with handicaps who may be frequently irritable, difficult to soothe, unresponsive, or who reject holding are at higher risk of victimization (Appel & Holden, 1998; Barnett, Miller-Perrin, & Perrin, 1997; Edleson, 2001).

—Cecilia Casanueva

See also Child and Adolescent Health; Violence as a Public Health Issue

Further Readings

Appel, A. E., & Holden, G. W. (1998). The co-occurrence of spouse abuse and physical child abuse: A review and appraisal. *Journal of Family Psychology*, *12*(4), 578–599.

Barnett, O., Miller-Perrin, C., & Perrin, R. (1997). *Family violence across the life span*. London: Sage.

Child Welfare Information Gateway. (2005). *Definitions of child abuse and neglect*. Retrieved June 28, 2007, from http://www.childwelfare.gov/systemwide/laws_policies/statutes/define.cfm.

Children's Bureau, U.S. Department of Health and Human Services. (2006). *Child maltreatment 2004*. Washington, DC: Government Printing Office.

Edleson, J. L. (2001). Studying the co-occurrence of child maltreatment and domestic violence in families. In S. A. Graham-Bermann & J. L. Edleson (Eds.), *Domestic violence in the lives of children: The future of research, intervention, and social policy*. Washington, DC: American Psychological Association.

Federal Child Abuse Prevention and Treatment Act, 42 U.S.C.A. § 5106g, amended by the Keeping Children and Families Safe Act of 2003, P.L.108-36. (1998).

National Survey of Child and Adolescent Well-Being. (2005). *CPS sample component wave 1 data analysis report*. Retrieved June 28, 2007, from http://www.acf.hhs.gov/programs/opre/abuse_neglect/nscaw/reports/cps_sample/cps_report_revised_090105.pdf.

Sedlack, A. J., & Broadhurst, D. D. (1996). *The Third National Incidence Study of Child Abuse and Neglect (NIS-3)*. Washington, DC: National Clearinghouse on Child Abuse and Neglect Information.

Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W., & Runyan, D. (1998). Identification of child maltreatment with the Parent-Child Conflict Tactics Scales: Development and psychometric data for a national sample of American parents. *Child Abuse & Neglect*, *22*(4), 249–270.

Child and Adolescent Health

The 20th century saw important and dramatic progress in child and adolescent health. With the recognition of child health as a distinct field and the development of pediatrics as a specialty of medicine, infections have been successfully treated and prevented, congenital anomalies have been classified and ameliorated, and newborn care has markedly increased the number of successful outcomes for premature or ill infants. Child development and behavioral studies have allowed for innovations in injury prevention, more effective parenting, and therapy of mental health issues for the young. This entry reviews assessment of child and adolescent health with special emphasis on prevention and discusses the roles of developmental disabilities, injury, and chronic diseases in child health.

Child Health

Routine, periodic health supervision visits are the cornerstone of preventive care for infants and children among pediatric health care workers worldwide.

Newborn Screening

In the past 50 years, developed nations have successfully implemented population-wide newborn laboratory screening programs that detect conditions such as phenylketonuria, sickle cell disease, and hypothyroidism early enough that treatment can prevent or ameliorate many of the consequences of the disease. Dozens of conditions can now be identified within days of birth using a few drops of blood; the large scale of the programs allows ultimate savings of money and human life for recognition of even very rare metabolic, hormonal, and other disorders.

Growth Surveillance

Monitoring child growth parameters of height, weight, and head circumference allows early warning and intervention in problems such as malnutrition, congenital malformations, chronic infection, and other chronic diseases. Such monitoring can be done easily and with minimal cost; it provides invaluable information to guide further care. Optimal maternal nutrition and prenatal care are foundations for infant growth; prevention of premature birth is one of the most important challenges in health care today. Human milk is the optimal food for newborns.

Poor growth is evaluated based on deviations from the expected growth curve or trajectory. For example, infants with poor weight gain, but normal height and head circumference growth, may not be receiving adequate caloric intake. They may suffer starvation in cases of dire poverty or maternal neglect, or they may not absorb the nutrition that is provided (infectious diarrhea)—this is often called "failure to thrive." Children with suboptimal weight and height, but relatively normal head circumference, may have had

prolonged protein calorie deprivation or a chronic medical condition such as chronic renal failure. In the United States, the incidence of overweight and obesity now exceeds failure to thrive and undernutrition.

Immunizations

Prevention of infectious disease by childhood immunization is one of the triumphs of medical science. Smallpox has been eradicated. Poliomyelitis may be near worldwide eradication. At present, effective immunization against tetanus, diphtheria, pertussis, hepatitis A and B, influenza, measles, mumps, rubella, varicella, hemophilus influenza B, and pneumococcus have reduced or nearly eliminated these conditions in many geographic areas. Newer immunizations against rotavirus, meningococcus, and human papillomavirus have recently begun widespread use. In the near future, other infections will also be targeted. Unfortunately, vaccination programs worldwide are often hindered by political, economic, and cultural barriers as well as practical issues such as requirements for vaccine refrigeration during storage.

Developmental Assessment

Developmental assessment ensures that children are progressing on appropriate developmental trajectories or that they receive appropriate support if there are areas of concern. Many tools can be used to assess developmental progress. One of the most important roles of the care provider is to listen to parents and address their concerns with appropriate screening tools, evaluation, and referrals to the appropriate subspecialists for further evaluation.

Developmental assessment should consist of a review of the child's progress in motor, cognitive, and social/emotional streams of development. Screening tools can be helpful to assure that all areas of development are addressed. Patterns of developmental delay are helpful both to clarify areas of concern and to assist with diagnosis.

Social Assessment

Safety/Abuse

Child abuse and neglect are worldwide, transcultural phenomena that can be recognized by perceptive and well-trained health care workers, treated with medical and psychological methods, and effectively prevented with parental education and support from home health visitors. Physical abuse often results when isolated, poorly prepared caretakers face predictable developmental crises such as crying and toilet training. Neglect of basic needs may indicate emotional or intellectual problems in the family. Both physical abuse and neglect more often occur in the settings of poverty, substance abuse, domestic violence, and other stressors. Sexual abuse of children may be incestuous, may involve family acquaintances, or may be related to pedophilia and child prostitution; a myriad societal and legal strategies will be needed to prevent these incidents.

Parenting

Child development and discipline issues are common. As fewer extended families live together, the common intergenerational knowledge base of normal development and effective child-raising techniques is replaced by media information, professional opinion, and information gained in schools. Scientific attention to parenting greatly increased over a century ago with psychoanalytic investigations by Sigmund Freud and Carl Jung and, later, the behavioral modification theories of B. F. Skinner and others. At mid-century and beyond, popularizers of parenting advice such as Benjamin Spock and Berry Brazelton advised a child-centered, relaxed parenting style. Currently, cognitive science and genetic ideas influence popular understanding of parenting practice and effects.

Adolescent Health

Growth/Puberty

Adolescence is marked by the rapid growth to adult height, development of secondary sexual characteristics and ability to reproduce, and psychological and cognitive changes leading to maturity. As developed societies require longer years of education before economic independence from parents, behavioral adolescence is extended to the 20s while, in a not fully understood trend, the onset of puberty (pubarche) has seemed to begin earlier.

Social Assessment

Adolescent health assessment importantly includes the adolescent's friends, social milieu, risky behaviors and habits, psychological state, academic and vocational progress, and family setting. Such a global view reflects

the reality that in the United States, adolescent mortality and serious morbidity are most often due to consequences of unintentional injury, homicide, suicide, sexual behaviors, and drug and alcohol abuse. Furthermore, many of these behaviors, once established in adolescence, are the basis for later problems in adult life. Teenage pregnancy, more common in underprivileged populations, makes continued educational and vocational successes less probable.

Developmental Disabilities

Developmental disabilities are a diverse group of severe chronic conditions that are associated with cognitive and/or physical impairments. People with developmental disabilities may have difficulty with communication, mobility, learning, self-care, and independent living. Developmental disabilities begin anytime from birth through adolescence and usually last throughout a person's lifetime.

Cognitive Developmental Disabilities

Important cognitive developmental disabilities include mental retardation and autism. Mental retardation is characterized both by a significantly below-average score on a test of mental ability or intelligence and by limitations in the ability to function in areas of daily life, such as communication, self-care, and getting along in social situations and school activities. Children with mental retardation can and do learn new skills, but they develop more slowly than children with average intelligence and adaptive skills. There are different degrees of mental retardation, ranging from mild to profound. A person's level of mental retardation can be defined by their intelligence quotient (IQ) and by the types and amount of support they need.

Children with autism spectrum disorders (ASDs) have significant impairments in social interaction and communication and exhibit unusual behaviors and interests. Many people with ASDs also have unusual ways of learning, paying attention, or reacting to different sensations. The cognitive abilities of people with ASDs can vary from profoundly mentally retarded to gifted ability levels.

Motor Developmental Disabilities

Motor developmental disabilities include cerebral palsy and neuromuscular disorders. *Cerebral palsy* is an umbrella term covering a group of nonprogressive, but often changing, motor impairment syndromes secondary to lesions or anomalies of the brain arising in the early stages of its development. Children and adolescents with cerebral palsy have alterations in muscle tone that make it difficult to maintain posture or be independently mobile.

Neuromuscular disorders can be caused by alterations in the function anywhere in the neuromuscular system from the brain, spinal cord, anterior horn cells, nerves, neuromuscular junction, and muscles. Prognosis and treatment depend on the cause and location of the abnormality. Examples of neuromuscular disorders include muscular dystrophy, Guillain-Barré syndrome, myasthenia gravis, and peripheral neuropathies.

Injury

As medical advances have decreased the effects of many diseases, the recognition of injuries as a major cause of child and adolescent morbidity and mortality has engendered injury control and prevention efforts. As one example, child car seats have saved many children from avoidable death or injury. Awareness of unintentional injury allows the possibility of even more prevention, while trauma care centers have improved survival among the injury victims. Intentional injury, including homicide and suicide, continues to be a major source of trauma; prevention efforts may involve better screening for depression, alcohol and drug treatment programs, and community education efforts.

Brain Injury

Brain injury is the most common cause of acquired disability in childhood, and most primary care physicians will encounter a child who has experienced brain injury. It is difficult to know the exact numbers of children with brain injury because children with milder injuries may not present for evaluation. However, almost all children with severe brain injury are seen in emergency departments prior to referral for further care. Between 100,000 and 200,000 children are hospitalized with brain injury every year in the United States. Of the children hospitalized with brain injury, many will have long-term disability.

Traumatic brain injury is most commonly caused by motor vehicle collisions in teenagers and by falls and physical abuse in younger children. Nontraumatic brain injury is caused by metabolic disorders (organic

acidemias, fatty acid oxidation defects), cerebral vascular accidents (stroke), anoxic injury (cardiac or respiratory arrest, near drowning, near hanging), central nervous system tumors, and hypoxic seizures.

Many of the causes of brain injury are preventable, and improvements in car safety have already made great improvements in decreasing the morbidity and mortality of brain injury. Increased awareness of the dangers of all-terrain vehicles and the risks associated with impaired driving will continue to decrease the incidence of new injury.

Spinal Cord Injury

Spinal cord injury (SCI) is due to a trauma causing a contusion or a partial or complete transaction of the spinal cord. SCI is a common cause of permanent disability and death in children. About 11,000 people a year sustain a SCI, and more than half of SCIs occur among young people between the ages of 16 and 30 years. The majority of SCI victims are boys and young men.

Like brain injury, many incidents of SCI are preventable with improved car safety and the use of appropriate protective gear during sports activities. Gunshot wounds and stab wounds are also both common and preventable causes of spinal cord injury.

Chronic Diseases

During the epidemiologic transition of the last century, major causes of mortality in the developed world shifted from infectious diseases to degenerative processes such as cancer and heart disease. In the childhood and adolescent years, a similar shift occurred away from now preventable or treatable infections to prematurity or congenital conditions. Children with special health care needs, including cancer, asthma, and consequences of prematurity, are the focus of much tertiary pediatric care and require substantial support services, including in-home therapies and hospice care.

Obesity/Diabetes Mellitus

The epidemic of obesity and its complications (including type 2 diabetes mellitus) reflect nutritional abundance and a mismatch between energy intake and expenditure. Medical and surgical approaches to this problem are inadequate; societal changes will be necessary for its solution. All organ systems are adversely affected, and early onset of obesity portends a more problematic course in later life.

Asthma

An improved armamentarium of pharmacologic treatments has made control of asthma possible for the great majority of affected children. Still, many children have restricted lifestyles with this illness for a variety of reasons: Adherence to suggested treatments is often incomplete due to inadequate patient or parent education, availability of care is limited by lack of health insurance, environmental factors are incompletely controlled (pollution, plant allergens, home dust, and animal danders). The roles of heredity, early antigen exposures, and infant nutrition in the genesis of asthma are currently undergoing intensive study.

Cancer

Although an uncommon childhood disease, cancer is a major cause of childhood mortality and morbidity. Leukemia is the most common childhood malignancy and, in the case of acute lymphoblastic leukemia, the most dramatic success story, with cure attained in more than 70% of children. Other cancers have proved less susceptible to the chemotherapy, radiation, and surgery forms of therapy being continually studied and improved.

Infections

Successful treatment and, even more significantly, prevention of childhood infections has been the most dramatic success story in child and adolescent health. Smallpox has been eradicated from the world; polio transmission occurs only in isolated pockets in the world. Diphtheria, tetanus, pneumococcal, hemophilus influenzae, measles, and chickenpox (varicella) vaccines have drastically reduced the incidence of these illnesses in the developed world; efforts in less developed areas have the potential for similar lifesaving results. Efforts to better control and prevent malaria through insect control, medicated sleep netting, medication, and development of a vaccine could result in substantially bettered childhood morbidity and mortality. Control and treatment of other parasitic diseases, tuberculosis, and human immunodeficiency virus are also acute needs.

Mental Illness

As life-threatening infections and other illnesses threatening children have diminished, more attention has been given toward child mental health and what Robert Haggerty (1998) called "the new morbidity" of developmental and psychosocial concerns (p. 1327). Attention deficit hyperactivity disorder is the most common early childhood diagnosis; psychostimulant medication and counseling help significantly. Depression affects children and adolescents as well as adults; suicide remains a leading cause of death in developed countries among teenagers, and prevention efforts are not yet systematized to acceptable levels of effectiveness. Eating disorders such as anorexia nervosa and bulimia are uncommon but serious problems in teenagers.

—William N. Marshall and Sydney Rice

See also Child Abuse; Chronic Disease Epidemiology; Genetic Disorders; Injury Epidemiology; Newborn Screening Programs

Further Readings

American Academy of Pediatrics Committee on Psychosocial Aspects of Child and Family Health. (2001). The new morbidity revisited: A renewed commitment to the psychosocial aspects of pediatric care. *Pediatrics*, *108*, 1227–1230.

Batshaw, M. L. (Ed.). (2002). *Children with disabilities* (5th ed.). Baltimore: Paul H. Brookes.

Behrman, R. E., Kliegman, R. M., & Jenson, H. B. (Eds.). (2004). *Nelson textbook of pediatrics* (17th ed.). Philadelphia: W. B. Saunders.

Bryce, J., Boschi-Pinto, C., Shibuya, K., Black, R. E., & WHO Child Health Epidemiology Reference Group. (2005). WHO estimates of the causes of death in children. *Lancet*, *365*, 1147–1152.

Guyer, B., Freedman, M. A., Strobino, D. M., & Sondik, E. J. (2000). Annual summary of vital statistics: Trends in the health of Americans during the 20th century. *Pediatrics*, *106*, 1307–1317.

Haggerty, R. J. (1998). International perspectives in early emotional development. *Pediatrics*, *102*, 1327–1329.

Jones, G., Steketee, R. W., Black, R. E., Bhutta, Z. A., Morris, S. S., & Bellagio Child Study Group. (2003). How many child deaths can we prevent this year? *Lancet*, *362*, 65–71.

Olness, K. (2003). Effects on brain development leading to cognitive impairment: A worldwide epidemic. *Journal of Developmental & Behavioral Pediatrics*, *24*(2), 120–130.

Chi-Square Test

The chi-square test is used for categorical data. There are three situations in which one can use the chi-square test: to test for independence, for equality of proportions, and for goodness of fit. The test statistic and the expected values for the first two cases are identical, but the hypotheses and sampling situations are distinctly different. For the goodness-of-fit situation, the test statistic and expected values are similar but not identical, and the hypotheses and sampling situation are also different from the other cases. The first two cases are presented first. All data tables were created by the author for this entry.

Test for Independence

Situation. One has a single random sample, and this sample is cross-categorized by two variables each with 2 or more categories. The null hypothesis is that there is no relationship between Variable 1 and Variable 2, that is, the two variables are independent of each other. The alternative hypothesis is that there is a relationship between the two variables, that is, the two variables are dependent.

For example, one has a random sample of people who are cross-categorized by race and blood type. The null hypothesis is that there is no relationship between race and blood type, and the alternative hypothesis is that there is a relationship between race and blood type.

Test for Equality of Proportions

Situation. One has a single variable of interest with 2 or more categories and multiple independent random samples. The null hypothesis is that there is no difference in the proportion of each category across the different populations. The alternative hypothesis is that there is at least one difference in a proportion across the different populations.

For example, one has four treatment groups. These treatment groups are the multiple independent random samples. The variable of interest has two categories *survived* or *died.* The null hypothesis is that the proportions that survived are the same for each treatment group and the proportion that died are the same for each treatment group. The alternative hypothesis is

that at least one proportion is different between the treatment groups. The test for equality of proportions is analogous to the one-way analysis of variance test for equality of means for quantitative data.

Expected Values and Tests Statistic: Independence and Equality of Proportions

With the tests for independence and equality of proportions, the data should be given in table form where i represents the ith row and j represents the jth column of the table. One finds the expected value (E_{ij}) for each ijth cell. The expected value for the ijth cell is the value that one would *expect* if the null hypothesis is true.

$$E_{ij} = \frac{i\text{th row total} \times j\text{th column total}}{\text{grand total}}.$$

Reasoning Behind the Expected Value for Test for Independence

For the test for independence, the null hypothesis states that Variable 1 and Variable 2 are independent. By independence,

$$\begin{aligned} &P[i\text{th row category}] \cdot P[j\text{th column category}] \\ &\quad = P[i\text{th row category} \cap j\text{th column category}] \\ &\quad = P[ij\text{th cell}]. \end{aligned}$$

So the best estimate of the $P[ij$th cell] is the estimate of the proportion of the ith row multiplied proportion of the jth column, that is,

$$\frac{i\text{th row total}}{\text{grand total}} \times \frac{j\text{th column total}}{\text{grand total}}.$$

To find the expected *count*, one would multiply by the grand total, which gives the above formula for E_{ij}.

Reasoning Behind the Expected Value for Test for Equality of Proportions

For the test for equality of proportions, the null hypothesis states that the proportion of each category is the same across the different populations. Let the columns represent the different samples representing the populations and the rows represent the categories of the variable of interest. Consider the ith category. The best estimate of the probability of the ith category is ith row total/grand total. To find the expected *count* for the jth column, one would multiply by the jth column total, which gives the above formula for E_{ij}.

Test Statistic

The observed value (O_{ij}) denotes the observed count in the ijth cell. The test statistic for both cases is $\chi^2 = \sum_i \sum_j \left(O_{ij} - E_{ij}\right)^2 / E_{ij}$ and degrees of freedom, $df = (r-1)(c-1)$ where r is the number of rows and c the number of columns. One rejects the null hypothesis if the $\chi^2 >$ critical value from the chi-square distribution with $df = (r-1)(c-1)$ and desired significance level (usually, $\alpha = 0.05$).

The chi-square test should not be used if any cell has an expected value less than 1. Also, when one has expected values less than 5, the results may be incorrect. Then the use of a nonparametric test such as Fisher's exact test should be considered.

Example 1

If one's situation was the test for independence, then one would have a random sample of 330 items

Table 1 Example 1: Table of Observed Values

	Column 1	*Column 2*	*Column 3*	*Column Total*
Row 1	28	45	27	100
Row 2	25	32	33	90
Row 3	17	33	30	80
Row 4	30	10	20	60
Row total	100	120	110	330

Table 2 Example 1: Table With Observed Values, Expected Values, and Contribution to the Test Statistic

Observed Count Expected Value Contribution	*Column 1*	*Column 2*	*Column 3*	*Column Total*
Row 1	28	45	27	100
	30.30	36.36	33.33	
	0.175	2.051	1.203	
Row 2	25	32	33	90
	27.27	32.73	30.00	
	0.189	0.16	0.300	
Row 3	17	33	30	80
	24.24	29.09	26.67	
	2.164	0.525	0.417	
Row 4	30	10	20	60
	18.18	21.82	20.00	
	7.682	6.402	0.000	
Row total	100	120	110	330

that are cross-classified by two variables, Variable 1 (row) and Variable 2 (column). The null hypothesis is that Variable 1 and Variable 2 are independent of each other.

If one's situation was the test for equality of proportions, then one would have three independent samples (columns) that are classified by a variable (row) that has four categories. The null hypothesis is that the probability of Row 1 category is the same for all three populations (represented by the columns), the probability of Row 2 category is the same for all three populations, the probability of Row 3 category is the same for all three populations, and the probability of Row 4 category is the same for all three populations.

The expected cell count for the first row and the first column is

$$E_{11} = \frac{100 \cdot 100}{330} = 30.30.$$

and its part of the test statistic is

$$\chi^2 = \frac{(O_{11} - E_{11})^2}{E_{11}} = \frac{(28 - 30.30)^2}{30.30} = 0.175.$$

The following table gives the expected cell counts and the cells contribution to the test statistic.

The test statistic $\chi^2 = 0.175 + 2.051 + \ldots + 6.402 + 0.000 = 21.124$.

For this example, $df = (4-1)(3-1) = (3)(2) = 6$. If $\alpha = 0.05$, the critical value from a chi-square distribution is 12.59. Therefore, one would reject the null hypothesis.

After one has rejected the null hypothesis, one should investigate which cells contributed the most to the test statistic. In the above table, cells (4, 1) and (4, 2) contributed 7.682 and 6.402 to the test statistic, that is, the observed count for (4, 1) was much greater than the expected count, and for (4, 2) the observed count was much smaller than what would have been expected. One may also wish to observe the cells that had little contribution to the test statistic.

Goodness of Fit

Situation. One has *A* single random sample and *A* single variable of interest with 2 or more categories. The null hypothesis is that the population proportions follow a specific distribution, and the alternative hypothesis is that the population proportions do not follow the specific distribution.

Table 3 Example 2: Table of Observed Values

Category 1	*Category 2*	*Category 3*	*Category 4*	*Total*
12	25	50	13	100

Table 4 Example 2: Table With Observed Values, Expected Values, and Contribution to the Test Statistic

Category 1	*Category 2*	*Category 3*	*Category 4*	*Total*
12	25	50	13	100
10	30	40	20	
0.4	0.83	2.50	2.45	

Table 5 Chi-Square: Summary

	Test for Independence	*Test for Equality of Proportions*	*Goodness of Fit*
Sample	A single sample	Multiple samples	A single sample
Variable(s)	Two variables of interest	A single variable of interest	A single variable of interest
Null hypothesis	Variable 1 and Variable 2 are independent of each other.	For each category, the proportion is equal across the different populations.	The proportion of each category is given.
Expected value	$\frac{i\text{th row total} \times j\text{th column total}}{\text{grand total}}$	$\frac{i\text{th row total} \times j\text{th column total}}{\text{grand total}}$	Proportion given in the null sample size
Test statistic	$\sum_i \sum_j \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$	$\sum_i \sum_j \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$	$\sum_i \frac{(O_i-E_i)^2}{E_i}$
df	$(r-1)(c-1)$	$(r-1)(c-1)$	No. of categories -1

Note: In all the three cases, the chi-square test is suspect if expected values are below 5 and should not be performed if you have expected values below 1.

For example, one believes that the distribution of certain traits in offsprings will follow a specific pattern. One has a random sample of offsprings and classifies the offsprings according to the traits. The null hypothesis is that the traits follow a specific distributional pattern, and the alternative is that the traits do not follow the pattern.

The expected value for a category is equal to the proportion given in the null hypothesis for the category multiplied by the sample size. The test statistic is $\chi^2 = \sum_i (O_i - E_i)^2/E_i$ and $df =$ No. of categories -1. One rejects the null hypothesis if the $\chi^2 >$ critical value from the chi-square distribution with df and desired significance level (usually, $\alpha = 0.05$).

One should not use the goodness-of-fit test if one does not have a large enough sample size to ensure that the expected values are 5 or above.

Example 2

The null hypothesis is $p_1 = 0.10$, $p_2 = 0.30$, $p_3 = 0.40$, and $p_4 = 0.20$ versus the alternative that at least one probability is not equal.

The expected value for category 1 is $0.10 \times 100 = 10$, and the contribution to the test statistic is $(12-10)^2/10 = 0.40$. The following table contains the observed count, expected values, and contribution to the test statistic.

The test statistic is $\chi^2 = 0.4 + 0.83 + 2.50 + 2.45 = 6.18$. The critical value from the chi-square distribution with $df = 4 - 1 = 3$ and $\alpha = 0.05$ is 7.81, so one would not reject the null hypothesis and conclude that the data do not give enough evidence to say that the distribution is different from $p_1 = 0.10$, $p_2 = 0.30$, $p_3 = 0.40$, and $p_4 = 0.20$.

Table 5 summarizes the three cases of the chi-square test.

—Marjorie E. Bond

See also Fisher's Exact Test; Nonparametric Statistics; Study Design

Further Readings

Moore, D. S. (2007). *The basic practice of statistics* (4th ed.). New York: W. H. Freeman.

Samuels, M., & Witmer, J. (2003). *Statistics for the life sciences.* Upper Saddle River, NJ: Pearson Prentice Hall.

Weiers, R. (2002). *Introduction to business statistics* (4th ed.). Belmont, CA: Duxbury.

Cholesterol

Cholesterol was first described near the end of the 18th century by a French chemist, Antoine Francois de Fourcroy, and then named "cholesteroline" by Michel Eugene Chevreul in 1815. Cholesterol is an insoluble constituent of animal fats found among the lipids in the bloodstream and in all cells of the human body. As an essential component of cell membranes and serum lipoproteins, cholesterol enables transmembrane transport and the transport of triglycerides. As a precursor of bile acids, cholesterol aids in the absorption of fat in the intestine. And as a precursor to adrenal steroids and sex hormones, cholesterol aids in endocrine regulation. When regulated properly by the body, cholesterol ensures survival. When regulated improperly, cholesterol threatens good health. Elevated serum cholesterol levels often lead to the buildup of arterial plaques, heart attack, and even death.

Cholesterol originates in one of two ways: either in the bile or through the diet. The intake of biliary cholesterol is typically 600 to 1,000 mg per day, while the intake of dietary cholesterol is only 250 to 500 mg per day. Biliary cholesterol is primarily synthesized from acetyl CoA through the HMG-CoA reductase pathway in many cells and tissues.

Since it is insoluble, cholesterol cannot travel freely through the blood. Cholesterol transport in the body is achieved through the use of lipoproteins. Lipoproteins consist of a core of neutral lipids surrounded by a polar surface coat that allows for the transport of cholesterol and other insoluble triglycerides. There are five major classes of lipoproteins: chylomicrons, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). LDL carry triglycerides and cholesterol on to other body cells, while HDL transport cholesterol back to the liver for excretion.

Of the lipoprotein fractions, LDL, IDL, and VLDL are considered atherogenic. Conversely, increased concentrations of HDL correlate with lower rates of atherosclerosis. LDL pass through the arterial walls and become modified to form fatty streaks that in turn become fibrous plaques and finally lesions. These lesions often bring about calcification, hemorrhage, and ulceration. Atherosclerosis usually remains asymptomatic until an atheroma obstructs the bloodstream in the artery, and angina or myocardial infarction may subsequently develop.

Physiological and behavioral risk factors for atherosclerosis and coronary heart disease include age, heredity, diabetes, obesity, high blood pressure, and smoking, among others. In 1986, the statins, a class of pharmaceuticals that disrupt cholesterol biosynthesis by inhibiting HMG-CoA reductase, first became commercially available. Since then, the statins have emerged as the most effective therapeutic regimen for controlling a patient's blood cholesterol level, and clinical trials have repeatedly confirmed that altering lipoprotein transport significantly lowers a patient's risk for and incidence of cardiovascular disease.

—Todd M. Olszewski

See also Chronic Disease Epidemiology; Framingham Heart Study; Keys, Ancel

Further Readings

Grundy, S. M., Cleeman, J. I., Bairey Merz, C. N., Brewer, H. B., Jr., Clark, L. T., Hunninghake, D. B., et al. (2004). Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. *Circulation*, *110*, 227–239.

Rothstein, W. G. (2003). *Public health and the risk factor: A history of an uneven medical revolution.* Rochester, NY: University of Rochester Press.

Chromosome

A chromosome is a large macromolecule that functions as the structural unit of the genetic material. In eukaryotes, chromosomes are molecules consisting of linear, double-stranded deoxyribonucleic acid (DNA) and associated proteins. In prokaryotes, chromosomes are typically single-stranded, circular molecules. In epidemiology, knowledge of chromosomes is essential because chromosomal abnormalities are a leading cause of human genetic diseases. These abnormalities can include deletions (where part or all of a chromosome is missing), duplications (part or all of a chromosome is duplicated, resulting in excess genetic material), translocations (part of a chromosome is transferred to another chromosome), or inversions (part of the chromosome has detached, ''flipped over,'' and reattached, resulting in the genetic material being in the wrong order).

The totality of all the chromosomes in an individual is referred to as its genome. Each chromosome consists of genes (functional regions of the DNA that encode proteins), noncoding DNA, and associated structural proteins and ribonucleic acid (RNA). The sum of the material that makes up the chromosome is called chromatin. The number of chromosomes present varies greatly between species, ranging from a single chromosome (in the case of many bacteria) to more than 50 chromosomes in many animals. Humans have 46 paired chromosomes, receiving 23 from each parent. These include 22 pairs of autosomal chromosomes and one pair of sex chromosomes, which determine the gender of the individual: Females receive an X chromosome from each parent (giving them an XX genotype), while males receive an X chromosome from their mother and a Y chromosome from their father (resulting in an XY genotype).

One of the most important genetic diseases is Down's syndrome, which is the result of a trisomy (a duplication, leading to three copies) of chromosome 21. Individuals with Down's syndrome typically have mild to moderate mental retardation, decreased muscle tone, and shortened limbs. The incidence of Down's syndrome is approximately 1 in 800 births in the United States, and is most common in mothers who are above the age of 40 at birth. However, genetic testing can be carried out during pregnancy to inform parents if their fetus is positive for this chromosomal abnormality.

While Down's syndrome results from an extra autosomal chromosome, other genetic conditions can result from duplication or deletion of all or portions of the sex chromosomes (X and Y). Men with Klinefelter's syndrome possess an extra X chromosome, leading to an XXY genotype. Physically, they tend to be sterile, and tall with long arms and legs. Females with Turner syndrome, on the other hand, lack a second sex chromosome, and genetically are X0. Female sex characteristics may be present but are underdeveloped.

Chromosomal abnormalities also play a role in some cancers. Tumor cells frequently are aneuploid, meaning they have an abnormal number of chromosomes. They may also contain translocations or portions of chromosomes that have been copied not just once but dozens or hundreds of times.

—Tara C. Smith

See also Genetic Disorders; Genetic Epidemiology; Genomics

Further Readings

Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C., & Gelbart, W. M. (2004). *Introduction to genetic analysis* (8th ed.). New York: W. H. Freeman.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Molecular cell biology* (4th ed.). New York: W. H. Freeman.

University of Utah Genetic Learning Center. (2007). *Understanding genetic disorders*. Salt Lake City, UT: Author. Retrieved June 28, 2007, from http://learn.genetics.utah.edu/units/disorders/index.cfm.

Chronic Disease Epidemiology

With important advances in medicine and public health and overall increases in average life expectancy in the developed and developing world in the 20th century, chronic diseases have now reached epidemic proportions globally. This entry provides an overview of factors that have contributed to the worldwide emergence of chronic diseases; describes the epidemiology of the major chronic diseases of

cardiovascular disease, cancer, and diabetes; and highlights some of the core epidemiologic and statistical methods for studying chronic diseases.

Global Emergence of Chronic Diseases

In industrialized countries, the latter half of the 20th century bore witness to the emergence of chronic diseases as major contributors to morbidity and mortality. In recent years, similar patterns have been taking place in the developing world. These trends are largely due to significant advancements in medicine and public health—including improvements in sanitation, nutrition, and the discovery of antibiotics such as penicillin—that led to overall reductions in perinatal and childhood mortality, declines in the incidence of infectious diseases, and rises in life expectancy. Together, these changes caused a shift in the incidence of infectious diseases to chronic, noncommunicable diseases, particularly because many chronic diseases are highly age dependent. This shift has been labeled the "epidemiologic transition." Simultaneous increases in economic development and urbanization have also spurred changes in lifestyle factors, including diet, physical activity, and stress, and compounded the burden of morbidity and mortality from chronic diseases. Globally, approximately 58 million deaths are estimated to have taken place in 2005, with 60% (35 million deaths) attributed to chronic diseases. The leading causes of death from chronic diseases are cardiovascular diseases (primarily heart disease and stroke), cancer, chronic respiratory diseases, and diabetes. With the large populations and the epidemiologic transition in the developing world, chronic diseases have become a global epidemic. In 2005, 80% of all deaths from chronic diseases had occurred in low- and middle-income countries. Some of these countries, such as India, face "double burdens" of acute infectious diseases (e.g., malaria) and chronic diseases.

Major Chronic Diseases

Cardiovascular Disease

Cardiovascular disease (comprised mainly of heart disease and stroke) accounted for 30% (17 million) of all deaths, and 10% of disability-adjusted life-years worldwide in 2005, and thus represents the leading cause of mortality and morbidity globally. It is anticipated that the developing world will experience the greatest increases in the burden of cardiovascular disease in the decades to come. Between 1990 and 2020, the developing world is projected to have 137% and 120% increases in the number of deaths from coronary heart disease among men and women, respectively. In contrast, these increases are estimated to be 29% and 48% in developed countries. Likewise, 124% and 107% increases of stroke deaths among men and women in developing countries are expected over the same time period, with lesser respective increases of 78% and 56% in developed countries.

In Western developed nations, including the United States, Canada, and Australia, mortality rates from coronary heart disease rose until the late 1960s, after which there was a secular decline in mortality rates. A similar pattern was observed for stroke mortality rates, although the decline began as much as two decades earlier and was more pronounced than for heart disease. Presently in these countries, coronary heart disease mortality rates exceed stroke mortality rates. Despite the declines in annual rates of death for both coronary heart disease and stroke, the absolute number of deaths from each outcome has substantially increased over the past decade. This is mainly due to population aging, with mortality rates increasing successively with age. Notably, disparities in these outcomes, particularly along racial/ethnic lines, are well established. For example, in the United States, compared with whites, in both men and women, heart disease and stroke death rates are higher among blacks, yet lower among Asians/Pacific Islanders and Hispanics.

Established risk factors for coronary heart disease and stroke include age, male gender, smoking, elevated low-density lipoproteins, low high-density lipoproteins, hypertension, diabetes, physical inactivity, obesity, and low socioeconomic status. Risk factors for which consensus is less established include novel inflammatory markers, elevated homocysteine and elevated lipoprotein(a) levels, psychological factors such as depression and hostility, and physical and social environmental factors such as residence in a low-income neighborhood. Strokes are classified as ischemic (occurring mainly as a result of atherosclerosis) or hemorrhagic (for which the most important risk factor is hypertension), with approximately 80% of strokes in Western developed countries being ischemic. In countries in sub-Saharan Africa and Asia (such as China and Japan), hemorrhagic strokes are relatively more common than ischemic strokes. Unlike Western developed nations, these regions are

also characterized by higher mortality rates from strokes than from coronary heart disease.

Cancer

Cancer accounted for 13% (7.6 million) of all deaths, and 5% of disability-adjusted life-years worldwide in 2005, and is the second leading cause of mortality and morbidity internationally. In 2000, among men, the leading causes of cancer deaths were lung, stomach, liver, and colon and rectal cancer. Among women, the main contributors were breast cancer, followed by lung, stomach, and colon and rectal cancer. For both sexes combined, lung cancer accounted for the greatest proportion (17%) of cancer deaths. The number of new cases of cancer is estimated to escalate by 50% to 50 million in 2020. This dramatic increase is largely driven by population aging globally and the epidemiologic transition in countries in the developing world.

In the United States, among men, lung, prostate, and colon and rectal cancer were the top three contributors to cancer death in 2001, with liver and stomach cancer ranking fifth and sixth, respectively. Among women, lung, breast, and colon and rectal cancer were the leading causes of cancer death, with stomach cancer ranking fifth. Death rates for the top three contributors to cancer death for men and women have been on the decline (for lung and prostate cancer, since the early 1990s, and for colon and rectal cancer, since around 1980). These trends for lung cancer and colon and rectal cancer in part reflect similar falls in incidence rates. In contrast, breast cancer and prostate cancer incidence rates have continued to rise steadily; declines in mortality rates from these cancers may be the result of earlier screening detection and improved medical treatments. Like cardiovascular disease, racial/ethnic disparities also exist for cancer outcomes. Mortality rates are higher among blacks than whites for cancers of the colon and rectum, lung (men only), breast, and prostate, and incidence rates are higher among blacks for all these cancers except breast cancer. Some of these disparities may be due to lower socioeconomic status, with 24% of blacks living below the poverty line (vs. 8% in whites). Socioeconomic status likely influences the adoption of risk factors for cancer, as well as access to cancer screening and high-quality medical treatment.

Approximately 35% of deaths from cancer globally have been attributed to nine potentially modifiable risk factors. Both worldwide and in low- and middle-income countries, the leading risk factors for death from cancer in 2001 were smoking, alcohol consumption, and low fruit and vegetable intakes. In high-income countries, smoking, alcohol consumption, and overweight and obesity were the main contributors to cancer mortality. These modifiable risk factors offer a vast opportunity for preventing substantial morbidity and mortality globally.

Diabetes

Diabetes comprised 2% (1 million) of all deaths, and 1% of disability-adjusted life-years worldwide in the year 2005. It is projected that the total number of adults with diabetes will more than double from 171 million in the year 2000 to 366 million in 2030. This rapid increase can be attributed to aging of the population globally, the epidemiologic transition in the developing world, and the increasing prevalence of poor diets, sedentary behaviors, and obesity in developed nations. India, China, and then the United States are estimated to have the highest numbers of diabetes cases in both 2000 and 2030.

In the United States, nearly 1 in 10 Americans above the age of 20 had diabetes in 2005, with a slightly lower prevalence among women (8.8%) than men (10.5%). Based on death certificate records, diabetes ranked as the sixth leading cause of death in 2002. However, diabetes is most likely underreported as an underlying cause of death on death certificates.

Diabetes is associated with both macrovascular complications (heart disease and stroke) and microvascular complications (including blindness and kidney disease). Established risk factors for diabetes include age, obesity, physical inactivity, and a family history of diabetes. Furthermore, in the United States, blacks and Hispanics are 1.5 to 2 times more likely to develop diabetes than whites.

Epidemiologic and Statistical Methods

The investigation of potential risk factors for chronic diseases, including estimates of the strength of the associations of these factors with diseases, is based on epidemiologic studies that range in design from ecologic studies (in which only data at a group level and not individual level are compared), to case-control studies (in which cases of disease are compared with respect to the potential risk factor to individuals from

the same source population), prospective cohort studies (in which data on the risk factor are measured at one point in time and the individuals are then followed over time through regular examinations, tests, or surveys), and randomized clinical trials (in which individuals are randomly assigned to a treatment or placebo group and followed over time). The occurrence of chronic disease can be identified through surveys (i.e., with self-report) or through medical records or disease registries (for cancer diagnoses), while reports of deaths from chronic diseases are often requested from the next of kin and then confirmed through medical records or national death registries.

In these studies, statistical methods are applied to estimate the associations between the risk factor under study and the chronic disease outcome, controlling for age and other factors that predict the outcome and that may also be associated with the risk factor under study (i.e., potential confounders). The analyses hence typically use multivariable regression models to derive estimates. These models often take the specific form of logistic regression models, or Poisson or Cox regression models, which differ according to how the model structure is specified. In instances where a physical or social environmental characteristic is being examined as a main predictor or is included as a confounder, applying multilevel models (such as two-level models in which similarities on the outcome in individuals within the same spatial area are taken into account) is appropriate to obtain more valid estimates of the statistical significance of associations.

Moreover, when different populations are being compared (such as across countries) for a chronic disease, it is important to calculate and compare age-standardized disease rates, so that different age structures between populations do not bias the comparisons. This is achieved through a standardization procedure, whereby age-specific disease rates are weighted according to a standard population, such as the World Health Organization World Standard population, to produce a summary of age-standardized incidence or mortality rate from the disease.

Apart from mortality figures, disability-adjusted life-years (DALYs) provide summary measures of disease burden in terms of both mortality and morbidity to allow for comparisons across countries or regions. DALYs combine years of life lost due to premature mortality (YLL) with years of life lived with disability (YLD) for a given disease or group of diseases. Like mortality, to permit more valid comparisons across countries or regions, it is appropriate to age-standardize DALYs using a standard population to produce summary measures.

Finally, epidemiologic studies that examine the risk factors for chronic diseases, or the determinants of trends in chronic disease incidence or mortality rates, are more likely to be valid when they consider that a true risk factor typically does not immediately cause the outcome, but rather precedes the outcome by a certain number of years. For example, based on biological mechanisms, a prospective cohort study on smoking and heart disease incidence with 10 to 15 years of follow-up is more likely to be valid and to find a stronger association than a study that measures both smoking exposure and heart disease status at the same point in time. Similarly, studies that explore the possible effects of economic growth on trends in stroke mortality rates years later are more likely to be valid and sensitive than those that model economic growth and stroke mortality for the same time period.

Future Patterns of Chronic Disease

Chronic diseases have become the major contributor to mortality and morbidity globally. Over the next few decades, the burden of chronic diseases will grow fastest in countries in the developing world, largely due to the epidemiologic transition, economic development, and urbanization. Because the major chronic diseases worldwide—cardiovascular disease, cancer, and diabetes—share several modifiable risk factors, there exist tremendous opportunities to prevent these diseases. Epidemiologic studies have played an important role in identifying such risk factors, and in investigating other factors that may determine patterns of chronic disease within countries and around the world.

—Daniel Kim

See also Cancer; Cardiovascular Disease; Diabetes; Regression; Study Design

Further Readings

American Cancer Society. (2006). *Cancer facts and figures 2006.* Atlanta, GA: Author.

Danaei, G., Vander Hoorn, S., Lopez, A. D., Murray, C. J. L., & Ezzati, M. (2005). Causes of cancer in the world:

Comparative risk assessment of nine behavioral and environmental risk factors. *Lancet*, *366*, 1784–1793.
Shibuya, K., Mathers, C. D., Boschi-Pinto, C., Lopez, A. D., & Murray, C. J. L. (2002). Global and regional estimates of cancer mortality and incidence by site: II. Results for the global burden of disease 2000. *BMC Cancer*, *2*, 37.
Wild, S., Roglic, G., Green, A., Sicree, R., & King, H. (2004). Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. *Diabetes Care*, *27*, 1047–1053.
World Health Organization. (2005). *Preventing chronic diseases: A vital investment*. Geneva, Switzerland: Author.
Yach, D., Hawkes, C., Gould, C. L., & Hofman, K. J. (2004). The global burden of chronic disease: Overcoming impediments to prevention and control. *Journal of the American Medical Association*, *291*, 2616–2622.
Yusuf, S., Reddy, S., Ôunpuu, S., & Anand, S. (2001). Global burden of cardiovascular diseases: Part I: General considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation*, *104*, 2746–2753.

Circumcision, Male

Male circumcision is a practice that has been carried out across societies of the world for thousands of years. Recent scientific and epidemiological research provides evidence of protective effects against penile foreskin-related disorders, urinary tract and sexually transmitted infections, penile and cervical cancer, and HIV. This entry provides a comprehensive overview of male circumcision's prophylactic role as a low-risk, low-cost procedure with significant potentially long-term benefits to the individual and society.

Male circumcision is the removal of the foreskin (prepuce) from around the head (glans) of the penis. The amount of skin removed varies drastically by individual due to different foreskin sizes. A short prepuce does not completely cover the glans and exposes the tip of the head even when flaccid, whereas a long prepuce is loose and droops down from the end of an unerect penis. Many of these differences are genetic, so there are generalized trends among like populations. When erect, the glans emerges from the foreskin sleeve. Circumcision removes the foreskin, thus always exposing the head of the penis.

Who Gets Circumcised and Why?

Male circumcision dates back several thousand years. The earliest documentation comes from Egyptian tomb artwork dated to the Sixth Dynasty (2345–2181 BCE). The book of Genesis (17:11) speaks of circumcision as a rite of passage for Jews during the age of Abraham, who lived around 2000 BCE. Although its origin is unknown, male circumcision practice is widespread today, extending from Africa to the Middle East, the islands of the Pacific to the West.

Male circumcision practices differ greatly by culture. Its practice is often associated with rites of passage into adulthood, religious sacrifice, and hygiene promotion. Approximately 25% of males are circumcised globally. The age at time of circumcision also varies with culture, extending from infancy through puberty and into adulthood. Jews and Muslims mandate circumcision as a part of their religious practice and account for 100,000 and 10 million annual circumcisions, respectively. Pacific Islanders are traditionally circumcised as a rite of initiation, as are Australian aboriginals. In Africa, circumcision practice is disparate, dictated largely by influences ranging from colonization to tribal rituals.

In the United States, 65% to 90% of males are circumcised. This wide range is attributed to differences between the statistical reporting of birthing centers and the observed rate of practice that includes adult circumcision. From 1988 to 2000, the U.S. newborn circumcision rate increased by 12.8%. This increase is most prominent in states where immigration is low, because most immigrants to the United States, particularly Hispanics, are traditionally not circumcised.

Outside religious or medical influences, circumcision decisions are based most heavily on parental preference. For mothers, there is a strong correlation between their son's circumcision status and the woman's ideal male partner's circumcision status. Likewise, fathers make the decision based on personal experience.

Histological and Biological Effects of Male Circumcision

To fully comprehend the protective effects of male circumcision, one needs to have a general understanding of the histological and biological differences between a circumcised and uncircumcised penis.

Keratin is a protein found in skin cells that acts as a primary line of immunological defense against infection. According to a histological study conducted

in 2006 by McCoombe and Short, different regions of the penis are keratinized to varying degrees. In an uncircumcised male, the outer surface of the foreskin is heavily keratinized while the inner surface is not and closely resembles the mucosal epithelia of the cervix and nasal passageways. When the penis is erect, this weak inner prepuce is exposed and stretched, thus putting it in direct contact with potentially infectious agents during unprotected sex.

Furthermore, the warm mucosal environment under the foreskin favors growth of microorganisms. The preputial sac must be properly and frequently cleaned to prevent infection. Multiple studies show lower levels of penile hygiene in uncircumcised men, as compared with those who are circumcised. This poor genital cleanliness contributes to higher incidence of penile discomfort and infection.

Circumcision removes the weakly keratinized inner prepuce and cesspool-like preputial sac. The basis for male circumcision's prophylaxis is believed to be in the keratinization of the penis and the reduction of bacterial growth.

Health Benefits of Male Circumcision

Substantial observational, epidemiological, and biological research points to the prophylactic role of male circumcision. These benefits include a reduced incidence of

1. Foreskin-related disorders
2. Urinary tract infections
3. Sexually transmitted infections
4. Human papillomavirus
5. Penile cancer
6. Cervical cancer
7. HIV

Foreskin-Related Disorders

The following three foreskin-related disorders are common among uncircumcised men.

- *Phimosis* is a condition in which the foreskin of the penis becomes constricted and difficult to retract. Although congenital phimosis is common in young boys while the epithelial layers between the glans and foreskin keratinize and separate, 90% are no longer affected by age 3. After age 6, phimosis is considered problematic and is most commonly owed to poor hygiene. Repetitive forceful retraction of a congenital phimosis by parents in an attempt to clean under their child's foreskin may also contribute to the onset of acquired phimosis. At least 10% of uncircumcised adult males are afflicted by acquired phimosis.

- *Paraphimosis* is a condition where the retracted foreskin cannot be brought back again over the glans of the penis to its naturally occurring position. Paraphimosis is a true urologic emergency and is most often relieved by circumcision or foreskin slitting. Frequent bladder catheterizations in adults, without replacing the foreskin in its resting position, is a common cause of adult paraphimosis.

- *Balanoposthititis* is an inflammation of the foreskin and glans, most often caused by acute infection underneath the foreskin of an unhygienic penis. It is also common among uncircumcised, diabetic men owing to a weakened, shrunken penis.

Urinary Tract Infections

Infections of the urinary tract are regarded as common among the pediatric population with the highest prevalence and greatest severity in boys prior to 6 months of age. A study at Kaiser Permanente hospitals in northern California revealed that 86% of the urinary tract infections (UTIs) among boys occurred in those who were uncircumcised. A meta-analysis of nine studies found an average 12-fold increase of UTI incidence in uncircumcised boys. The pathophysiological basis for this high incidence is the ease in which uropathic bacteria cling to the mucosal lining underneath the foreskin and travel up the urethra, thereby infecting the urinary tract. With long-term morbidity and potential mortality associated with untreated UTIs, preventive measures are highly desirable.

Sexually Transmitted Infections

The association between circumcision status and sexually transmitted infection (STI) is not yet entirely conclusive. A range of studies conducted in recent years show differential patterns of association. In a 2006 meta-analysis, Weiss, Thomas, Munabi, and Hayes reviewed 26 studies and found a strong

association for syphilis and chancroid infection but no association with genital herpes (HSV-2) in men. Another study showed women with uncircumcised male partners to be 2.2 times more likely to contract HSV-2 and 5.6 times more likely for chlamydia infection.

The proposed biological rationale for increased STI incidence in uncircumcised men and their partners is multifold. The area underneath the prepuce is susceptible to microabrasions during intercourse and is a moist breeding ground for pathogens. Incidence appears to be influenced by genital hygiene, availability of running water, sexual practices, and socioeconomic status. As a result, developed countries are less likely to see male circumcision's impact on STI incidence. In poorer settings, where penile hygiene suffers, incidence is much higher for uncircumcised males and their partners.

Human Papillomavirus

Incidence and Prevalence

Human papillomavirus (HPV) is a sexually transmitted virus that affects both men and women. It spreads primarily through genital contact and can remain undetected for years. Approximately 20 million people are currently infected with HPV. More than 50% of sexually active men and women acquire genital HPV infection at some point in their lives, and, by age 50, at least 80% of women will have had HPV. Nearly 6.2 million Americans get a new HPV infection each year.

Link to Penile and Cervical Cancers

High-risk types of HPV (namely, Types 16 and 18) are associated with the incidence of penile cancer in men and cervical cancer in women. Nearly 50% of penile cancer cases are coupled with HPV infection, and the relationship between HPV-positivity and cervical cancer is almost 100% in women.

Prophylactic Benefits of Male Circumcision

Recent studies point to a link between circumcision status and HPV infection, the most convincing of which was led by Castellsague in 2002. It shows HPV detection in 19.6% of uncircumcised men but in only 5.5% of circumcised men. After adjusting for potential confounding factors, such as sexual behavior, circumcised men are only one third as likely to have HPV infection.

Penile Cancer

Incidence and Prevalence

In the United States, the total annual incidence of penile cancer is approximately 1 per 100,000. Incidence is even lower in Israel—0.1 per 100,000—where almost all men are circumcised. The prevalence is much higher in poor countries, especially African countries where male circumcision is not routinely practiced. Here, penile cancer rates can be 10 times more common than in developed countries, representing 10% to 22% of all male cancers.

Etiology

The majority of penile malignancies are cancers of the skin and can be found anywhere along the shaft of the penis although they are most often found on the foreskin and glans. It is treatable if detected early. The relationship between male circumcision and penile cancer is believed to be attributed to poor hygiene. The decaying cells on the undersurface of the foreskin (smegma) produce irritation and lead to cancer development. The presence of venereal warts and HPV has also been implicated in the etiology of penile cancers.

Prophylactic Benefits of Male Circumcision

The evidence pointing to male circumcision's benefit in penile cancer prevention is astounding. Males who are circumcised at birth almost never get penile cancer. In fact, of the approximately 50,000 reported cases of penile cancer in the United States since the 1930s (10,000 of which resulted in death), only 10 cases were reported in circumcised men. Those who are circumcised within the first few years of life have a decreased incidence, and adult circumcision confers almost no protective effect.

Cervical Cancer

Incidence and Prevalence

Cancer of the cervix is the second most prevalent cancer in women worldwide with 450,000 new diagnoses per year. With more than 80% occurring in less

developed countries, it is the leading cause of cancer-related death for these women. In the United States, an estimated 10,000 women will be diagnosed with cervical cancer in 2006 and nearly 4,000 will die from it.

Prophylactic Role of Male Circumcision

A number of studies have documented higher rates of cervical cancer in women who have had one or more uncircumcised male sexual partners. Castellsague et al.'s multinational study revealed that women are 5.6 times more likely to have cervical cancer if their partner is uncircumcised and at high risk for HPV exposure. Penile HPV infection is associated with a fourfold increase in the risk of cervical HPV infection in the female partner, which then has a 77-fold increase in the risk of cervical cancer. Thus, the cervical cancer epidemic worldwide appears to be contributed, at least in part, to the uncircumcised male.

Human Immunodeficiency Virus

Incidence and Prevalence

More than 25 million people have died from AIDS. To date, 60 million have been infected and 40 million are currently living with HIV. The overwhelming majority of these individuals—98% of women and 94% of men—live in developing countries. More than 80% of these infections arise from vaginal intercourse. An extrapolation model estimates 1 billion to be HIV-positive by 2050 with an exponential increase in HIV-orphaned children.

Prophylactic Role of Male Circumcision

As previously discussed, histological analysis of an uncircumcised penis shows the inner prepuce to be weakly keratinized and more susceptible to minor trauma during intercourse. The distribution of HIV-susceptible Langerhans's cells is greatest on the outer surface of the foreskin, with high levels also at the inner foreskin layer. The weakly keratinized inner prepuce with frequent microabrasions, coupled with high-density Langerhans's cells and their HIV-1 receptors, substantially increases susceptibility to infection. Moreover, selective entry of HIV at the inner prepuce has been shown by direct experimentation using punch biopsies. Live uptake was observed at the inner preputial Langerhans's cells and none on the keratinized tissue of the outer foreskin, histologically similar to the shaft of the circumcised penis.

To further compound the risk, the preputial sac underneath the foreskin is speculated to act as a storage site for HIV after intercourse, therefore prolonging time of exposure and allowing for greater infection. Ultimately, the weakest point (inner prepuce) of an uncircumcised penis is exposed to potentially infected vaginal secretions during and after vaginal intercourse.

Epidemiological Studies

Since the 1980s, more than 40 studies have been conducted to understand the role of male circumcision in HIV incidence. Observational studies have consistently shown a considerably lower incidence of HIV in localities where male circumcision is practiced. Weiss's large systematic meta-analysis of 27 of these studies showed the HIV incidence to be 2.4-fold higher in uncircumcised men when adjusted for confounding factors. A study in Dar es Salaam, Tanzania, analyzed the differential risk to women due to their partner's circumcision status. It was found that married women with one sex partner were at a four times higher risk of HIV infection if their husband was uncircumcised.

To study the absolute protective effect of male circumcision, a prospective study followed heterosexual couples in which one partner was HIV-positive. The incidence of seroconversion was 17 per 100 person-years among the 137 uncircumcised males, whereas none of the 50 circumcised men seroconverted. The effect was equally apparent in Muslim (who wash after intercourse) and non-Muslim men, suggesting religious and cultural behaviors were not involved in transmission rates.

In 2005, the results of the first randomized controlled trial (RCT) in Orange Farm, South Africa, were published. The protective benefit of circumcision was so great that the Data and Safety Monitoring Board halted the trial 6 months early so that the control group could immediately be offered circumcision. Male circumcision prevented more than 60% of HIV infections, an efficacy rate similar to that of a vaccine. Moreover, 99% of men were "very satisfied" with their new circumcision status. The two other RCTs—Kenya and Uganda—are still in progress with projected completion dates for 2007 and 2008, respectively.

In response to Auvert et al.'s (2005) Orange Farm results, a study estimated the potential impact of male circumcision in sub-Saharan Africa. Assuming

full coverage of male circumcision and a 60% protective effect, Williams et al. (2006) predicts male circumcision could avert 2 million new HIV infections and 0.3 million deaths over the next 10 years, with even greater effects in the 10 years to follow.

Implementation of Male Circumcision as a Public Health Measure

In response to these studies, some countries are promoting male circumcision as a means of protection. In these locations, circumcision clinics are overwhelmed with huge demand from young men. In Swaziland, for example, there is an 8-month waiting list for adult male circumcision. Not only does the implementation of male circumcision yield protection from HIV, it also provides a means for accessing males in reproductive health issues.

Many public health officials share a valid concern that male circumcision interventions may be misinterpreted as a guarantee against HIV infection, thus encouraging risky behavior in newly circumcised men. This circumstance can be avoided if male circumcision implementation and integration is carried out carefully. It must be promoted as part of an educational package reiterating that it reduces, but does not eliminate, the risk of HIV infection.

Sociosexual Perspectives

Socially, preference for male circumcision differs by culture. A survey of women in the United States indicates preference for male circumcision because of cleanliness, visual appearance, and benefits to intercourse. In Korea, male circumcision is viewed as necessary for penile hygiene measures. African women often prefer male circumcision because of its known reduced STI risk.

Functional differences due to circumcision status are difficult to quantify. Somatosensory tests yield no differences in penile sensitivity, and histological research reveals similar keratin levels for uncircumcised and neonatally circumcised penises. A national study among 1,400 men in the United States found that uncircumcised men are more likely to experience sexual dysfunctions later in life, including a two times greater difficultly in maintaining an erection.

Whether or not sexual pleasure is enhanced or diminished via circumcision is subjective, differing by individual, partner, and sexual behaviors. However, an English study of 150 adult-circumcised men revealed that 38% claimed an enhanced sensation in vaginal intercourse, 18% reported a decreased sensation, and 44% said it was unchanged. Of these men, 61% were pleased with their new circumcision status and 17% were not.

A Note on Female Genital Mutilation

There is an important distinction to be made between male circumcision and female genital mutilation (FGM), sometimes under the misnomer "female circumcision." The most widely practiced form of FGM entails removal of the clitoris and labia for traditional or religious practices. This genital scarring results in dyspareunia and life-threatening complications during labor and delivery. Male circumcision, in contrast, has major health benefits.

Risks of Male Circumcision

As with any surgical procedure, there are inherent risks with circumcision. Clinical studies performed over the past decades have provided overwhelming evidence that when performed by an experienced operator, circumcision is a very safe surgical procedure. Surgical complications range from 0.2% to 0.6% and include excessive bleeding (1 per 1,000 cases) and local infections (1 to 10 per 1,000). Both complications can be easily addressed and solved. Greater concerns exist in poor countries where access to clean surgical tools and running water is difficult. Overall, male circumcision is one of the top 40 most frequently performed surgical procedures with thousands of years of practice.

—Hilarie Martin and Julia Walsh

See also Cancer; Clinical Trials; Confounding; Etiology of Disease; HIV/AIDS; Meta-Analysis; Sexually Transmitted Diseases

Further Readings

Auvert, B., Taljaard, D., Lagarde, E., Sobngwi-Tambekou, J., Sitta, F., & Puren, A. (2005). Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: The ANRS 1265 trial. *PLoS Medicine*, *2*(11), 1112–1122. Retrieved June 28, 2007, from http://medicine.plosjournals.org/archive/1549-1676/2/11/pdf/10.1371_journal.pmed.0020298-S.pdf.

Castellsague, X., Bosch, F. X., Munoz, N., Meijer, C. J. L. M., Shah, K. V., de Sanjose, S., et al. (2002). Male

circumcision, penile human papillomavirus infection, and cervical cancer in female partners. *New England Journal of Medicine*, *346*, 1105–1112.
McCoombe, S. G., & Short, R. V. (2006). Potential HIV-1 target cells in the human penis. *AIDS*, *20*(11), 1491–1495.
Morris, B. (2006, September). *Benefits of circumcision: Medical, health and sexual*. Retrieved September 30, 2006, from http://circinfo.net.
Schoen, E. J. (2006). Ignoring evidence of circumcision benefits. *Pediatrics*, *118*, 385–387.
Weiss, H. A., Quigley, M. A., & Hayes, R. J. (2000). Male circumcision and risk of HIV infection in sub-Saharan Africa: A systematic review and meta-analysis. *AIDS*, *14*, 2361–2370.
Weiss, H. A., Thomas, S. L., Munabi, S. K., & Hayes, R. J. (2006). Male circumcision and risk of syphilis, chancroid and genital herpes: A systematic review and meta-analysis. *Sexually Transmitted Infections*, *82*, 101–110.
Williams, B. G., Lloyd-Smith, J. O., Gouwx, E., Hankins, C., Getz, W. M., Hargrove, J., et al. (2006). The potential impact of male circumcision on HIV in sub-Saharan Africa. *PLoS Medicine*, *3*(7), 1032–1040. Retrieved June 28, 2007, from http://medicine.plosjournals.org/archive/1549-1676/3/7/pdf/10.1371_journal.pmed.0030262-S.pdf.
Wiswell, T. E., & Hachey, W. E. (1993). Urinary tract infections and the circumcision state: An update. *Clinical Pediatrics*, *32*, 130–134.

Classification and Regression Tree Models

Classification and regression tree models, also known as recursive partitioning or CART™, are a class of nonparametric regression models becoming increasingly popular in epidemiology and biomedical data analysis, as well as in the computer science and data-mining fields. These models became popular when this methodology was formalized by Leo Breiman and colleagues in their book *Classification and Regression Trees*. Subsequent availability of commercial software (e.g., Salford Systems, Inc.) and academic freeware (R ''tree'' and ''rpart'' functions) for fitting these models helped make this approach practical to data analysis. One of the most common uses of classification and regression tree models in epidemiology is to develop predictive rules for diagnosis; other uses include developing screening guidelines and creating prognostic models.

The goal of regression and classification is to fit a mathematical model that takes categorical or continuous input (independent or predictor variables) and returns a predicted value for an output (dependent or outcome variable). To take a simple example, the analyst may want to predict a person's weight as a function of their height using a simple linear regression model where height is the input and weight is the output. In collecting data, the analyst will have measured height and weight on numerous people and are likely to have several individuals with nearly or exactly the same height. For this subset, the weights will follow a distribution with some people being heavier or lighter than others. A simple linear regression of weight on height gives a formula that for a particular height, the model returns the ''expected'' or ''mean'' weight for individuals at that height. The concept of the ''expected'' or ''mean'' weight for individuals at a particular height is essential for understanding regression. In statistics, a simple linear regression gives the conditional mean of Height for a given value of Weight $= w$, which we write as $E(\text{Weight}|\text{Height} = h) = \beta_0 + \beta_1 \times \text{Height}$, where $E(\text{Weight}|\text{Height} = h)$ is the expected value of weight for a person with height h, and β_0 and β_1 are the parameters of the equation used to make this prediction.

A similar concept applies to the classification problem where the input variables are used to predict which group an individual or some other object belongs to. The proper statistical approach for this type of analysis includes discriminant analysis, though often logistic regression is used by many applied data analysts. In classification, the idea of conditional expectation, that is, $E(\text{Weight}|\text{Height} = h)$, is replaced with a statement of probability of belonging to one of the classes. The simplest problem considers two classes, say, patients who either responded to treatment or did not. In a treatment study comparing response rates for patients receiving drugs versus patients receiving a placebo, the analyst would fit models to obtain estimates of conditional probabilities of response, for example, Prob(Patient responded|Patient received drug) versus the Prob(Patient responded|Patient received placebo).

Nonparametric recursive partitioning has the same goals as regression and classification, as above, but does not assume a particular parametric model. Using a nonparametric approach allows more flexibility in the model fitting adjusting for fluctuations in the data, however, at giving up computational simplicity and formal tests of hypothesis (e.g., testing for significance of a coefficient). In the case above, where weight was regressed on height, a linear relationship might be

reasonable for a homogeneous population, for instance, within a particular age range and gender group. It might become less reasonable if all ages from infants to the elderly were included, or if outliers such as weight lifters and marathon runners were included. (While a more complex parametric model with more terms and interactions might fit the data well, confirming that the parametric model is correct cannot always be done easily, especially as the number of input variables increase.)

Example: Kyphosis Following Spinal Surgery

To illustrate recursive partitioning we will consider a simple two-class classification problem and then the regression problem. These techniques will be illustrated using the kyphosis data set available through the R software package (www.r-project.org), which represents data on children who have had corrective spinal surgery. The outcome variable `Kyphosis` is a binary variable with levels `absent` or `present` indicating if a kyphosis (a type of deformation) was present after the operation. The input variables include `the age` in months of the child, the number of vertebrae involved, and the number of the first (topmost) vertebra operated on, that is, the starting vertebra. The first 10 rows of data are presented in Table 1.

Recursive partitioning methods attempt to find ranges over the input variables where the value of the outcome variable is more likely to be one class rather than the other. In the entire data set, 17 of the 81 rows had an outcome of Kyphosis "present," so the proportion of Kyphosis was 21% overall. By finding the appropriate ranges, or partitions of the covariate space, we can see if the proportion of present and absent outcomes increases or decreases over different values and combinations of values of these variables. Figure 1 presents a partition using the variables age and starting vertebrae.

In this figure, each case is plotted at their age and start value and labeled as kyphosis absent (A) or present (P). The covariate space is partitioned into five regions and the proportion of outcomes indicated. The top region consists of 27 cases all with kyphosis absent. This region is defined by the starting vertebrae above number 14, while the bottom region defined by the starting vertebrae less than 9 is more likely to have kyphosis present (11 of 19, or 58%).

In brief, recursive partitioning finds subregions such that the class membership within that region is more homogeneous than in the data set overall.

Calculating a recursive partitioning model is computationally intensive but conceptually easy. The algorithm takes each input variable and goes through each possible cut point that splits the data into two groups. The cut points are determined from the data and are the midpoints between the ordered values (e.g., for age these cut points might be 10.5 months,

Table 1 First 10 Rows of Kyphosis Data Set

Row	*Kyphosis*	*Age*	*Number*	*Start*
1	Absent	71	3	5
2	Absent	158	3	14
3	Present	128	4	5
4	Absent	2	5	1
5	Absent	1	4	15
6	Absent	1	2	16
7	Absent	61	2	17
8	Absent	37	3	16
9	Absent	113	2	16
10	Present	59	6	12

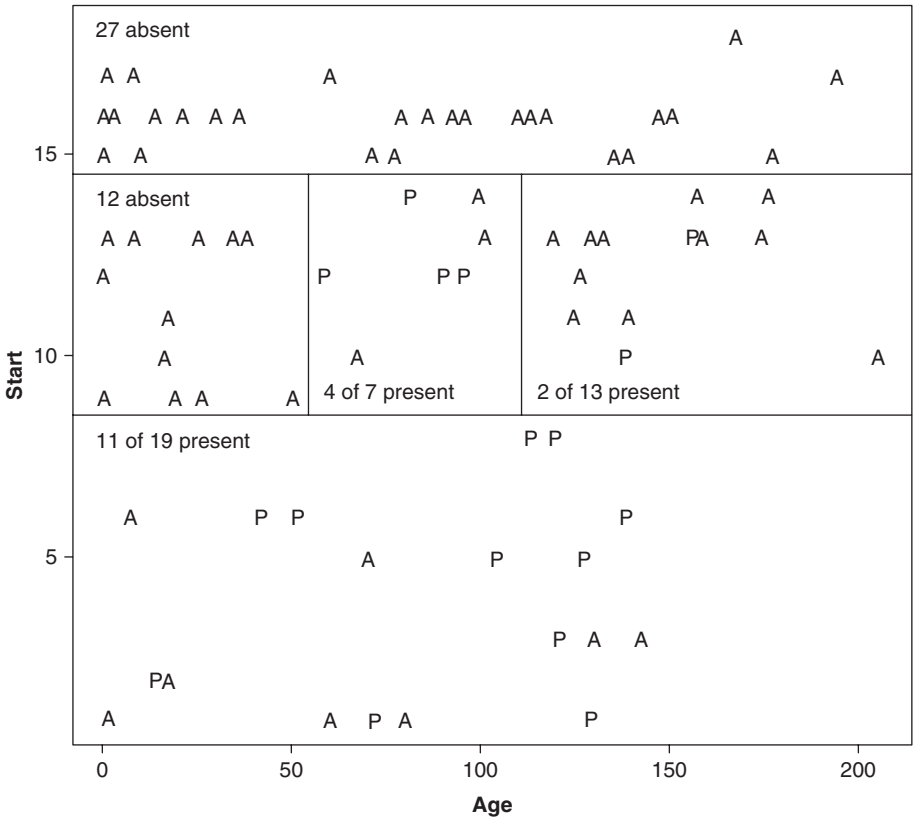


Figure 1 Partition of Kyphosis Data Set by Age and Starting Vertebrae

11.5 months, 12.5 months). For each split examined, the outcome variable is compared across the two sides—the side with cases where the age is greater than the cut point and the side where the age is less than the cut point. In the kyphosis data, 21% of the cases have kyphosis present. If a split of a variable resulted in one side having only kyphosis present (100%) and the other side having only kyphosis absent (0%), then this would be considered a perfect split and the algorithm would stop. This result would suggest that the most accurate prediction of kyphosis would be based entirely on the age of the child.

In practice, perfect splits are rarely found in data. However, criteria for deciding on the best split are available. One, the Gini diversity index, is a popular algorithm that is maximized for perfect splits and minimized for cases where half the cases in a partition are in one class and half are in the other. Any criteria used to determine the best split has the qualities that it can be quickly calculated and can be used to rank how well the split performed in classifying the outcome variable.

To recap, at the start of the algorithm recursive partitioning goes through every cut point for every input variable and records which variable by cut point combination produced the best split of the data into more homogeneous classes. This variable by cut point combination is selected as the first or initial split dividing the data into two partitions. In recursive partitioning, the algorithm is then repeated on both partitions independently to find the best variable by cut point combination that produces the best split for each side. One of the great strengths of the recursive partitioning approach is that this process proceeds independently on either side of a split, so that subsequent splits can be made on the same or different variables.

Recursive partitioning models produce a clear data representation in the form of a tree that details the partitioning process and data. For the kyphosis data, the recursive partitioning algorithm produces the tree presented in Figure 2.

At the top of the tree, all 81 cases are present and initially split into two subgroups based on the starting

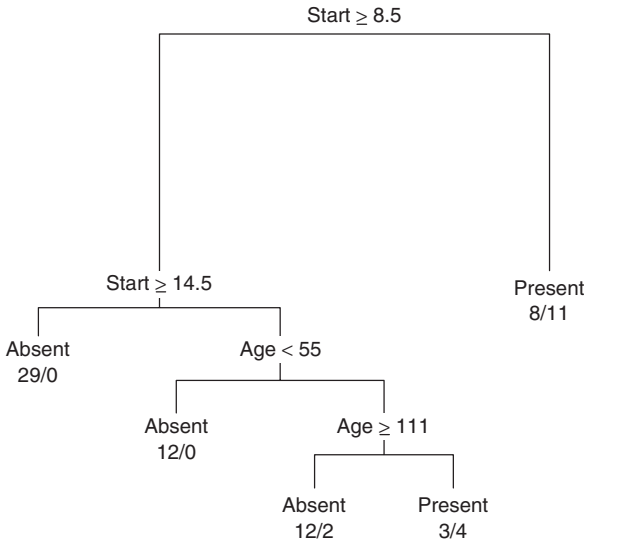


Figure 2 Classification Tree for Kyphosis Data

number of the vertebrae operated on. Cases whose starting number is greater than 8.5 (i.e., the ninth vertebrae or further) are classified into the left branch of the tree and those cases whose starting number is less than 8.5 (i.e., the eighth or lower vertebrae) are classified into the right branch. There were 19 cases classified into the right branch—8 with kyphosis absent and 11 with kyphosis present. These are not further partitioned, because no further splits significantly increased the proportion of cases with kyphosis present. This group is labeled kyphosis present because that diagnosis applies to the majority of its cases. If this model were used for prediction, future cases whose starting vertebrae number is less than 8.5 would be predicted to be kyphosis present. Cases classified into the left branch (starting number is greater than 8.5) were subsequently split on starting number greater than or less than 14.5. In the far left terminal node, which consists of cases whose starting number is greater than 14.5, all cases are always kyphosis absent (29 out of 29) indicating that this deformity is possibly limited to cases where the starting vertebrae is a low number. The other splits are followed as described above and result in all 81 cases being assigned to a single partition.

Regression trees, used when the outcome variable is continuous, are fit and interpreted in a similar manner. In this case, however, the criteria for making a split is calculated by some measure in the reduction in variance (e.g., reduction in mean squared error) and the terminal nodes are labeled by the mean of the observations for that partition. In terms of a conditional mean, as in parametric regression, the splits along a branch define the conditions.

—William D. Shannon

See also Clinical Epidemiology; Regression; Study Design

Further Readings

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (Eds.). (1984). *Classification and regression trees.* Boca Raton, FL: Taylor & Francis.

Calvocoressi, L., Stolar, M., Kasl, S. V., Claus, E. B., & Jones, B. A. (2005). Applying recursive partitioning to a prospective study of factors associated with adherence to mammography screening guidelines. *American Journal of Epidemiology*, *162*(12), 1215–1224.

Chambers, J. M., & Hastie, T. J. (Eds.). *Statistical models in S.* Pacific Grove, CA: Wadsworth.

Zhang, H., & Singer, B. (1999). *Recursive partitioning in the health sciences.* New York: Springer.

Clinical Epidemiology

Clinical epidemiology is the application of the methods and principles of epidemiology, which is focused on population health, to the practice of clinical medicine, which is focused on the health of particular individuals. Modern medical practice should at all times be predicated on the best available scientific evidence. But at no time is there, in fact, any scientific evidence that pertains directly and reliably to a particular, individual patient. Rather, scientific evidence is derived from the past experience of other patients, or of subjects in clinical trials of various design. The relative correspondence between the one patient now receiving care and the many patients or subjects from whom scientific evidence has been derived is therefore an assessment prerequisite to clinical decisions, and an interface where science in medicine must inescapably collide with judgment. Clinical epidemiology is both a philosophical approach to medical decision making and a collection of tools and techniques that may inform the practice of medicine. This entry reviews the principles of clinical epidemiology, describes the relationship between this field and clinical practice, and discusses the use of clinical epidemiology in determining what diagnostic tests to perform. It also

explores fundamental principles of statistics and probability, such as Bayes's theorem, that inescapably underlie medical decisions and can and should be used to inform and fortify them.

Evidence has securely claimed its place among the dominant concepts in modern medical practice. To the extent possible, clinicians are expected to base their decisions (or recommendations) on the best available evidence. The consistent application of evidence to clinical decision making rests on the traditional fault line separating clinical practice from public health. Despite efforts to bridge that gap, the philosophical divide between a discipline devoted to the concerns of populations and another devoted to the advocacy of an individual can seem impassable. But evidence is, in fact, the bridge, because evidence-based practice is population-based practice.

Evidence applied clinically derives from the medical literature, where the standards of evidence, and therefore practice, continuously evolve. But apart from case reports, what is reported in the literature is not the experience of *an* individual patient and certainly not the experience of *the* individual patient, but rather the experience of a population of other patients. Therefore, the practice of evidence-based medicine requires the application of population-based data to the care of an individual patient, different in ways both discernible and not, from the subjects whose experience is reported. All decisions made on behalf of (or, preferably, with) individual patients are extrapolation or interpolation from the prior experience of other patients. Clinical medicine is evidence based only if it is population based.

To base clinical decisions for an individual on the individual alone, the outcome of an intervention would need to be known in advance. In other words, medicine would need to borrow from astrology, or some alternative system of predicting future events. The choice of an initial antihypertensive drug for a hypertensive patient cannot be based, before the drug is prescribed, on the response of the patient in question. Nor can the benefits to the patient be known in advance. The drug is chosen based on the published results of antihypertensive therapy in other patients. The particular drug is selected based on the characteristics of the patient, and how closely they approximate those of others who have benefited from specific therapies. Once the drug is selected, while the therapeutic effect on the surrogate measure (e.g., blood pressure) is detectable, most outcome benefits to the patient (e.g., stroke prevention) remain unknowable. The physician can never identify the stroke that was prevented in an individual but can only surmise that the risk of stroke has been reduced for that individual.

Implicit in the concept of evidence being the derivative of population experience is the need to relate that experience back to the individual patient. The inapplicability of some evidence to some patients is self-evident. Studies of prostate cancer are irrelevant to female patients; studies of cervical cancer are irrelevant to male patients. Yet beyond the obvious is a vast sea of gray. If a patient is older than, younger than, sicker than, healthier than, taller, shorter, simply different from the subjects of a study, do the results pertain? Because an individual patient will never be entirely like the subjects in a study (unless they were, in fact, a subject, and even then their individual experience might or might not reflect the collective experience), will the results, the evidence, ever be truly pertinent? No degree of evidence will fully chart the expanse of idiosyncrasy in human health and disease, and to work skillfully with evidence is to acknowledge its limits.

Principles of Clinical Epidemiology

Every clinical decision derives in whole or in part from the tacit comparison of a patient to a population of patients, and the skill with which that comparison is made becomes a fundamental element in the skill with which medicine is practiced. Integral to that comparison is the capacity to recognize the defining characteristics of both patients and populations as the basis for defining the bounds of similarity and dissimilarity. Equally important is the capacity to evaluate the context in which evidence was gathered, to determine its applicability to a given clinical decision. An ability to evaluate the quality as well as the pertinence of evidence is essential. And of course, an ability to find reliably the best available evidence when one is uncertain about a clinical decision is prerequisite to its interpretation.

While clinical choices (for both testing and treatment) are predicated on, at a minimum, the knowledge, judgment, values, preconceived notions, experiences, preferences, and fears of both clinician and patient, clinical decision making is greatly influenced by three discrete considerations: probability, risk, and alternative. Probability is fundamental to such decisions, as patients are evaluated and treated only for a specific

condition or conditions it seems they might have. The physician will recommend a particular test or treatment when the patient seems likely enough to need it.

Some low-probability diagnoses are pursued because they pose such high risk. Physicians admit some patients to the hospital to evaluate for myocardial infarction (MI or heart attack) although they believe the probability of MI to be low, because the risk associated with undetected MI is high. Similarly, physicians are taught to do a lumbar puncture (LP, also known as a spinal tap) whenever they wonder, "Should I do an LP?" because of the devastating consequences of failing to identify a case of meningitis.

Finally, alternatives need to be factored in. When chest pain seems atypical for angina, but no alternative explanation is at hand, the physician is more apt to treat the pain as angina. When pneumonia is present to explain shortness of breath, the physician will be less inclined to work up pulmonary embolism (PE), despite pleuritic chest pain and tachycardia. By a process to which physicians are, for the most part, comfortably incognizant, they make every decision factoring in considerations of probability, risk, and alternatives.

But a process to which one is more or less inattentive is a process that cannot be optimally regulated. Because the physician's decisions rely on their evaluation of probability, risk, and alternatives, these parameters should become of sufficient interest to warrant conscious oversight. And each of these parameters is population based. There is no probability of genuine relevance to an individual; there is the rate of occurrence in populations and the degree of concordance between individual and population characteristics. There is no true individual risk; for an individual, an event occurs (100% risk) or does not (0% risk). There is, however, the comparability of the patient to groups in whom the event rate in question is higher or lower. And the alternatives available for an individual patient are those options and interventions applied under similar circumstances to other patients, with varying degrees of success.

Clinical Epidemiology and Clinician Performance

Consider that a patient presents to a medical office. How likely is it that the patient has coronary disease? Naturally, at this point, the clinician can't answer the question. There is as yet virtually no information about the patient. Yet the clinician would have probably already begun the process of generating an estimate. If the provider were a pediatrician, patients in his or her practice are unlikely to have coronary disease. Therefore, this patient presenting to you is unlikely to have coronary disease. A similar situation occurs if the provider is a specialist (other than a cardiologist) to whom patients are referred after passing through the filter of primary care. But even if the clinician in this case were an internist or family practitioner, he or she would have already started to consider the probability of coronary disease. For practitioners in the United States, many of their adult patients will have coronary disease; so, too, might the patient in question. In certain other countries, the probability of coronary disease may be so low that practitioners there seldom need to consider it and, thus, would not consider it in this patient.

Of note, even at this very preliminary stage of evaluation, is the role of bias, or, more bluntly, prejudice in clinical decision making. Clinicians reach decisions based on experience—either their own or that of others. Making inferences about an individual based on the prior experience one has had with others in the same population as that patient is the essence of prejudice, or prejudging. But this is not meant to have negative connotations in clinical practice. Prejudice—or rather a tendency to judge the probability of a diagnosis in an individual based on the probability of that condition in the population the patient comes from—is appropriate and essential. It would be foolish to consider coronary disease routinely in individual patients from a population in which coronary disease almost never occurred. The prejudice born of experience, and familiarity with the population in question, would influence clinical judgment, and decisions, in an appropriate way.

So one immediately begins to formulate an impression of probability based on the characteristics of other patients one has seen. But we want more information. In this case, we would like to know whether or not the patient has chest pain suggestive of angina. We would like to know the patient's age and gender; whether or not the patient is hypertensive or diabetic; whether the patient smokes, is sedentary, has hyperlipidemia; whether the patient has a family history of heart disease or is obese. In attempting to determine how probable coronary disease is in the patient in question, we ask questions that sequentially allow us to place the patient in the context of

populations in which coronary disease is more or less probable. If the patient has chest pain typical of angina pectoris and happens to be a 70-year-old male smoker with diabetes, hypertension, hyperlipidemia, and a family history of heart disease, we can now answer the question with considerable confidence: The probability of coronary disease is high. Our inferences about an individual patient are derived from the historical experience of other patients whom we believe to be much like our own.

The history and physical examination can be considered a process of sequential hypothesis testing. Each question asked tests hypotheses about the population from which the patient might come. Once this is acknowledged, there comes a point in the history when additional questions (and answers) cannot dissuade the physician from a particular conclusion. In the case under consideration, a point in the history would be reached when coronary disease would seem sufficiently probable to warrant further investigation. Even if the answers to subsequent questions were negative, lowering the probability of coronary disease, a suspicion—based on both probability and risk—might be great enough to warrant commitment to investigating the possibility of that condition. Recognizing this semi-quantitative element in decision making is essential to manage the result. For example, if the patient seemed very likely to have coronary disease, would the physician abandon that belief if the ECG were normal? Probably not. The answer would depend on how robust the clinical suspicion of coronary disease were, compared with the negative results of any testing done.

In the process of sequentially testing hypotheses about the patient, health care providers are in essence endeavoring to define as narrowly as possible the population of which the patient is representative. Once that goal is achieved, epidemiology can offer a fairly stable estimate of the probability of the particular condition under consideration. That estimate is the *prevalence* of the condition in the population on which we've settled. Prevalence, the proportion of a specified population with a particular condition at a particular point in time, is related to the probability of disease in an individual member of that population. *Incidence*, the number of new cases of a particular condition in a defined population during a given period of time (typically a year), is related to the risk of that condition in an individual member of that population.

Clinical epidemiology allows the physician to convert the population data of the epidemiologist into a concept of practical utility to patient care. The analogue of prevalence for the individual patient is the *prior probability*, the probability of the condition in question *prior to* any subsequent assessments that might be indicated to further test our hypothesis(es). In essence, there is a discrete probability of a condition prior to every question posed during the history that is modified by each answer to become an estimate of the *posterior probability*, the probability resulting from, or following, a test. Each such posterior probability becomes the prior probability estimate in advance of the next question. The questions should be tailored to the continuously revised probability estimate, so that the pertinent hypotheses are tested. The process of questioning, physical examination and laboratory testing, sequential hypothesis testing, and sequential conversion of prior probability, to posterior probability, back to prior probability, is in advance of the next maneuver.

This process of reasoning is guided both by knowledge of the reliability of particular medical tests and by knowledge of the prior probability of a particular result. If the prior probability of a particular condition is very close to 0, a much higher standard of technological evidence should be required to conclude with confidence that a condition has been *ruled in.* If a prior probability is very close to 1, a high standard of evidence should also be required to support the conclusion that disease can be *ruled out.* A test not reliable enough to alter a very high or low prior probability estimate is of little clinical utility. But worse, if the principles of clinical epidemiology are overlooked, the test is potentially harmful. A positive test, if interpreted identically in patients in whom probability of disease is high or low, will provide misinformation as often as information. Dissociated from the principles of clinical epidemiology, medical technology, whether invasive or noninvasive, poses very real threats to patient care.

Clinical Epidemiolgy and Test Performance

The performance of tests used to refine the probability of any particular diagnosis can be cast in terms of sensitivity and specificity. *Sensitivity* is the ability of a test to detect disease when it is present. *Specificity* is the ability of a test to exclude disease when it is absent.

The denominator for sensitivity is the presence of the condition in question (e.g., disease); sensitivity can tell us nothing about the test's performance in patients who are condition (disease) free. In terms of a population, the denominator for sensitivity is the prevalence, the proportion of the population with the condition. Of those with the condition, some will test positive (true positives), and some will test negative (false negatives). Sensitivity is the proportion of disease positives in whom the test is positive. If a test is negative in a patient with disease, it is a false negative result.

Specificity pertains to the proportion of the population that is free of disease. In comparable terms, the denominator for specificity, all disease-free individuals, is 1 − prevalence. A test is specific when it reacts only to the singular condition under investigation. The proportion of those free of disease identified as disease free is the specificity; those who are disease free but have a positive test result are false positives. Note, then, that among those with disease there are true positives and false negatives; sensitivity defines the rate at which true positives are identified. Among those free of disease, there are true negatives and false positives.

Test characteristics are an important element in medical screening programs, defined as efforts to assess for early disease in a population with no outward signs of pathology. While intuition might suggest otherwise, highly sensitive tests are useful for ruling out pathology, and highly specific tests for ruling it in. The logic is as follows. A highly sensitive test reliably detects disease when present. A highly sensitive test is thus almost always positive when disease is present, and hardly ever negative when disease is present. Therefore, if a highly sensitive test is negative, it is almost certainly because disease is truly absent. A positive result of a highly specific test reliably rules in disease for precisely the same reasons.

Most diagnostic studies are subject to interpretation in light of a cutoff point. Virtually any laboratory test result is more suggestive of disease when it is well outside the normal range than when it is just outside the normal range. Does the abnormal test result warrant treatment or further workup? If the result is just outside the normal range, perhaps not. Why? Because at such a level, many disease-free individuals might also have a "positive" test. In other words, a cutoff point that makes it easy to detect disease when it is present makes it comparably easy to detect disease mistakenly when it is absent.

Conversely, what happens when the apparently healthy ("normal") patient has an extreme result on a routine laboratory test? Because a cutoff far outside the normal range makes it likely that some true cases of disease will be missed, but unlikely that disease-negative individuals will test positive, a positive result almost certainly indicates disease (or laboratory error). Further investigation is clearly warranted. While such interpretations are second nature to any experienced clinician, knowing the statistical properties that underlie such decision making fortifies the process. When a test result is sufficiently outside the normal range to warrant careful evaluation, but not so clearly abnormal as to provide definitive evidence of disease, an understanding of the variation in sensitivity and specificity with variation in the cutoff point can help determine when to test further, when to treat, and when to observe expectantly.

Generally, when a disease is serious, when detection is important, and when false positives can be distinguished from true positives with follow-up testing that is readily available, a relatively low cutoff point is desirable. When a disease is less serious or indolent, when detection of disease is not urgent, and/or when distinguishing false from true positives is difficult without costly or invasive procedures, a relatively high cutoff point is desirable.

Quantifying Clinical Uncertainty

Clinical medicine remains challenging, and always will, because no two patients are just alike, and an exact match between any patient and a population is unachievable. Nonetheless, the ability to diagnose at all relies on approximating a match. How does a clinician know that someone with rhinorrhea (a runny nose) has a viral upper respiratory infection rather than a cerebrospinal fluid (CSF) leak? In fact, they do not actually *know*; they *infer* that because colds are common, CSF leaks are rare, and patients with colds share characteristics with all the other patients treated for colds, that this particular patient with rhinorrhea has a cold. Clinicians rarely can or do apply pathognomonic tests (i.e., tests that render a diagnosis with complete accuracy). Rather, providers overcome the discomfort of residual uncertainty with estimates of probability. And these estimates are population derived; all clinical decisions involve the application of what is known about populations to what is suspected about an individual.

Bayes's Theorem

This concept is rendered statistically useful by Bayes's theorem, developed in the 18th century by Thomas Bayes. In principle, the theorem asserts that the probability of any condition in an individual is related to the probability of that condition in the population of which the individual is a member (the underlying population prevalence of the condition). The theorem has been modified to indicate that the result of any diagnostic test alters the probability of disease in the individual patient, effectively because each successive result reclassifies the population from which the individual comes (i.e., a patient with chest pain and a negative cardiogram versus a patient with chest pain and an ischemic cardiogram). In its basic form, the theorem is expressed as a series of conditional probabilities:

$$PD+ = [(pT+|D+)pD+]/\{[(pT+|D+)pD+] + [(pT+|D-)pD-)]\}$$

where

$PD+$: the posterior probability of disease (aka posttest probability)
$|$: given that; a symbol of conditional probability
$pT+$: the probability of a positive test
$D+$: disease is present
$pD+$: the prior probability of disease (aka pretest probability or prevalence)
$D-$: disease is absent
$pD-$: the probability of nondisease; $1 -$ prevalence

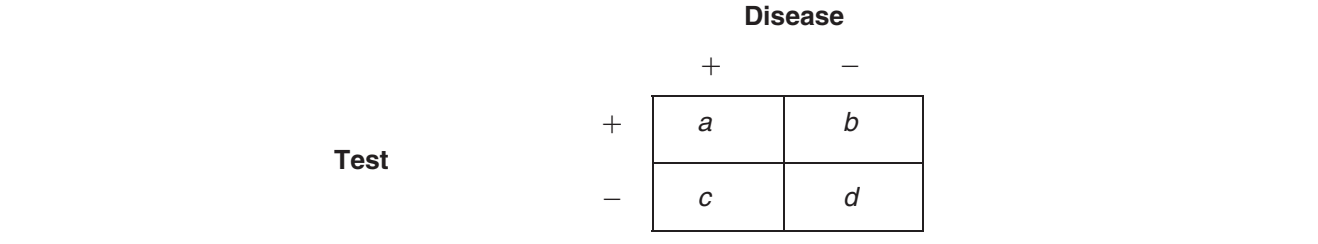
Deconstructed, Bayes's theorem becomes less intimidating. Consider the numerator $[(pT+|D+)\ pD+]$. The probability of a test being positive *given that* (conditional probability) the disease is present is the test sensitivity. Expressed in a 2×2 contingency table (Figure 1), the probability of a positive test in those who have disease is simply cell a, the true positives. Bayes's theorem is now

$$PD+ = \text{true positives}/\{\text{true positives} + [(pT+|D-)pD-]\}$$

or

$$PD+ = a/\{a + [(pT+|D-)pD-]\}$$

We've made good progress because the first term in the denominator of the theorem is the same as the numerator term. What of the second term in the denominator $[(pT+|D-)pD-)]$? As shown in Figure 1, this is the probability of a positive test result among those without disease. This term reduces to cell b of the table shown in Figure 1, the



$D+$:	$a+c$
$D-$:	$b+d$
$_PT+\|D+$:	a
$_PT+\|D-$:	B
$_PD+$:	$(a+c)/N$
$_PD-$:	$(b+d)/N$

Figure 1 The 2×2 Contingency Table and Bayes's Theorem

term representing the false positives. The theorem can now be expressed as

$$PD+ = \text{true positives}/(\text{true positives} + \text{false positives})$$

or

$$PD+ = a/(a+b)$$

Thus, Bayes's theorem asserts the following: The probability that an individual with a positive test result truly has disease is the proportion of all positive test results (true and false) that are true positives. This is the same formula as the positive predictive value, the probability that a positive test result truly indicates the presence of disease.

There is one other useful way of expressing the theorem. The probability of a positive test in those with the disease is the sensitivity, while the probability of disease is the prevalence. Thus, the formula can be expressed as

$$PD+ = [\text{sensitivity} \times \text{prevalence}/[\text{sensitivity} \times \text{prevalence} + [(pT+|D-)pD-)].$$

The probability of a positive test result among those truly free of disease is the false positive error rate, or 1 − specificity, and the probability of disease absence is 1 − prevalence. The formula can now be converted to

$$PD+ = [\text{sensitivity} \times \text{prevalence}/\{[\text{sensitivity} \times \text{prevalence}] + [(1-\text{specificity})(1-\text{prevalence})]\}.$$

In this form, Bayes's theorem permits calculation of the posterior probability of disease, the probability of disease after testing, provided one knows the prior probability, and the sensitivity and specificity of the diagnostic test. Working with Bayes's theorem rather than just the positive predictive value offers both conceptual and logistical advantages. Conceptually, Bayes's theorem emphasizes the importance of prior probability to the probability of disease after testing. Logistically, Bayes's theorem allows us to determine the posterior probability of disease without resorting to a 2 × 2 table.

Application

Consider three patients suspected of having deep venous thrombosis (DVT) on the basis of history and physical examination. Patient 1 has risk factors for DVT, such as a long period of immobility or a known hypercoagulable state. Patient 2 has no obvious risk factors for DVT. Patient 3 is therapeutically anticoagulated on warfarin for a prosthetic heart valve.

If each of these three patients presented with identical signs and symptoms of DVT in the left lower extremity, our estimates of the probability of DVT would likely vary. The first patient seems very likely to have DVT, because the patient comes from a population of patients in which the occurrence of DVT is known to be high. The second patient has a moderate or intermediate probability of DVT. The third patient, other things being equal, should be at low risk for DVT given that the INR (the International Normalized Ratio, a standard measure of how much blood has been "thinned" by warfarin) level is therapeutic.

Assume that the first patient is estimated to have an 80% probability of DVT. Further assume that the venous Doppler ultrasound is ordered and that it has a sensitivity of 80% and a specificity of 75% for clot above the knee. The test is positive. How does this affect our diagnosis?

Bayes's theorem provides an answer, while forcing us to consider that the test result must modify, rather than replace, our prior probability estimate. The posterior probability of disease is

$$PD+ = \frac{[\text{sensitivity} \times \text{prevalence}]}{\left[\begin{array}{l}\text{sensitivity} \times \text{prevalence}] \\ +[(1-\text{specificity})(1-\text{prevalence})\end{array}\right]}.$$

In this scenario, the sensitivity is 80%, the prevalence is 80%, and the specificity is 75%. The formula then becomes

$$PD+ = (0.8 \times 0.8)/\{(0.8 \times 0.8) + (0.25 \times 0.20)\}$$
$$PD+ = 93\%.$$

Thus, after a positive Doppler ultrasound test, the probability of DVT is 93%.

What if the same test, with the same operating characteristics, were applied to the second patient, who appears to be at intermediate risk for DVT? We can use a prior probability of 50% to convey our substantial uncertainty. In this case,

$$PD+ = (0.8 \times 0.5)/\{(0.8 \times 0.5) + (0.25 \times 0.5)\}$$
$$PD+ = 76\%.$$

Note that the posterior probability of DVT in this patient, that is, the probability after a positive diagnostic study, is lower than the prior probability in the patient in the first scenario. Also of note is the much greater change between prior and posterior probability estimates than in the first scenario. When our prior probability estimate was 80%, a fairly reliable diagnostic test increased our posterior probability by 13%. Using the same test, but with a prior probability estimate of 50%, the posterior probability increased 26%, or twice as much. In general, the greater the degree of diagnostic uncertainty (i.e., the closer the prior probability estimate is to 50%), the more helpful the diagnostic testing becomes, and the more test results will modify the posterior relative to the prior probability.

The patient who is therapeutic on warfarin but with a presentation suggestive of DVT has a low prior probability of the condition; we can estimate it at 10%. The same venous Doppler testing is applied, with the same operating characteristics. For this patient, the posterior probability is

$$PD+ = (0.8 \times 0.1)/\{(0.8 \times 0.1) + (0.25 \times 0.9)\}$$
$$PD+ = 26\%.$$

This demonstrates the hazard of simply replacing our clinical estimates with the results of diagnostic testing. Despite a positive Doppler ultrasound (given the operating characteristics provided), the patient in question is far more likely *not* to have DVT, than to have DVT. In fact, three out of four similar patients will be free of DVT.

Confronted with such a scenario, one is generally tempted to reconsider the prior probability estimate. If the test is positive, then the patient seems to have the disease in question, and therefore must have been more likely to have the disease in question than I originally thought.

This kind of rethinking is both important and dangerous. It is important because the prior probability estimate is just that, an estimate, and one we tend not to like making in a strictly quantitative manner in the first place. Reconsideration of the prior probability estimate is appropriate. But if the prior probability estimate is, in fact, reasonable, it should not be discarded or replaced based on the results of diagnostic testing. This is an inappropriate use of test results and gives them influence over the diagnostic process they should not have. Test results influence the probability of disease by converting the prior probability to a posterior probability; if the same test is used also to recalibrate the prior probability estimate, the same test result is effectively being used twice. Test results should be interpreted in light of the prior probability of disease, not used to replace that estimate. That estimate may be revisited but on the basis of clinical judgment, rather than a test result.

Implications of Bayes's Theorem for Diagnostic Testing

Does judicious application of Bayes's theorem to the process of workup and diagnosis require that we include a calculator in our black bags? The answer is a resounding *no* (although there are times it can come in handy) because there is generally substantial uncertainty about the prior probability estimate. Therefore, the use of the theorem to generate a very accurate posterior probability of disease is unreliable. What is reliable is the theorem's capacity to demonstrate the interaction of prior probability estimate and test result. If one is very confident that disease is present, only a truly exceptional test result can make one substantially more certain. Similarly, if the test is negative, it can refute that impression reliably. Bayes's theorem provides mathematical evidence to support the concept that the value of diagnostic testing depends both on the performance characteristics of the test and the clinical scenario in which the test is ordered.

Astute use of the Bayes's theorem, or at least the concepts behind it, should at times result in more diagnostic testing, and at times less. When one might be inclined to order a test despite considerable confidence that the disease is present or absent, the theorem should highlight the extent to which test results will modify that level of confidence. When the test result is unlikely to change the clinical impression, the test may well be unnecessary. Alternatively, when a needed test yields a surprise result, the theorem would argue against abandoning the prior probability and replacing it with the test result. In this situation, further, confirmatory testing is likely to be required.

Finally, how reliably can the theorem, and its application, be when the whole process begins with a prior

probability estimate that is seldom more than an educated guess? This seeming weakness is a strength in the application of Bayes's theorem, because it forces clinicians to confront the issue of uncertainty rather than ignore it. That the estimates among a group of clinicians vary is in a way reassuring; no matter how statistics and evidence are applied, much of clinical practice will be shaped by judgment. Application of Bayes's theorem is not intended to eliminate uncertainty in the diagnostic process, but rather to disclose it, so we can wrestle with it more effectively.

One way to "wrestle" directly with the uncertainty in the prior probability estimate is to conduct a sensitivity analysis. This term refers to any situation in which the a priori assumptions are modified across some range of plausible values to determine whether conclusions are altered. In the case of Bayes's theorem, one can vary the prior probability estimate over a range of reasonable values. What is reasonable is the product, again, of judgment, but is bounded by 0 and 1 (100%).

While one might be ill at ease in any given clinical situation to say, for example, that the prior probability of disease was 22%, one might be quite secure that it was between 5% and 50%. By interpreting the test result over the range of prior probability estimates, one can determine if the conclusion remains the same or changes.

Consider a test with 95% sensitivity and specificity, which is used to diagnose a condition for which treatment is absolutely necessary, but invasive and painful. If the test, for example, an MRI, is positive and the prior probability is set at 5%, the positive test results in a posterior probability of 0.5, or 50%. While this is a substantial increase, it hardly is adequate to justify invasive therapy. If the prior probability is set at 50%, the posterior probability is 0.95, or 95%. We might establish a "rule of thumb": If the prior probability estimate is reliably greater than 50%, and the MRI is positive, the disease is almost certainly present, and treatment is indicated. If the prior probability is substantially lower than 50%, a positive MRI will suggest a need for further, confirmatory testing, before invasive therapy is initiated.

Conceptual Factors Influencing Probability Estimates

Bayes's theorem is concerned with probability. But as noted earlier, probability is just one of the three concepts that should share comparable prominence in the diagnostic process. The other two are risk and alternatives. When a disease is very serious and can be modified with treatment (e.g., meningitis), the need to pursue a workup cannot be discounted just because the condition is relatively improbable. However, if a condition is truly trivial, then no matter how probable, diagnostic testing is unlikely to be valuable. And if a condition is extremely improbable, then no matter how serious, it should probably not be pursued. That we conduct our workups this way is clear; not every patient with fever or headache has a lumbar puncture.

And this is where the third concept comes in: alternatives. When an alternative diagnosis is very probable, then the diagnosis under consideration becomes less so. If the patient with fever and a headache appears to have a viral upper respiratory illness (i.e., a cold), then the probability that the symptoms are due to meningitis is that much less. Once a "cold" is diagnosed with confidence, the consideration of meningitis is resolved. But what of situations in which no good alternative suggests itself? For example, what of the patient with chest pain atypical for angina, but at least equally atypical for a gastrointestinal or musculoskeletal source? The lack of alternatives makes the principal diagnosis(es) in question more probable. Thus, the building blocks of a solid diagnosis are considerations of probability, alternatives, and risk.

The practice of clinical medicine, dedicated as it is to the care of individuals, cannot avoid population-based principles. Physicians cannot practice clinical medicine and avoid the practice of clinical epidemiology. But the discipline of evidence-based practice or clinical epidemiology (the terms might reasonably be used interchangeably) is not one in which most clinicians have had the necessary initiation. The art of clinical decision making is judgment, an even more difficult concept to grapple with than evidence. As the quality and scope of evidence to support clinical interventions is, and will likely always remain, limited in comparison to the demands of clinical practice, the practice of evidence-based medicine requires an appreciation for the limits of evidence, and the arbiters of practice at and beyond its perimeters. Judgment fortified by the highest standards of decision-making science is a force to be reckoned with, enabling clinicians to achieve the best possible results. And ultimately, that is the validation of evidence-based

practice, or population-based practice, or clinical epidemiology: the outcomes to which such concepts contribute. Rigorous reasoning is the means, desirable outcomes the ends.

—David L. Katz

See also Bayesian Approach to Statistics; Bayes's Theorem; Evidence-Based Medicine; Sensitivity and Specificity

Further Readings

Dawson-Saunders, B., & Trapp, R. G. (1994). *Basic and clinical biostatistics* (2nd ed.). Norwalk, CT: Appleton & Lange.

Eddy, D. M. (1996). *Clinical decision making. From theory to practice*. Sudbury, MA: Jones & Bartlett.

Fletcher, R. H., Fletcher, S. W., & Wagner, E. H. (1996). *Clinical epidemiology: The essentials* (3rd ed.). Baltimore: Williams & Wilkins.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1996). *Fundamentals of clinical trials* (3rd ed.). St. Louis, MO: Mosby-Year Book.

Guyatt, G. H., Rennie, D., & American Medical Association. (2002). *Users' guides to the medical literature*. Chicago: American Medical Association. Retrieved June 28, 2007, from http://books.google.com/books?id=UOnNshaw8KcC&dq=Users%27%Guide+to+the+Medical+Literature&pg=PA191&ots=_MxmuHSWmh&sig=6bh22NEZWkfcViVYj6bNMMUvFaQ&prev=http://www.google.com/search%3Fsourceid%3Dnavclient%26ie%3DUTF-8%26rls%3DTSHA,TSHA:2006-02,TSHA:en%26q%3DUsers%27%2BGuide%2Bto%2Bthe%2BMedical%2BLiterature&sa=X&oi=print&ct=result&cd=1).

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., Hearst, N., & Newman, T. B. (2001). *Designing clinical research* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

Jekel, J. F., Elmore, J. G., & Katz, D. L. (1996). *Epidemiology, biostatistics, and preventive medicine*. Philadelphia: W. B. Saunders.

Katz, D. L. (with Nawaz, H., & Grecie, L.). (2001). *Clinical epidemiology and evidence-based medicine: Fundamental principles of clinical reasoning and research*. Thousand Oaks, CA: Sage.

Lurie, J. D., & Sox, H. C. (1999). Principles of medical decision making. *Spine*, *24*, 493–498.

Motulsky, H. (1995). *Intuitive biostatistics*. New York: Oxford University Press.

Petitti, D. B. (2000). *Meta-analysis, decision analysis, and cost-effectiveness analysis. Methods for quantitative synthesis in medicine* (2nd ed.). New York: Oxford University Press.

Reiser, S. J. (1997). Topics for our times: The medicine/public health initiative. *American Journal of Public Health*, *87*, 1098–1099.

Reiser, S. J. (1999). Medicine and public health. Pursuing a common destiny. *Journal of the American Medical Association*, *276*, 1429–1430.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P. (1991). *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston: Little, Brown.

Sackett, D. L., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (1997). *Evidence-based medicine. How to practice and teach EBM*. New York: Churchill Livingstone.

Speicher, C. E. (1998). *The right test* (3rd ed.). Philadelphia: W. B. Saunders.

U.S. Preventive Services Task Force. (1996). *Guide to clinical preventive services* (2nd ed.). Alexandria, VA: International Medical Publishing.

Clinical Trials

Clinical trials are experimental studies performed mainly on humans, but sometimes on animals, tissues, and cultures as well, to assess the effectiveness and safety of an intervention under investigation such as a new drug, diagnostic procedure, surgical procedure, prophylactic procedure, or screening program. There are many types of clinical trials. This entry focuses on randomized controlled clinical trials and randomized crossover designs for studying treatment effects. Rigorously designed clinical trials such as randomized concurrently controlled clinical trials conducted with human subjects, usually patients, have become well established as the scientific method based on empirical evidence that investigators must use to assess new treatments if their claims are to find widespread acceptance. Clinical trials have become indispensable in discovering new techniques to prevent and treat diseases, and their applications have been largely responsible for the compression of morbidity and decline in mortality rates at advanced ages in recent years.

Types of Trials

In *experimental studies*, the investigator manipulates the study factor (exposure groups) and randomly allocates experimental units to different exposure groups. An experimental study is also called a trial. A trial must progress longitudinally in time from exposure to outcome. This together with the ability to manipulate the study factor and to randomize the experimental units makes a stronger causal inference using experimental studies than using *quasi-experimental studies*,

which involve manipulation of the study factor but not random allocation of experimental units, or using *observational studies*, which involve neither manipulation of the study factor nor randomization of experimental units. Experimental studies without control or comparison groups (e.g., Phase I/II clinical trials) are called *uncontrolled trials*, while experimental studies with control or comparison groups are called *controlled trials.* A controlled trial in which the allocation of experimental units to different exposure groups is done randomly so that all experimental units have an equal chance of being allocated to each of the exposure groups is known as a *randomized controlled trial* (RCT), which can be a randomized controlled clinical trial (when the unit of randomization is a patient such as the Phase III clinical trial); a randomized controlled field trial (when the unit of randomization is a normal individual, rather than a patient, such as the vitamin C trial of Karlowski et al., 1975); or a randomized controlled community (cluster—a group of people in a community) trial (when the unit of randomization is a community or cluster such as Project Burn Prevention trial of Mackay and Rothman and water fluoridation trial of Ast and Schlesinger (1956). The less common field trials and community (cluster) trials are aimed at the evaluation of primary preventives, while the more common clinical trials are used to evaluate treatment effectiveness of a disease or to find a preventive of disease recurrence or death. All randomized trials are controlled. In nonrandomized controlled trials, the groups may not be comparable due to selection and confounding biases and so estimates of effects of the intervention may not be valid without further statistical adjustments.

Clinical Trial Protocols

Researchers for each clinical trial follow a protocol reviewed and approved by an institutional review board (IRB), a separate board of scientists, statisticians, physicians, and nurses who are not associated with the clinical trial. A Clinical Trial Protocol contains a study plan that describes the organization of a clinical trial, the background, rationale, objectives, and hypotheses; how the subjects are to be selected and how data are to be collected; primary exposures and methods of their measurements (an exposure is a factor that either causes, prevents, or treats an outcome), outcomes of interest and methods of their measurements; as well as intervening variables and methods of their measurements, type of study design, method of randomization, methods to control confounding bias prior to data analysis if randomization is not used, measures of association to be used, statistical methodology and analysis, including methods to control confounding bias during data analysis, and power calculations. It may also address issues of noncompliance, dropout as well as selection and information biases, and how nondifferential misclassification may affect the interpretation of results. The protocol also states the number of participants, eligibility requirements, agents that will be used, dosages, schedule of tests, the length of the study, and the larger population to which the results can be generalized. Note that both issues of internal validity and external validity have been covered in the description of clinical trial protocol given above. This protocol will also serve as the basis for writing the final report.

Randomized Controlled Clinical Trials

Essentially, a randomized controlled clinical trial is conducted by recruiting a group of patients from a target population. With adequate allocation concealment to protect against selection bias, the consenting eligible patients from the group of recruited patients are then randomly allocated to the treatment and control arms, which are then followed to the end of the trial with outcomes between the different arms compared. The patients recruited from the target population must meet the required eligibility criteria (inclusion and exclusion criteria) and, for ethical reasons, must give their informed consent before their randomization into different treatment groups to avoid selection and confounding biases. This ensures that the difference in treatment groups is caused by the difference in treatments alone. A concurrent control arm is needed so that outcomes with and without treatment(s) can be compared.

The choice of the control group will have an impact on research question and sample size. Use of a placebo in the control arm would help achieve blind treatment allocation and exclude placebo or Hawthorne effects, but ethical considerations demand that the control group should be the established treatment for the disease under study provided that its therapeutic effects have been well documented. Blinding is also used to exclude detection and performance biases, which occur when the investigator/patients know the treatment being given and which could

affect assignment, assessment, or compliance. These biases occur particularly when a subjective outcome variable such as pain or quality of life is measured.

Noncompliance—the failure to follow protocol requirements—can (1) result in a smaller difference between the treatment and control arms than truly exists, thereby diluting the real impact of a treatment, and (2) reduce study power, making it harder to detect an effect when it exists. Compliance measures can be used to improve estimates of treatment effects. The real strength of a randomized controlled clinical trial lies in the randomization. With sufficient sample size, randomization in which each treatment group is equally likely to be allocated to each patient would produce close similarity across the groups in all respects other than the intervention, including the unmeasured and unknown factors. That is to say, by randomization, both known and unknown confounders are controlled at the outset. This strengthens the validity of the causal inference. To preserve the baseline comparability, to maintain the statistical power of the original study population, and to ensure unbiasedness against noncompliance, intent-to-treat analysis should be conducted. Such analysis, also known as treatment assignment analysis, gives information on the effectiveness of the treatment under real-life conditions, unlike efficacy analysis, which determines the treatment effects under ideal conditions. In an intent-to-treat analysis, all participants who are randomized to a treatment are analyzed, regardless of whether they complete or even receive the treatment. Given the covariate measures at baseline and during follow-up monitoring, the statistical methods for analyzing the treatment effects of randomized controlled clinical trials depend on what endpoints are chosen to measure outcomes. For example, if disease incidence or death or occurrence of some event is the endpoint, logistic or Poisson regression models may be used to analyze the data; such analysis would also serve to control the residual confounding during the analysis. If survival time is the endpoint, then the Cox regression models as well as counting processes martingale methods may be used. If the endpoint consists of continuous response measures, then simple two-sample t tests or, when several treatments are involved, analysis of factorial experiments may be used. If repeated measures are involved, then methods for longitudinal data analysis such as random coefficient analysis may be used.

Parallel Group Designs

Most clinical trials belong to the so-called parallel group designs in which subjects in different arms are followed in parallel. Parallel group designs have four phases. Phase I tests a new drug or treatment in a small group of people (usually less than 10 normal volunteers), using an uncontrolled trial, the purpose being to learn how to administer a treatment safely and to determine optimal dosage or the so-called maximally tolerated dose (MTD) based on dose escalation investigations. Phase II expands the study to a larger group of people (often 30 to 40 patients or normal volunteers), the purpose being to test patient responses, to monitor side effects, and to determine the minimum effective dose (MED). These may be either uncontrolled trials or RCTs with patients who are expected to benefit from the treatment as experimental units. Phase III expands the study to an even larger group of people (usually running from hundreds to thousands of patients) for a full-scale evaluation of treatment using a randomized controlled clinical trial, as described above, with control group being either placebo or standard treatment. Patients are also closely monitored for severe adverse side effects for possible cancellation of the trial. Phase IV takes place after the drug or treatment has been licensed and marketed—postmarketing surveillance. It typically compares two treatments that are approved for similar uses to determine which one is more effective. It may also be conducted to study long-term safety and efficacy and to study new uses or cost-effectiveness of FDA-approved treatment.

Crossover Designs

In the design of RCTs, the opposite of parallel group designs are the crossover designs, which involve a switch of study treatments for each participant in a clinical trial. Thus, the participants serve as their own controls, which allow more precise estimates of treatment effects by making comparison within subjects rather than between subjects. Most crossover studies are *planned* crossover studies in which the predetermined period of treatment before switching to another treatment is specified in advance in the protocol. Noncurable medical conditions such as asthma and diabetes are suitable candidates for planned crossover studies as they present the possibility of giving more than one treatment to each patient. Other crossovers

are *unplanned* such as when patients in medical care are switched to surgical treatment because of deterioration in their condition. In planned crossover studies, there are as many groups as there are permutations of treatment sequences so that each group is uniquely determined by its first treatment. Each participant is randomized to receive his or her first treatment so as to produce roughly equal numbers of participants in each group. They are then followed for a predetermined period of time and then switched to a different treatment for another predetermined period of time and so on. Each period is of the same length for all participants, with the crossover point being blinded where possible. At the end of the study, outcomes are compared across treatments by combining the responses to each treatment from different groups. Thus, in parallel group designs, each group receives just one treatment and the treatments are administered concurrently, while in crossover designs, each group receives all treatments one after another and the treatment order differs across the groups. To avoid the so-called carryover effects in planned crossover studies, a washout period may intervene between treatments to allow the human body time to metabolize and excrete the previous treatment. The double-blind, two-treatment, two-period crossover designs are particularly popular in clinical pharmacology. These can be analyzed as special cases of repeated measure analysis of variance. The hypothesis of no treatment difference is tested simply by using a two-sample t test to compare the two sets of within-patient difference summed over all individuals. The main advantage of this design is the gain in statistical power, namely, that it can achieve statistically significant results with fewer subjects than would be required with a parallel design because the sample size required for this design depends on the variability within subjects, not between subjects as in the case of parallel design. (For most response variables, the within-subject variance is smaller than the between-subject variance.) Note that in the statistical analysis of treatment differences of crossover studies we assume no treatment-period interaction (which implies no carryover effects).

Sequential Trials

Clinical trials are costly and time-consuming. The idea of stopping such trials early when the treatments are being found to be unsafe or ineffective has been pursued on ethical and economic grounds. Wald's sequential procedure has been used to achieve this aim. Sequential trials in which the data are analyzed after each patient's results become available, and the trial continues until a clear benefit is seen in one of the comparison groups, or it becomes clear that no significant difference between the groups will emerge serve to achieve this aim. Thus, unlike parallel group and crossover designs, the number of subjects studied is not fixed in advance in sequential trials. Flexible sequential trials based on the discretization of Wiener processes and the group sequential designs implemented with a boundary for the null hypothesis (futility boundary) and a boundary for the alternative hypothesis (efficacy boundary) allow early stopping favor in either of the null or of the alternative as soon as conclusive interim evidence of being efficacious or not became available. Sequential trials and development of monitoring strategy that allows for interim looks at the accumulating data so as to shorten the average length of the trial are discussed in the articles by Jennison and Turnbull, Lan and DeMets, Pampallona and Tsiatis, Pocock, O'Brien and Fleming, and Whitehead listed in the further readings below.

—John J. Hsieh

See also Analytic and Descriptive Epidemiology; Bias; Community Trial; Randomization; Study Design

Further Readings

Ast, D. B., & Schlesinger, E. R. (1956). The conclusion of a ten-year study of water fluoridation. *American Journal of Public Health*, *46*, 265–271.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials* (3rd ed.). New York: Springer.

Jennison, C., & Turnbull, B. W. (2000). *Group sequential methods with application to clinical trials.* New York: Chapman & Hall.

Jones, B., & Kenward, M. G. (1989). *Design and analysis of cross-over trials.* London: Chapman & Hall.

Karlowski, T. R., Chalmers, T. C., Frenkel, L. D., Kapikian, A. Z., Lewis, T. L., & Lynch, J. M. (1975). Ascorbic acid for the common cold: A prophylactic and therapeutic trial. *Journal of American Medical Association*, *231*, 1038–1042.

Lan, K. K. G., & DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, *70*(3), 659–663.

MacKay, A., & Rothman, K. (1982). The incidence and severity of burn injuries following Project Burn Prevention. *American Journal of Public Health*, *72*, 248–252.

Meinert, C. L. (1986). *Clinical trials. Design, conduct, and analysis.* New York: Oxford University Press.

Newell, D. J. (1992). Intent-to-treat analysis: Implication for quantitative and qualitative research. *International Journal of Epidemiology*, *21*, 837–841.
O'Brien, P. C., & Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, *35*, 549–556.
Pampallona, S., & Tsiatis, A. A. (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favour of the null hypothesis. *Journal of Statistical Planning and Inference*, *42*, 19–35.
Piantadosi, S. (2005). *Clinical trials: A methodological prospective* (2nd ed.). New York: Wiley.
Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, *64*(2), 191–199.
Pocock, S. J. (1983): *Clinical trials: A practical approach.* Chichester, UK: Wiley.
Senn, S. (1993). *Cross-over trials in clinical research*. Chichester, UK: Wiley.
Whitehead, J. (1999). A unified theory for sequential clinical trials. *Statistical Medicine*, *18*, 2271–2286.

Cluster Analysis

Cluster analysis (CA) is an exploratory data analysis set of tools and algorithms that aims at classifying different objects into groups in a way that the similarity between two objects is maximal if they belong to the same group and minimal otherwise. In biology, CA is an essential tool for taxonomy of plants, animals, or other specimens. In clinical medicine, it can be used to identify patients who have diseases with a common cause, or patients who should receive the same treatment or who should have the same level of response to treatment. In epidemiology, CA has many uses, such as finding meaningful conglomerates of regions, communities, or neighborhoods with similar epidemiological profiles, when many variables are involved and natural groupings do not exist. In general, whenever one needs to classify large amounts of information into a small number of meaningful categories, CA may be useful.

Researchers are often confronted with the task of sorting observed data into meaningful structures. CA is an inductive exploratory technique in the sense that it uncovers structures without explaining the reasons for their existence. It is a hypothesis-generating rather than a hypothesis-testing technique. Unlike discriminant analysis where objects are assigned to preexisting groups on the basis of statistical rules of allocation, CA generates the groups, or "discovers" a hidden structure of groups within the data.

Classification of Methods

In a first broad approach, CA techniques may be classified as *hierarchical*, if the resultant grouping has an increasing number of nested classes, which resemble a phylogenetic classification, or *nonhierarchical*, if the results are expressed as a unique partition of the whole set of objects.

Hierarchical algorithms can be divisive or agglomerative. A divisive method begins with all cases in one cluster. This cluster is gradually broken down into smaller and smaller clusters. Agglomerative techniques usually start with single-member clusters that are successively fused until one large cluster is formed. In the initial step, the two objects with the lowest distance or highest similarity are combined into a cluster. In the next step, the object with the lowest distance (or highest similarity) to either of the first two is identified and studied. If it is closer to a fourth object than to either of the first two, the third and fourth objects become the second two-case cluster; otherwise, the third object is included in the first cluster. The process is repeated, adding cases to existing clusters, creating new clusters, or combining those that have emerged until each object has been examined and allocated to one cluster, or stands as one separate cluster by itself. It should be noted that at each step of this process, a different partition is formed that is nested in the partition generated in the following step. Usually, the researcher chooses the partition that turns out to be the most meaningful for a particular application.

Distance and similarity are key concepts in the context of CA. Most algorithms, particularly those yielding hierarchical partitions, start with a distance or similarity matrix. The cell entries of this matrix are distances or similarities between pairs of objects. There are many types of distances of which the most common is the *Euclidean distance.*

The Euclidean distance between any two objects is the square root of the sum of the squares of the differences between all the coordinates of the vectors that define each object. It can be used for variables measured at an interval scale. Euclidean distance is calculated as

$$d(x,y) = \left\{ \sum_i (x_i - y_i)^2 \right\}^{1/2}.$$

When two or more variables are used to calculate the distance, the variable with the larger magnitude

will dominate. To avoid this, it is common practice to first standardize all variables. The choice of a distance type is crucial for all hierarchical clustering algorithms and depends on the nature of the variables and the expected form of the clusters. For example, the Euclidean distance tends to yield spherical clusters.

Other commonly used distances include the following:

- *The Manhattan distance.* It is defined as the average distance across variables. In most cases, it yields results similar to the simple Euclidean distance. However, the effect of single large differences (outliers) is dampened (since they are not squared). It is computed as

$$d(x, y) = \sum_i |x_i - y_i|.$$

- *The Chebichev distance.* It may be appropriate when objects that differ in just one of the variables should be considered different. It is calculated as

$$d(x, y) = \max_i |x_i - y_i|.$$

- *The power distance.* It is used when it is important to increase or decrease the progressive weight that is assigned to variables on which the respective objects are very different. The power distance is controlled by two parameters r and p as shown by the following expression:

$$d(x, y) = \sum_i \left(|x_i - y_i|^p\right)^{1/r}$$

  where r and p are user-defined parameters. Parameter p controls the progressive weight that is placed on differences on individual variables, while parameter r controls the progressive weight that is placed on larger differences between objects. If r and p are equal to 2, then this distance is equal to the Euclidean distance.
- *The percent disagreement.* It may be used when the data consist of categorical variables. It is computed as

$$d(x, y) = \text{number of } x_i \neq y_i / i.$$

Linkage Rules

When clusters are composed of a single object, the distance between them can be calculated with any of the distances shown above. However, when clusters are formed by two or more objects, rules have to be defined to calculate those distances.

The distance between two clusters may be defined as the distance between the two closest objects in the two clusters. This linkage rule is known as *the nearest neighbor rule* and it will string objects together and tend to form chain-like clusters.

Other popular linkage rules are the pair-group average and the pair-group centroid. The first of these rules is defined as the average distance between all pairs of objects in the two different clusters. This method tends to form natural distinct clumps of objects. The pair-group centroid is the distance between the centroids or centers of gravity of the clusters.

The most frequently used nonhierarchical clustering technique is the *k-means algorithm*, which is inspired by the principles of analysis of variance. In fact, it may be thought of as an analysis of variance "in reverse." If the number of clusters is fixed as k, the algorithm will start with k random clusters and then move objects between them with the goals of minimizing variability within clusters and maximizing variability between clusters.

—Jorge Bacallao Gallestey

See also Analysis of Variance; Discriminant Analysis; Factor Analysis; Inferential and Descriptive Statistics

Further Readings

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis* (Quantitative Applications in the Social Sciences, Series No. 44). Beverly Hills, CA: Sage.

Corter, J. E. (1996). *Tree models of similarity and association* (Quantitative Applications in the Social Sciences, Series No. 112). Thousand Oaks, CA: Sage.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). London: Edward Arnold.

Kachigan, S. K. (1982). *Multivariate statistical analysis.* New York: Radius Press.

Kaufman, L., & Russeeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

Coefficient of Determination

The coefficient of determination, R^2, is a useful measure of the overall value of the predictor variable(s) in predicting the outcome variable in the linear regression setting. R^2 indicates the proportion of the overall sample variance of the outcome that is predicted or explained by the variation of the predictor

variable, or in the case of multiple linear regression, by the set of predictors.

For example, an R^2 of 0.35 indicates that 35% of the variation in the outcome has been explained just by predicting the outcome using the covariates included in the model. Thirty-five percent might be a very high portion of variation to predict in a field such as the social sciences; in other fields such as rocket science, one would expect the R^2 to be much closer to 100%. The theoretical minimum R^2 is 0; however, since linear regression is based on the best possible fit, R^2 will always be greater than zero, even when the predictor and outcome variables bear no relationship to one another.

Just as R^2 will virtually always be greater than zero, R^2 will also always increase when a new predictor variable is added to the model, even if the new predictor is not associated with the outcome. To combat this effect, the adjusted R^2 incorporates the same information as the usual R^2 but then also penalizes for the number of predictor variables included in the model. As a result, as new predictors are added to a multiple linear regression model, R^2 will always increase, but the adjusted R^2 will increase only if the increase in R^2 is greater than one would expect from chance alone. In such a model, the adjusted R^2 is the most realistic estimate of the proportion of the variation in Y that is predicted by the covariates included in the model.

When only one predictor is included in the model, the coefficient of determination is mathematically related to the Pearson's correlation coefficient, r. Just as one would expect, squaring the correlation coefficient results in the value of the coefficient of determination. The coefficient of determination can also be found with the following formula: $R^2 = MSS/TSS = (TSS - RSS)/TSS$, where MSS is the model sum of squares, TSS is the total sum of squares associated with the outcome variable, and RSS is the residual sum of squares. Note that R^2 is actually the fraction of the proportion of variability explained by the model out of the total variability in Y.

The coefficient of determination shows only association. As with linear regression, it is impossible to use R^2 to determine whether one variable causes the other. In addition, the coefficient of determination shows only the magnitude of the association, not whether that association is statistically significant.

In summary, the coefficient of determination provides an excellent one-number summary of how clinically relevant the predictor variable is in a given linear regression model; the adjusted R^2 provides the same one-number summary when more than one predictor is included in the model.

—Felicity Boyd Enders

See also Analysis of Variance; Pearson Correlation Coefficient; Regression

Further Readings

Bring, J. (1996). A geometric approach to compare variables in a regression model. *The American Statistician*, *50*(1), 57–62.

Srivastava, A. K., Srivastava, V. K., & Ullah, A. (1995). The coefficient of determination and its adjusted version in linear models. *Econometric Reviews*, *14*, 229–240.

Cohort Effects

Cohort effects are variations in factors such as health status or mortality that are attributed to the unique physical and social environment to which a cohort is exposed during its lifetime. A commonly used synonym is *generation effects.*

A cohort is defined as an aggregate of individuals within a specified population who experience the same event (e.g., birth) at the same time or within the same time interval, and who are observed over time. A cohort is usually identified by the event itself and by the time period during which the event occurs, and also, at least implicitly, by geographic location, that is, the physical environment. For example, the U.S. birth cohort of 2000 includes all persons born in the United States in the calendar year 2000. The time period that defines a cohort can be very short, or it can extend over several years; a decade is commonly used to identify a cohort.

The importance of taking cohort membership into consideration was first noted by demographers, and subsequently by other social and behavioral sciences, including epidemiologists. Cohort studies (and cohort analysis) are conducted to separate the effects due to cohort membership from those due to age or period. However, there is an inherent problem since cohort, age, and period are all related to time, and the interpretation of findings is subject to confounding. Since each of these time-related variables is dependent on the other two, attributing an observed

effect to any one of them usually requires knowledge of biological, historical, or behavioral factors that might produce age, period, or cohort effects. Multiple regression techniques can help disentangle age, period, and cohort effects. However, the question remains whether cohort membership has any effect apart from age and period, since cohorts are defined by the sharing of age and period.

A classic example of the attempt to separate age, period, and cohort effects is Wade Hampton Frost's examination of tuberculosis mortality rates in Massachusetts between 1880 and 1930, which showed that apparent changes in the age distribution of tuberculosis mortality could be better interpreted as a decline in mortality over cohorts, rather than as changes in age-specific mortality over time. In a graph of mortality rates, with age along one axis and year of death along the other, the cohort effect can be seen by tracing the diagonal axis of the graph, which plots the mortality of each cohort as it ages over time.

A cohort effect was discerned in the pattern of deafness in New South Wales, Australia; prevalence was especially high in 1911 among those who were 10 to 14 years old, in 1921 among 20- to 24-year-olds, and in 1931 among 30- to 34-year-olds. The most likely explanation for this prevalence pattern was the 1899 measles epidemic in New South Wales, which resulted in congenital deafness among a large proportion of those born to women exposed to measles during pregnancy.

—Judith Marie Bezy

See also Birth Cohort Analysis; Confounding; Frost, Wade Hampton; Interaction; Longitudinal Research Design

Further Readings

MacMahon, B., & Trichopoulos, D. (1996). *Epidemiology: Principles and methods* (2nd ed.). Boston: Little, Brown.

Mason, W. M., & Fienberg, S. E. (1985). *Cohort analysis in social research: Beyond the identification problem.* New York: Springer-Verlag.

Cohort Studies

See Study Design

Collinearity

Collinearity is a concern when two regression covariates are associated with one another. In essence, because the covariates are correlated, their prediction of the outcome is no longer independent. As a result, when both covariates are included in the same regression model, each one becomes less statistically significant because they are explaining some of the same variance in the dependent variable. Cues that collinearity may be a concern are (1) high correlation or association between two potential covariates, (2) a dramatic increase in the p value (i.e., reduction in the significance level) of one covariate when another covariate is included in the regression model, or (3) high variance inflation factors. The variance inflation factor for each covariate is $1/(1 - R^{2}*)$, where $R^{2}*$ is the coefficient of determination for the model excluding only that covariate. Variance inflation factors of 1 or 2 show essentially no collinearity; 5 represents moderate collinearity. Variance inflation factors greater than 10 suggest that collinearity is such a concern that one might consider removing one of the collinear covariates. Variance inflation factors of 20 and higher show extreme collinearity.

In the schematic on the left in Figure 1, X_1 and X_2 are nearly unassociated with one another, so the overlapping gray area is quite small. In the schematic on the right, the overlapping area is larger; in this gray area, the two covariates are battling to predict the same variance in the outcome, so that each prediction has lower statistical significance. If the outcome were income level, one might expect covariates such as education and previously winning the lottery to be similar to the schematic on the left. Alternatively, covariates such as education and parental income level would likely be more similar to the schematic on the right.

Multicollinearity describes a situation in which more than two covariates are associated, so that when all are included in the model, one observes a decrease in statistical significance (increased p values). Like the diagnosis for collinearity, one can assess multicollinearity using variance inflation factors with the same guide that values greater than 10 suggest a high degree of multicollinearity. Unlike the diagnosis for collinearity, however, one may not be able to predict multicollinearity before observing its effects on the multiple regression model, because

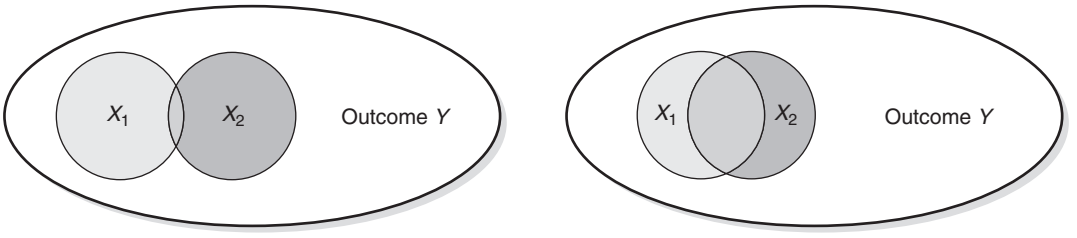


Figure 1 Collinearity Schematic

Note: Low collinearity is shown on the left, while moderate to high collinearity is represented on the right.

any two of the covariates may have only a low degree of correlation or association.

Sometimes the goal of a multiple regression model is to provide the best possible prediction of the outcome. In this situation, even with high variance inflation factors, one might choose to include several somewhat collinear covariates; the result will likely be that one or more fails to achieve statistical significance. However, care should be taken in this situation. A high degree of multicollinearity goes hand in hand with instability in the regression coefficients themselves. If the goal is simply to predict the outcome, then this instability need not be a major concern since the predicted value of Y would be unlikely to change much with a slightly different model. However, if one also cares about the regression equation used to make the prediction, multicollinearity can become a grave concern.

—Felicity Boyd Enders

See also Pearson Correlation Coefficient; *p* Value; Regression; Significance Testing

Further Readings

Sengupta, D., & Bhimasankaram, P. (1997). On the roles of observations in collinearity in the linear model. *Journal of the American Statistical Association*, *92*(439), 1024–1032.

Stine, R. A. (1995). Graphical interpretation of variance inflation factors. *The American Statistician*, *49*(1), 53–56.

Community-Based Participatory Research

Community-based participatory research (CBPR) in public health is a collaborative approach to research that equitably involves diverse partners, including community members, organizational representatives, and academic researchers, in all aspects of the research process. All partners contribute their expertise and share responsibility and ownership to enhance understanding of a topic of importance to the community, and to translate the knowledge gained into action to improve community health and well-being. Through research findings that can guide interventions and policy change, CBPR can contribute substantially to epidemiology's key functions of identifying and understanding the distribution and determinants of health and disease, and applying this knowledge to improve the public's health.

CBPR is an established approach to inquiry rooted in several research traditions across multiple disciplines, including participatory research and participatory action research. Common to these approaches is an emphasis on engaging community and academic partners in understanding and improving community well-being. CBPR is increasingly applied in epidemiological research and is well suited for investigating complex relationships between the social, economic, physical, and biological environments, and how they interact to influence health across multiple levels. Because CBPR is grounded in equitable collaboration between community members and academic researchers, CBPR is particularly relevant to the investigation of health disparities rooted in racial, economic, and social inequality.

As an approach to research rather than a particular research design or method, CBPR can be applied to a wide range of research efforts, including observational research, exposure assessment, risk assessment, intervention research and evaluation, and as a guide to policy development. CBPR draws on the full range of study designs, including cross-sectional, longitudinal,

experimental, and quasi-experimental, and may include both quantitative and qualitative methods of data collection and analysis.

CBPR can also form the foundation for comprehensive etiologic research and intervention research investigation of a particular health issue over time. For example, an ongoing community-academic partnership effort to examine and address exposure to environmental triggers of childhood asthma in Detroit, Michigan, as described by Parker et al. (2003), included prioritizing the health issue, successful competition for multiyear federal funding, a laboratory study of allergens and asthma, an epidemiological study of air quality and health indicators, and an intervention to reduce environmental triggers for childhood asthma at the household, neighborhood, and policy levels.

What distinguishes CBPR methodologically from other approaches to research is the engagement of the community as coinvestigator in determining what is being studied, how, by whom, and for what purpose. CBPR provides a framework for applying traditional and innovative research methods in ways that involve partners equitably, address power and cultural differences, maximize strengths and resources of partners, and build individual, organizational, and community capacity. Participatory approaches can enhance research quality and applicability and contribute to innovation in research methods. See Viswanathan et al. (2004) for a systematic review of existing evidence on the conduct and evaluation of CBPR, commissioned by the Agency for Healthcare Research and Quality.

Principles of Community-Based Participatory Research

CBPR entails a shared commitment to conducting research based on a set of principles and procedures developed by each specific partnership and intended to promote equitable engagement of all partners. The process of developing shared guidelines for practice forms the foundation of a particular CBPR partnership and guides the design, conduct, and application of the research. While no one strategy is applicable to all CBPR endeavors, there are core underlying principles derived from theory and practice that inform the conduct of CBPR. The following principles or key elements synthesized by Israel and colleagues together comprise a set of goals toward which research partnerships may strive within a CBPR orientation, with the potential for enhancing the quality and application of the research.

- *CBPR recognizes community as a unit of identity.* Understanding complex public health issues requires examining the interaction between people and their social, economic, and physical environments. CBPR processes contextualize investigation and action within specific communities, which may be defined geographically (e.g., neighborhood), sociopolitically (e.g., ethnic group), or based on some other aspect of shared identity. With an emphasis on engaging community members in identifying issues of local concern, CBPR can be an effective means for epidemiologists to conduct studies that matter to the health of the public.

- *CBPR builds on existing knowledge, strengths, and resources within the community.* By drawing on community assets, CBPR challenges a deficit model that emphasizes community problems and identifies strengths and resources that can enhance the design and implementation of research and facilitate its translation into interventions or policy change.

- *CBPR facilitates a collaborative, equitable partnership that engages community members and researchers in all phases of research.* This partnership involves an empowering and power-sharing process that addresses social inequalities, particularly those based on race, class, and gender. CBPR can link communities that have historically been marginalized or excluded from research with researchers committed to eliminating social inequalities and is particularly relevant to understanding and addressing health disparities. Engaging in equitable relationships recognizes the importance of local knowledge, helps build a more complex understanding of the phenomenon of interest, and shifts power from outside researchers toward the community as experts in their own experience of the determinants of health.

- *CBPR promotes colearning and capacity building among all partners.* CBPR brings together individuals and organizations with diverse knowledge, skills, and resources with the intent of reciprocal exchange of these capacities. For example, community members have knowledge and insights about dynamics that influence health in their communities of which outside researchers may not be aware. Researchers from outside the community can strengthen community capabilities as researchers and agents of change.

By partnering, community members and researchers can develop research that is relevant to the local community within historical and cultural contexts and that also addresses important public health questions; combine knowledge and perspectives to strengthen the quality of the data; pool their expertise to strengthen the effectiveness of interventions; and bring additional resources and employment opportunities for marginalized communities. This colearning process helps build the capacity of all partners to address public health concerns.

- *CBPR achieves a balance between knowledge generation and action for the mutual benefit of all partners.* CBPR aims to both generate knowledge and translate research findings into actions that benefit the community. CBPR may include development of interventions or policies that direct resources and influence determinants of health.

- *CBPR focuses on the local relevance of public health problems within a broader ecological framework that recognizes multiple determinants of health.* CBPR can strengthen epidemiology's goal of understanding the distribution and determinants of disease or health in populations and applying this knowledge to controlling health problems, particularly those associated with racial and socioeconomic disparities in health. CBPR is particularly relevant for examining interactions among determinants of health and identifying mechanisms for change across downstream, midstream, and upstream levels.

- *CBPR involves all partners in interpreting and disseminating results within the community and to wider audiences.* The partnership process enables collective interpretation of results, facilitates framing of findings in ways that are respectful of the community, and emphasizes communication of findings to multiple audiences, including community residents, scientists, funders, and policymakers. By engaging all partners in planning for dissemination and including both community and academic partners as coauthors and copresenters, CBPR recognizes community members for their ownership and role in knowledge production and provides expanded opportunities for dissemination.

CBPR is a long-term process that involves systems development within the partnership and between partner organizations, as well as a commitment to sustainability. CBPR relationships extend beyond a single research project and promote continued engagement in efforts to improve community well-being over time.

Recent edited volumes that detail principles and practice of CBPR and describe specific applications in public health include *Community-Based Participatory Research for Health* (Minkler & Wallerstein, 2003) and *Methods in Community-Based Participatory Research for Health* (Israel, Eng, Schulz, & Parker, 2005). In addition, several review articles (Israel, Schulz, Parker, & Becker, 1998; Viswanathan et al., 2004), as well as special journal issues, such as *Health Education and Behavior* (Volume 29, Issue 3, 2002) and *Environmental Health Perspectives* (Volume 110, Supplement 2, 2002), describe specific applications of CBPR for public health research and practice.

—Chris M. Coombe, Barbara A. Israel, Amy J. Schulz, Edith A. Parker, and Angela Reyes

See also Community Health; Health Disparities; Participatory Action Research

Further Readings

Israel, B. A., Eng, E., Schulz, A. J., & Parker, E. (Eds.). (2005). *Methods in community-based participatory research for health.* San Francisco: Jossey-Bass.

Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). Review of community-based research: Assessing partnership approaches to improve public health. *Annual Review of Public Health*, *19*, 173–202.

Lantz, P. M., Israel, B. A., Schulz, A. J., & Reyes, A. (2006). Community-based participatory research: Rationale and relevance for social epidemiology. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in social epidemiology* (pp. 239–266). San Francisco: Jossey-Bass.

Minkler, M., & Wallerstein, N. (2002). *Community-based participatory research for health.* San Francisco: Jossey-Bass.

Parker, E. A., Israel, B. A., Williams, M., Brakefield-Caidwell, W., Lewis, T. C., Robins, T., et al. (2003). Community action against asthma: Examining the partnership process of a community-based participatory research project. *Journal of General Internal Medicine*, *18*(7), 558–567.

Schulz, A. J., Krieger, J., & Galea, S. (2002). Addressing social determinants of health: Community-based participatory approaches to research and practice. *Health Education and Behavior*, *29*(6), 287–295.

Viswanathan, M., Ammerman, A., Eng, E., Gartlehner, G., Lohr, K. N., Griffith, D., et al. (2004, July). *Community-based participatory research: Assessing the evidence. Evidence Report/Technology Assessment No. 99*

(Prepared by RTI and University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). AHRQ Publication 04-E022-2. Rockville, MD: Agency for Healthcare Research and Quality.

Community Health

The term *community health* refers to the health status of a community and to the public and private activities undertaken to protect or improve the health of its members. The public and private activities are divided into three domains: health promotion, health protection, and health services. Each of these is discussed in greater detail below. *Health* can be defined as the physical, mental, emotional, and social resources for a productive and satisfying life. A *community* can be defined as a group of people who live under the same regulations, social norms, values, and organizational structure. Community members share a sense of identity and belonging, a common system of symbols, language, rituals, and ceremonies. They also share common needs and a commitment to meeting them. Communities can be based on geography, such as neighborhoods and towns, or they can be based on some other unifying factor, such as religion, ethnicity, or ancestry.

Population health differs from *community health* in that it describes the health of groups of people who do not represent a community. That is, they do not share a geographical identity, values, or any of the unifying factors mentioned above. Yet these populations as a group share similar health characteristics and concerns. The following are examples of such non-community-based populations that share common health characteristics and concerns: adolescents, adults 25 to 44 years of age, women 50 years of age or older, seniors living in public housing, prisoners, and blue-collar workers. Thus, a population can be based on age, gender, occupation, or any other characteristic. The health status of these groups and the public and private activities undertaken to protect and improve their health constitute population health.

Community and population health are affected by four factors: physical factors (e.g., community size and industrial development), social and cultural factors (e.g., politics and religion), individual behavior (e.g., getting immunized and willingness to recycle), and community organization (i.e., can a community come together to solve a problem).

History of Community Health Practice

Early History

Community health and the practice of community health have a long history. In all likelihood, the earliest community health practices went unrecorded. While there is archeological evidence of human concern about health dating as early as 25,000 BCE, the first inscribed health-related laws can be traced to Hammurabi, the king of Babylon, who in 1900 BCE issued a code of conduct that included laws pertaining to physicians and health practices.

During the years of the classical cultures (500 BCE to AD 500), the Greeks promoted men's physical strength and skill and made advances in community sanitation. The Romans built on the Greeks' engineering, practiced street cleaning and refuse removal, and built aqueducts that transported water from distant places. Although the Romans did little to advance medical thinking, they were the first to build hospitals.

The Medieval and Renaissance Periods

In the Middle Ages (AD 500–1500), when formal learning in Western Europe was largely restricted to monasteries, most believed that diseases and other health problems arose from spiritual causes and required spiritual solutions. Little progress was made in community health. Leprosy, plague, and other communicable diseases were epidemic. The plague epidemic of the 14th century, also known as the Black Death, was the worst of these, killing an estimated 25 million people in Europe alone.

Epidemics continued during the Renaissance period (AD 1500–1700), but there was a growing recognition of the relationship between the physical environment and diseases. More accurate descriptions of the courses of diseases led to better identification of specific diseases.

The 18th and 19th Centuries

The 18th century brought industrial growth to England and Western Europe. Generally, however, living conditions remained unhealthy, and many workplaces were unsafe. Toward the end of the century, in 1796, the English physician Edward Jenner successfully

demonstrated the process of vaccination for smallpox. Meanwhile, in the United States, in 1798, a young government established the Marine Hospital Service to meet the need for hospital services for merchant marines. The Marine Hospital Service eventually became the U.S. Public Health Service.

During the first half of the 19th century, there were few advances in community health practice in the United States. Unsanitary living conditions and epidemics were still concerns, but improvements in farming practices resulted in a better food supply and improved nutrition. Except for major epidemics, community health problems were not addressed by governmental agencies at any level. This began to change in 1850, when Lemuel Shattuck drew up a health report for the Commonwealth of Massachusetts. The Shattuck Report outlined the public health needs of the state and marks the beginning of the modern era of public health in the United States.

The second half of the 19th century was highlighted by the remarkable works of Louis Pasteur of France (the germ theory of disease) and Robert Koch of Germany (demonstrated that a particular microbe, and no other, causes a particular disease). Because of the works of Pasteur, Koch, and many others, the period from 1875 to 1900 has become known as the bacteriological period of public health. As monumental as these scientific achievements were, at the beginning of the 20th century, communicable diseases remained the leading causes of death in the world.

Health Resources Development in the United States

In the United States, the period 1900 to 1960 was known as the health resources development period. During this period, the number of medical and nursing schools increased significantly, new hospitals were built, and the number of health care professionals grew rapidly. This period also saw a rapid growth in the number of voluntary health agencies. The National Association for the Study and Prevention of Tuberculosis (now the American Lung Association) was founded in 1904 and the American Cancer Society in 1913. The federal government finally became involved in health and safety regulation when it passed the Pure Food and Drugs Act of 1906. Twenty-nine years later, the passage of the Social Security Act of 1935 signaled the federal government's first major involvement in social issues. World Wars I and II accelerated the rate of medical discoveries, including the development of new drugs (i.e., penicillin) and new medical procedures. In 1946, Congress passed the National Hospital Survey and Construction Act (Hill-Burton Act) to improve the distribution and enhance the quality of American hospitals.

Social Engineering and Health Promotion in the United States

The realization that many Americans were still not benefiting from the medical progress made during the first half of the 20th century led to a period of social engineering (1960–1973) during which the government tried to assure that more Americans would receive health care. Passage of amendments to the Social Security Act that established Medicare (payment of medical bills for the elderly and certain people with disabilities) and Medicaid (payment of medical bills for the poor) highlighted this period. The health promotion period (1974 to the present) began when it was recognized that the greatest potential for improving the health of communities and populations was not through health care but through improving lifestyle behaviors that are best addressed by health promotion and disease prevention programs.

In 1980, the U.S. Department of Health Education and Welfare released *Promoting Health/Preventing Disease: Objectives for the Nation*, its first comprehensive, 10-year health plan for the nation. This "blueprint for health" set forth health goals and objectives to improve the health of all Americans. Progress toward achieving the plan's objectives was measured, and new goals and objectives were developed and announced for each succeeding decade: *Healthy People 2000* and *Healthy People 2010.* Similar plans have now been developed by the governments of other countries, states, provinces, and even local communities.

Essential Disciplines

Those who practice community health must also have some level of expertise in three other disciplines: epidemiology, community organizing, and health education. John Last has defined *epidemiology* as the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control health problems. The

practice of epidemiology involves the collection of data about cases of disease or other health-related events and the calculation of rates and other forms of measurement. Herbert and Irene Rubin define *community organizing* as the bringing together of people to combat shared problems and increase their say about decisions that affect their lives. For example, communities may organize to help control violence in a neighborhood. *Health education* is the process of providing individuals, groups, or communities with learning experiences to help them make decisions that improve their health. A stress management class for church members is an example of health education. Health education is a basic component of health promotion programming that is discussed in more detail below.

Community and Population Health Through the Life Span

Most communities include people at all stages of their lives from the youngest members (infants) to the oldest members (elders). A closer look at each age segment of the community reveals that each of these groups has characteristic health problems. The most frequently used groupings are mothers, infants (ages < 1), and children (ages 1 to 14); adolescents and young adults (ages 15 to 24); adults (ages 25 to 64); and elders (65 years of age and older).

Maternal, infant, and child health encompasses the health of women of childbearing age from prepregnancy through pregnancy, labor, delivery, and the postpartum period, and the health of a child prior to birth through adolescence. Health statistics for this group reveal much about access to health care and the community's commitment to its most vulnerable members. High rates of maternal, infant, and childhood morbidity and mortality indicate inadequate access to health care, too many unplanned pregnancies, insufficient prenatal care, inadequate nutrition, substance abuse problems, and inadequate levels of immunization.

Community health programs that address these problems include those offered by state and local health departments and other agencies. Early intervention with educational programs and preventive medical services for women, infants, and children can lower morbidity and mortality rates, reduce the necessity of costly medical and/or social assistance, and optimize health in later years.

Family planning is defined as the process of determining and achieving a preferred number and spacing of children. A major concern in many communities is teenage pregnancies. Nearly 1 million U.S. teenagers become pregnant each year and most ($\sim 85\%$) of these pregnancies are unintended. Family planning services such as those offered by Planned Parenthood and other agencies can reduce the number of unwanted pregnancies and the number of low-birthweight (LBW) infants. These services include health education, risk assessment, and medical services that begin before the pregnancy and continue through birth. Prenatal care that begins before or very early in pregnancy reduces the chances of an LBW infant and the poor health outcomes that are associated with it. A controversial way of dealing with unintended or unwanted pregnancies is with abortion. Abortion has been legal in the United States since 1973 when the Supreme Court ruled in *Roe v. Wade* that women have a constitutionally protected right to have an abortion in the early stages of pregnancy. According to the Centers for Disease Control and Prevention, approximately 850,000 legal abortions are performed each year in the United States.

Infant and child health are the result of parent health behavior during pregnancy, prenatal care, and the care provided after birth. Critical health concerns for infants and children include immunizations; protection from injuries, both unintentional and intentional; and proper nutrition. One of the most successful programs in the latter regard is the Special Supplemental Food Program for Women, Infants, and Children, known as the WIC program. This program, sponsored by the U.S. Department of Agriculture, provides food, nutritional counseling, and access to health services for low-income women, infants, and children. The WIC program serves more than 8 million mothers and children per month and has been shown to save approximately 3 dollars for each dollar spent.

During the adolescent and young adult years, people complete their physical growth, marry and start families, begin a career, and enjoy increased freedom and decision making. They also adopt beliefs, attitudes, and behaviors about health and make lifestyle choices that will either promote or hamper their health in later years. For many, this is a time for risk-taking behavior (e.g., use and abuse of drugs, risky sexual behavior, and sedentary lifestyle) that can result in unintended injuries and even permanent disabilities. Communities that establish sensible guidelines for numbers of bars and liquor stores,

provide and promote the use of parks and recreation facilities, and plan neighborhoods with safety in mind can foster better health choices for all. In communities where interventions have been successful, these interventions have been comprehensive and community-wide in scope and sustained over time.

The adult population (ages 25 to 64) represents about half of the U.S. population and can be subdivided into those who are 25 to 44 and those 45 to 64 years of age. For the younger of these two subgroups, the leading cause of death is unintentional injuries, followed by cancer and heart disease. For the older group, the leading cause of death is cancer, followed by heart disease; these two causes account for nearly two thirds of all deaths. For most individuals, however, these years of life are the healthiest. Community health interventions for this population are aimed at improving the quality of life rather than extending life.

Members of the community who have reached their 65th birthday are referred to here as elders. This is the fastest-growing segment of the population. By 2030, one in five is expected to be 65 years and older. In 2011, the first baby boomers (those born between 1946 and 1964) will turn 65. From a community and population health perspective, greater attention will need to be placed on the increased demands for affordable housing, accessible transportation, personal care created by functional limitations, and all segments of health care, including adult day care and respite care. Signs of preparation for the growing population of elders are evident in many communities (i.e., senior centers and Meals-on-Wheels), but the increase in demand is expected to accelerate. As this happens, it will be important for communities not to lose focus on the health of its most vulnerable members, its mothers, infants, and children.

The Domains of Community and Population Health

The public and private activities of community and population health practice are divided into three domains: *health promotion*, *health protection*, and *health services.*

Health Promotion

Health promotion comprises the educational, political, environmental, regulatory, and organizational efforts designed to motivate and empower individuals, groups, and communities to take action to improve or protect their health. Health promotion programs encourage people and communities to adopt healthy behaviors and lifestyles and participate in community or political actions that influence health. Formal health promotion programs involve assessing community needs, setting goals and objectives, planning an intervention, implementing the plan, and evaluating the results. Health education is an important part of most health promotion interventions.

An area of community health that is usually amenable to health promotion programming efforts is recreation and fitness. A needs assessment might suggest that a community enhance its quality of life by building or improving sidewalks, bike paths, and hiking trails and by providing and promoting organized recreational programs that meet the social, creative, aesthetic, communicative, learning, and physical needs of its members. Such programs would contribute to the mental, social, and physical health of community members and provide healthy alternatives to the use of tobacco, alcohol, and other drugs as leisure pursuits. At a prescribed time following implementation of the program, its goals and objectives should be evaluated and compared with baseline data collected during the assessment stage. The plan or its implementation should then be adjusted to maximize health outcomes.

Health Protection

Health protection comprises all the efforts and actions undertaken to maintain the health and safety of a community, including the enactment, implementation, and enforcement of ordinances, laws, rules, and policies that protect community members from injuries and diseases. Such protective measures might include efforts to reduce the number and seriousness of injuries; control disease vectors, such as mosquitoes; assure that the air is clean, and that food and water are safe to consume; dispose of wastes properly; and assure that public transportation and housing and occupational and recreational environments are safe and healthy. The development and implementation of such protective measures often require citizen involvement, community organization, and effective government. Examples of successful public health protection policy include the Clean Air Act, implementation of safety belt laws, and the lowering of the blood alcohol concentration limit for impaired driving to 0.08% from 0.10%. Although these examples represent either federal laws or federal

mandates, it is important to recognize that the impetus for them may have begun with local, community-based action.

Health Services and Other Resources

The organization and deployment of the services and resources necessary to plan, implement, and evaluate community and population health strategies constitute the third domain in community health practice, health services. Today's communities are much more complex that those of the past, so high levels of organization are required to respond effectively to community health challenges. The types of agencies that might respond to a community health challenge include both governmental and nongovernmental agencies. These may be local, statewide, national, or even international agencies.

Governmental health agencies are funded primarily by tax dollars, are managed by government officials, and have specific responsibilities that are outlined by the governmental bodies that oversee them. Governmental health agencies include the World Health Organization, the U.S. Department of Health and Human Services, state and territorial health departments, and local health departments.

Nongovernmental health agencies are funded primarily by private donations or, in some cases, by membership dues. Voluntary health agencies such as the American Cancer Society and the American Heart Association are funded primarily by private donations. Professional organizations such as the American Public Health Association, the British Medical Association, the Canadian Nurses Association, and the Society for Public Health Education are funded by membership dues.

Philanthropic foundations are privately endowed organizations, many of which provide funds for worthwhile health programs or services. Examples of such foundations are the Robert Wood Johnson Foundation, the Henry J. Kaiser Family Foundation, the Bill and Melinda Gates Foundation, and the W. K. Kellogg Foundation.

Service, social, and religious organizations also contribute to community and population health by raising money and funding health-related programs. For example, the Lions Club works to help prevent blindness, and a countless number of religious organizations feed, clothe, and provide shelter for those in need.

Finally, there is another type of agency, the quasi-governmental agency. Quasi-governmental health organizations have some official responsibilities, but they also operate, in part, like voluntary health organizations. An example of this type of community health organization is the American Red Cross (ARC). Its official duties include acting as a representative of the U.S. government during natural disasters and serving as the liaison between members of the armed forces and their families during emergencies. In addition to these official responsibilities, the ARC engages in many nongovernmental services such as blood drives and safety services classes such as first-aid and water safety instructions. Most of its funding comes through private donations.

—James F. McKenzie and Robert R. Pinger

See also Community-Based Participatory Research; Governmental Role in Public Health; Health Communication; Public Health, History of

Further Readings

Green, L. W., & Ottoson, J. M. (1999). *Community and population health* (8th ed.). Boston: WCB/McGraw-Hill.

Israel, B. A., Checkoway, B., Schulz, A., & Zimmerman, M. (1994). Health education and community empowerment: Conceptualizing and measuring perceptions of individual, organizational, and community control. *Health Education Quarterly*, *21*(2), 149–170.

Joint Committee on Terminology. (2001). Report of the 2000 Joint Committee on Health Education and Promotion Terminology. *American Journal of Health Education 32*(2), 89–103.

Last, J. M. (Ed.). (2001). *A dictionary of epidemiology* (4th ed.). New York: Oxford University Press.

Lee, P. R., & Estes, C. L. (2003). *The nation's health* (7th ed.). Sudbury, MA: Jones & Bartlett.

McKenzie, J. F., Neiger, B. L., & Smeltzer, J. L. (2005). *Planning, implementing, and evaluating health promotion programs: A primer* (4th ed.). San Francisco: Benjamin Cummings.

McKenzie, J. F., Pinger, R. R., & Kotecki, J. E. (2005). *An introduction to community health* (5th ed.). Sudbury, MA: Jones & Bartlett.

Minkler, M. (Ed.). (1997). *Community organizing and community building for health.* New Brunswick, NJ: Rutgers.

Pickett, G., & Hanlon, J. J. (1990). *Public health: Administration and practice* (9th ed.). St. Louis, MO: Times Mirror/Mosby.

Rosen, G. (1993). *A history of public health.* Baltimore: Johns Hopkins.

Community Trial

Community trials, also called community intervention studies, are (mostly preventive) experimental studies with whole communities (such as cities or states) as experimental units; that is, interventions are assigned to all members in each of a number of communities. These are to be distinguished from field trials where interventions are assigned to healthy individuals in a community and from clinical trials in which interventions are assigned to patients in a clinical setting. Except for the experimental unit, the conduct of controlled community trials follow the same procedures as controlled clinical trials, including the requirement of informed consent of the communities meeting the eligibility criteria (e.g., consent being given by city mayors or state governors), randomization to treatment and control groups, and follow-up and measurement of endpoints. In contrast to clinical trials, blinding and double blinding are not generally used in community trials. Community trials are needed to evaluate directly the potential efficacy of large-scale public health interventions, and they are indispensable for the evaluation of interventions that cannot be allocated to individuals as experimental units. For example, in studying the effects on dental caries of adding fluoride to the water supply (see Example 2 below), Ast and colleagues found it impossible to add fluoride to drinking water for selected individuals; instead, a controlled community trial was conducted in which whole towns were allocated to receive fluoride in their water or not. Just as with other randomized controlled clinical trials, a randomized controlled community trial requires a sufficient number of communities to involve in the experiment for the randomization to achieve its purpose of reducing confounding bias. In reality, considerably fewer study units are capable of being randomized in community trials than in clinical trials, simply because the study units for the former are considerably larger than those for the latter. For this reason, community trials are often quasi-experimental studies rather than experimental studies and so are considered to have lower validity than clinical trials. However, community trials still have higher validity than all observational analytic and descriptive studies. After enough information has been accumulated from descriptive or observational studies about the risk factors and their potential for modification, community trials may then be conducted to assess the benefit of new public health programs. Below are two real examples of community trials, only the first example being a bona fide community trial.

Example 1: The Indonesian Vitamin A Study

This was a randomized controlled community trial conducted to determine if vitamin A supplementation is effective in reducing childhood mortality in Indonesia. The experimental units were villages and the intervention (vitamin A supplements) was given to whole villages. The study was for 1 year during which time 200,000 IU vitamin A were given twice to children aged 12 to 71 months in 229 randomly allocated treatment villages, while children in 221 control villages were not given vitamin A until after the study. Mortality among children in the control villages was 49% higher than that in the villages given vitamin A (mortality risk ratio $RR = .0073/.0049 = 1.49$, $p < .05$). Many more vitamin A randomized controlled community trials have been carried out worldwide in addition to the Indonesian study. Most of them yielded significant mortality risk ratios, indicating that supplements given to vitamin A–deficient populations would increase survival.

Example 2: The Newburgh-Kingston Caries Fluoride Study

This was a controlled community trial conducted to determine if increasing fluoride levels in drinking water would reduce children's dental caries. Two experimental units, the towns of Newburgh and Kingston in New York State, were assigned to a treatment (fluoridation) arm and a control (unfluoridation) arm, respectively. Starting from 1945 fluoride was added to Newburgh's water supply but not Kingston's. After 10 years of fluoridation, the prevalence rates of decayed, missing, or filled teeth (DMF) for fluoridated Newburgh children from age 6 to 16 years were found to be from 57.9% to 40.9% (the percentage decreases as age increases) lower than those for the corresponding unfluoridated children of Kingston (the DMF rates were similar between the two towns in 1945). However, with only two experimental units, randomization cannot achieve the intended purpose of reducing confounding bias. Hence, this study is only

a quasi-experimental study. Appropriate statistical adjustment of potential confounding variables would have to be made at the analysis stage before valid conclusions can be reached.

—*John J. Hsieh*

See also Clinical Trials; Descriptive and Analytic Epidemiology; Randomization; Study Design

Further Readings

Ast, D. B., Finn, S. B., & McCaffrey, I. (1950). The Newburgh-Kingston caries fluorine study. *American Journal of Public Health*, *40*, 716–724.

Ast, D. B., & Schlesinger, E. R. (1956). The conclusion of a ten-year study of water fluoridation. *American Journal of Public Health*, *46*, 265–271.

Beaton, G. H., Martorell, R., Aronson, K. J., Edmonston, B., McCabe, G., Ross, A. C., et al. (1993). *Effectiveness of vitamin A supplementation in the control of young child morbidity and mortality in developing countries.* New York: United Nations.

Sommer, A., Tarwotjo, I., Djunaedi, E., West, K. P., Jr., Loeden, A. A., Tilden, R., et al. (1986). Impact of vitamin A supplementation on childhood mortality. A randomized controlled community trial. *Lancet*, *327*, 1169–1173.

Comorbidity

In medicine and in psychiatry, comorbidity is defined as a preexisting concomitant but unrelated disease or diseases, in addition to a primary disease, disorder, initial diagnosis, or index condition. The term *comorbidity* is also used to describe the effect of all other disorders or diseases an individual patient might have other than the primary diagnosis or disease of interest. Results from the first National Comorbidity Survey, released in 1994, revealed that 79% of all seriously ill people (inclusive of all diseases) were comorbid. Comorbidity has serious implications for the diagnosis, treatment, rehabilitation, and outcome of affected individuals. Comorbidity may also affect the ability of affected individuals to function and may be used as a prognostic indicator for length of hospital stay, cost, and mortality.

According to the *Diagnostic and Statistical Manual* (*DSM*) published by the American Psychiatric Association, anxiety and major depressive disorders commonly occur together or are common comorbid disorders. Such comorbidity is found among about half of all the individuals with these disorders. Comorbidity is also common among substance users, both physiologically and psychologically (e.g., substance use/misuse and bipolar disorder). The presence of mental disorders associated with substance use and dependence—the dually diagnosed—among those attending substance use treatment services has been reported to be between 30% and 90%. A survey in 1994 found that 65% of those attending mental health services reported alcohol use disorders. Moreover, alcohol use–related disorders are also common among persons diagnosed with schizophrenia.

There are currently no standardized means of quantifying or classifying prognostic comorbidity. Many tests attempt to standardize the "weight" or predictive value of specific complications or comorbid conditions, based on the presence of secondary or tertiary diagnoses. The Charlson Co-Morbidity Index attempts to consolidate each individual comorbid condition into a single, adjusted variable that measures or predicts the 1-year mortality or other outcomes for a patient who presents with a range of comorbid conditions. The Charlson Co-Morbidity Index has demonstrated excellent predictive validation and contains 19 categories of comorbid conditions, primarily defined using the *International Statistical Classification of Diseases and Related Health Problems*, *Version 9, Clinical Modification* (ICD-9-CM) diagnoses codes. The comorbidity score reflects the cumulative increase in likelihood of 1-year mortality due to the severity of the effect of comorbidities; the higher the score, the more severe the burden of comorbidity.

Social scientists, health care scholars, and policymakers advised caution in the use and reliance on the Charlson Index because the ICD-9-CM codes used in the indexes or composite variables often lead to difficulties in distinguishing between complications and comorbidities. A complication is usually defined as a medical condition that is acquired during a hospital stay and could have been prevented. If these disease categories cannot be reliably differentiated, it is possible that the burden of comorbid conditions might be overestimated. Those urging caution in using the Charlson Index coding schema have also noted that the comorbidity and complication codes often fall in the same disease category resulting in a lack of distinction between principal diagnoses present at admission versus those that developed or occurred during the hospital stay.

Roos augmented the Charlson Index by developing a comorbidity algorithm that does not use individual diagnoses directly in calculating the comorbidity score, thereby limiting the effect of complications in risk adjustment. In addition, Quan modified and updated the Charlson Index ICD-9-CM coding indices to ICD-10, thus reducing the inability to distinguish between related comorbidities and complications.

—Kevin Robinson

See also Anxiety Disorders; Chronic Disease Epidemiology; Drug Abuse and Dependence, Epidemiology of; Psychiatric Epidemiology; Schizophrenia

Further Readings

Charlson, M. E., Pompei, P., Ales, K. L., & McKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, *40*(5), 373–383.

de Groot, V., Beckerman, H., Lankhorst, G., & Bouter, L. (2003). How to measure comorbidity: A critical review of available methods. *Journal of Clinical Epidemiology*, *56*, 221–229.

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J. C., et al. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care*, *43*(11), 1130–1139.

Roos, L. L., Stranc, L., James, R. C., & Li, J. (1997). Complications, comorbidities and mortality: Improving classification and prediction. *Health Services Research*, *32*(2), 229–238.

Competencies in Applied Epidemiology for Public Health Agencies

The Centers for Disease Control and Prevention (CDC) and the Council of State and Territorial Epidemiologists (CSTE) developed the *Competencies for Applied Epidemiologists in Governmental Public Health Agencies* (applied epidemiology competencies or AECs for short) to improve the practice of epidemiology within the public health system. The document defines the discipline of applied epidemiology and describes expected competencies for different levels of practicing epidemiologists working in government public health agencies. This document describes the rationale for developing the AECs and provides an overview of the participants, the development process, the AECs, their target audience, and the intended uses by those audiences.

CDC, the lead federal governmental agency on this project, is part of the U.S. Department of Health and Human Services and has responsibility for prevention and control of disease and other health-related conditions. As the professional organization representing epidemiologists at the state and recently at the large-city and county levels, CSTE is a key partner for CDC, particularly regarding applied epidemiology and workforce development. In January 2004, CDC and CSTE hosted a workforce summit to address concerns affecting public health epidemiologists. Leaders in applied epidemiology were invited to discuss key workforce issues. Participants strongly supported the need to establish core competencies for applied epidemiologists. Subsequent to the summit, CDC and CSTE identified competency development as a priority.

The reasons for pursuing this competency-development process were myriad. Multiple studies have demonstrated a substantial shortage of epidemiologists needed by local and state public health agencies. In addition, epidemiologists practicing in public health agencies often do not have sufficient training to accomplish their responsibilities. Promotion and retention of trained epidemiologists is often hindered by a lack of clear career ladders for epidemiologists. Previous, smaller-scale efforts to define the field have been independent and uncoordinated. The only national efforts, spearheaded by organizations such as the American College of Epidemiology (ACE) and the Association of Schools of Public Health (ASPH), focused on academic or doctoral epidemiologists. The ACE and ASPH efforts underemphasize critical applied epidemiology competencies in the areas of public health surveillance and field or outbreak investigation.

Competencies are observable and are identified in action-oriented statements that delineate the essential knowledge, skills, and abilities required for the performance of work responsibilities. Defining competencies for applied epidemiologists provides

- a roadmap for training the existing workforce;
- guidelines for academia to use in training the future workforce;
- definition of the skills needed for hiring epidemiologists;
- a basis for evaluating, rewarding, and promoting epidemiologists;

- improved ability to define the discipline; and
- a tool useful for any potential certification process.

In October 2004, CDC and CSTE convened an expert panel to define AECs for local, state, and federal public health epidemiologists. This panel comprised representatives from state and local health agencies, schools of public health, and private industry and from throughout CDC. To facilitate rigorous integration of both epidemiologic and workforce development perspectives in the process, two cochairs were recruited: Guthrie Birkhead, MD, MPH, New York State Department of Health, Albany, New York, a highly acclaimed leader in applied epidemiology; and Kathleen Miner, PhD, MPH, CHES, Rollins School of Public Health, Emory University, Atlanta, Georgia, also an equally acclaimed leader in workforce development. Additional academic epidemiologists and public health workforce development specialists also played key roles as reviewers on this panel. Other specific organizational partners included the ASPH, the Association of State and Territorial Health Officials, the American Public Health Association, and the National Association of County and City Health Officials.

The panel used a structured process to define the AECs during multiple face-to-face meetings and teleconferences that were interspersed with input from surveys of other practicing and academic epidemiologists. Jac Davies, MS, MPH, CSTE consultant, captured and synthesized all the comments for expert panel discussion and prepared the final documents, with oversight from the leadership group that consisted of the cochairs and the CDC convener (Denise Koo, MD, MPH). CSTE employees Jennifer Lemmings, MPH, LaKesha Robinson, MPH, and Executive Director Patrick McConnon, MPH, provided critical staffing for the effort, organizing the meetings and conference calls, coordinating the surveys and collating their results, and posting preliminary and final versions of the competencies on the CSTE Web site.

The AECs were built on the framework of the Core Competencies for Public Health Professionals—a product of the Council on Linkages (COL) Between Academia and Public Health Practice—and thus are consistent with the larger field of public health practice. They include not only epidemiologic or analytic competencies and competencies in other basic public health sciences but also competencies in the other COL domains of policy development and program planning, communication, cultural competency, community dimensions of public health practice, financial planning and management, and leadership and systems thinking. The AECs resulted from 2 years of highly collaborative work by the expert panel. Epidemiologists at all levels of public health practice from throughout the country and from academia provided substantial input into the AECs.

The document defines competencies for four tiers of practicing epidemiologists, categorized on the basis of level of responsibility, experience, and education: entry-level or basic, midlevel, supervisory, and senior scientist/researcher. The expert panel intended that all persons practicing applied epidemiology gain minimal competency in all the defined skill domains within the tier that most closely matches their level of responsibility. However, it is not expected that every applied epidemiologist will be equally competent in all areas. Different public health programs might emphasize different competency areas, and a government agency's responsibilities, needs, and resources might require epidemiologists in individual positions to focus on particular competencies.

The target audience and intended uses of the AECs include the following:

- practitioners, using the AECs for assessing current skills, creating career development plans, and planning specific training to meet educational needs;
- employers, using AECs for creating career ladders for employees, developing position descriptions and job qualifications, developing training plans for employees, and assessing epidemiologic capacity of the organization; and
- educators, using AECs for designing education programs that meet the needs of public health agencies and incorporating critical elements of epidemiologic practice into existing coursework.

It is expected that the AECs will be used as the basis for instructional competencies for training government epidemiologists and as the framework for developing position descriptions, work expectations, and job announcements for epidemiologists practicing in public health agencies. CDC plans to use the competencies as the basis for its Epidemic Intelligence Service and other epidemiology training activities. These competencies should also provide impetus for additional partnerships between academia and public health practice and might also prove useful for the training of other health professionals in quantitative or

population health skills. After public health agencies have used them for a period of time, CDC and CSTE will evaluate their utility and effectiveness as part of an ongoing process to update and improve them.

—Denise Koo

See also Applied Epidemiology; Centers for Disease Control and Prevention; Outbreak Investigation; Public Health Surveillance

Further Readings

Association of Schools of Public Health. (2002). *ACE/ASPH Workshop on doctoral education in epidemiology*, Baltimore, December 9–11, 2002. Retrieved January 3, 2007, from http://www.asph.org/document.cfm?page = 798.

Council of State and Territorial Epidemiologists. (2004). *2004 national assessment of epidemiologic capacity: Findings and recommendations.* Atlanta, GA: Author. Retrieved December 15, 2006, from http://www.cste.org//Assessment/ECA/pdffiles/ECAfinal05.pdf.

Lee, N. L., & Samet, J. M. (2003). ACE forum report: The making of an epidemiologist: Necessary components for doctoral education and training. *Annals of Epidemiology*, *13*, 552–556.

Miner, K. R., Childres, W. K., Alperin, M., Cioffi, J., & Hunt, N. (2005). The MACH Model: From competencies to instruction and performance of the public health workforce. *Public Health Reports* (Suppl. 1), 9–15.

Public Health Foundation. (2001, April 11). *Council on linkages between academia and public health practice*. Retrieved December 15, 2006, from http://www.phf.org/competencies.htm.

Web Sites

Centers for Disease Control and Prevention (CDC) and the Council of State and Territorial Epidemiologists (CSTE) Web site, which contains more information and the complete competency set: http://www.cste.org/competencies.asp. The competency set is also available on the CDC Internet site: http://www.cdc.gov/od/owcd/cdd/aec/.

Complementary and Alternative Medicine

Complementary and alternative medicine (CAM) is a group of diverse medical and health care systems, practices, and products that are not presently considered a part of conventional medicine. There are three general types of CAM: (1) Western alternative medical therapies, (2) traditional medical systems, and (3) complementary approaches. Researchers estimate that 68% of U.S. adults will use at least one form of CAM in their lifetime. Research and policy initiatives in CAM are growing in response to the widespread use. This entry discusses definitions and utilization of CAM, public policy with regard to CAM, and the relationship between CAM and public health. It also examines challenges for research on CAM; concerns related to reimbursement, credentialing, and education; legal and ethical issues; and the need for an integrative approach to CAM.

Defining CAM

Although the terms *complementary medicine* and *alternative medicine* are often used interchangeably, they have unique meanings. Alternative medicine suggests methods used in place of conventional Western medicine, while complementary medicine implies methods used in combination with conventional health care. As the field of unconventional medicine has grown, the combined term (*complementary and alternative medicine* or CAM) has become the acceptable nomenclature to depict the field. The National Institutes of Health Center for Complementary and Alternative Medicine (NCCAM) defines CAM as a group of diverse medical and health care systems, practices, and products that are not presently considered a part of conventional medicine. It includes medical or preventive practices that are not routinely taught in medical schools nor underwritten by third-party payers. These definitions, however, are exclusionary. More inclusive definitions suggest that CAM is a heterogeneous set of health systems and practices from cultures around the world that share characteristics differentiating them from conventional medicine. Most share a holistic approach that emphasizes wellness and health maintenance; acknowledges and integrates body, mind, and energy/spirit as the core and coequal elements of the human system; appreciates the role of environmental factors in health; and affirms the centrality of personal attitudes and choices in the creation of health and illness. According to the Cochrane Collaboration, an international nonprofit group that maintains a database of health care information, CAM is defined as a broad domain of healing resources that encompasses all health systems,

modalities, and practices and their accompanying theories and beliefs, other than those intrinsic to the dominant health system of a particular society or culture in a given historical period. For example, Western medicine (also referred to as allopathic or conventional medicine), which is practiced by medical doctors (MDs) and is dominant in the United States, understands and treats illness from a biologic perspective. Acupuncture, which is neither taught in medical school nor has a recognized biologic basis for its effects, is therefore considered CAM.

CAM can be classified into three broad types: (1) Western alternative medical therapies, for example, homeopathy, naturopathy, herbal medicine, aromatherapy; (2) traditional medical systems, for example, Chinese medicine, Indian Ayurveda, Tibetan medicine, Native American healing, Latin American curanderismo; and (3) complementary approaches, for example, mind-body interventions, psychoneuroimmunology (the study of the relationship between social and psychological factors and biologic processes), energetic healing. Western alternative medical therapies are relatively newer systems of health and healing developed in Europe or America. Some stem from or are related to conventional Western medicine, while others are unique. Traditional medical systems stem from cultural perspectives and form a complete approach to health care. Complementary approaches rely on the mind-body connection or energy as the primary mode of cure and may be used in conjunction with biomedical approaches.

Use of CAM

David Eisenberg and his colleagues conducted random-digit-dial cross-sectional studies in 1990 and 1997 that assessed the use of CAM in the United States. They found that the proportion of U.S. adults using CAM practices increased from 34% in 1990 to 42% in 1997. A trend analysis of these data suggests that 68% of U.S. adults will use at least one form of CAM in their lifetime. The Eisenberg studies also found that the most commonly used forms of CAM are relaxation therapies (16.5%), herbal medicine (12.1%), massage (11.1%), and chiropractic (11%); other studies added lifestyle/diet and exercise/movement to the list.

CAM is used most frequently to treat chronic conditions such as musculoskeletal problems, anxiety, depression, and headaches that have not been successfully treated with conventional medicine. In general, individuals are more likely to use CAM if they are more educated, have experienced a transformational event that changed their life, are in poorer overall health, and believe in a holistic philosophy. The appeal of CAM may be due to its perception of being natural and less artificial. In addition, its connection to cosmic forces (vitalism) and spiritual roots may drive its use.

In the Eisenberg studies, only about 40% of respondents informed their primary health care provider (PCP) about their use of CAM. Reasons for the lack of disclosure include the belief that it is not important or relevant to PCPs to know about their use; the belief that PCPs would not understand; the belief that PCPs would disapprove of CAM use; and the fact that PCPs do not ask about CAM use. A review by Astin and colleagues found that physicians refer clients for CAM when (1) the patient is not responding to conventional treatment, (2) it is requested by the patient, and (3) the physician believes in its effectiveness and safety.

Public Policy and CAM

In 1993, the National Institutes of Health established the Office of Alternative Medicine with a budget of $20 million to fund projects that explored the impact of CAM approaches on a number of health conditions. Since its inception, the Office has become a Center (1999), allowing expanded support of research. The NCCAM continues to grow, with a current budget in excess of $110 million.

A number of national and international activities have been conducted that explore and respond to CAM use and safety. In 2002, the World Health Organization published its *Global Statement on Traditional or Complementary and Alternative Health* (T/CAM). This document reports on the worldwide use of T/CAM and delineates steps to increase the quality, efficacy, and accessibility of these approaches.

Also in 2002, the White House Commission on CAM published a policy statement that provides recommendations on (1) the coordination of research, (2) education and training of practitioners, (3) information dissemination to health care providers, and (4) strategies for increasing access to CAM. A set of 10 principles (including wholeness orientation in health care delivery, evidence of safety and efficacy, emphasis on health promotion) guided the development of the recommendations.

Finally, the Institute of Medicine released its report on complementary and alternative medicine in the United States in 2005. The report explains the use, users, and types of CAM in the United States. The report provides several recommendations regarding research, training, and translation and dissemination of accurate information. The core recommendation is that all health-related treatments, conventional or CAM, should use the same standards and principles to determine treatment effectiveness.

CAM and Public Health

Traditionally, the public health sector has not embraced CAM. In part, this may be due to the fact that CAM is "complementary" and "alternative" to conventional medicine and that its relationship to public health has been unclear. The field of public health emphasizes the health of the entire community through an emphasis on prevention and by linking interventions to multiple social and environmental determinants of disease. This approach includes a focus on disease prevention and health maintenance, a holistic focus on the person in the context of the natural and human-created environment, and attention to personal choices and health behaviors. Public health is therefore grounded in principles consonant with those of the CAM practices. It is incumbent on the public health community to recognize the similarities between public health and CAM and to increase its cultural competency through increased understanding and acceptance of the diverse beliefs and practices of community members.

Challenges for Research on CAM

The quantity and quality of scientific research on CAM have increased, particularly since the advent of funding through the NCCAM. The PubMed search engine contains 420,000 references to CAM-related articles in medical and scientific journals. The NCCAM currently funds several Centers of Excellence for Research on CAM in the areas of acupuncture, antioxidants, energy medicine, herbal therapy, mindfulness-based stress reduction, and traditional Chinese medicine; six Centers for Dietary Supplements Research; and nine Developmental Centers for Research on CAM that support developmental research on CAM through collaborations between CAM schools and conventional biomedical research institutions. Current funding priorities of the NCCAM include research on mechanisms of action, active ingredients, pharmacology, bioavailability, and optimal dosing, safety, and efficacy. Areas of special interest include anxiety and depression, cardiovascular diseases, ethnomedicine, immune modulation/enhancement, inflammatory bowel disease and irritable bowel syndrome, insomnia, liver, obesity/metabolic syndrome, and respiratory diseases.

There are several barriers to the use of conventional scientific approaches for evaluating CAM therapies. Since Western culture and mainstream biomedical practice rely on principles of scientific evidence, CAM practices are currently judged within the scientific community on the basis of evidence from controlled research studies. The "gold standard" for medical research is the double-blind randomized controlled trial, in which the treatment and a control (placebo) are administered randomly to patients; neither the patient nor the health care provider knows who is getting the real treatment. Since CAM refers to a large and heterogeneous grouping of health interventions, the nature of many modalities in CAM can make comparison to conventional practice difficult. There is diversity even within some therapies (e.g., acupuncture, which may use either a 5-element or an 8-principle model); others are designed to be tailored to an individual patient on the basis of nonmedical considerations, making them difficult to standardize (e.g., homeopathy). A common research concern is distinguishing the effects of treatment from the placebo effect; because some CAM interventions are intended to stimulate the body's self-healing potential, this may become a meaningless distinction in CAM. Identifying appropriate controls can be challenging, and the design of double-blind trials is often impossible when practitioner knowledge is required to administer a CAM intervention appropriately.

The Cochrane Collaboration has also found evidence of publication bias resulting in the exclusion of research on some CAM topics from scientific publications, as well as a pattern of some CAM topics being published exclusively in languages other than English. The Collaboration is standardizing the assessment of research quality, which will provide more comparability of results across biomedical and complementary health research.

Reimbursement for CAM Care

Pelletier and associates estimate that total expenditures for CAM in 1997 were $27 billion, of which $21.2

billion was paid to practitioners for their professional services; this was a 45% increase over what was reported in 1990 data. Most of those expenses are paid by consumers out of pocket. Insurers decide whether to cover CAM treatment based on evidence regarding the clinical efficacy, safety, and cost-effectiveness of the treatment; consumer demand in their service area; state mandates; and standards of practice and practitioner licensure. Reimbursement rates depend on the local market, procedure code, and practitioner education and licensure; in most instances, a procedure is covered only if it is determined to be medically necessary (rather than important for general wellness).

As of 1999, Medicaid coverage is provided for CAM treatment in 36 states; chiropractic care is the most common treatment covered. States paying for services other than chiropractic are concentrated in the Northeast, Mid-Atlantic, and Pacific regions of the United States. In a few states, naturopaths or chiropractors are reimbursed as primary care providers under Medicaid.

Credentialing and Education

Licensure of both conventional and CAM health care practitioners is under the purview of each state legislature, and state policies on professional licensure and scope of practice vary. Chiropractic is licensed in every state, and acupuncture in 42 states; other practices (such as naturopathy and massage) are licensed in some states, whereas still others are rarely licensed (homeopathy is licensed in only 3 states, and practice is restricted to licensed medical doctors). Unlicensed CAM practitioners are liable to prosecution for the unlicensed practice of medicine. Nonlicensing professional credentialing in CAM modalities is provided by professional associations and educational institutions, and so credentialing is not widely standardized or legally recognized.

A 1997 survey by Wetzel and colleagues indicated that 64% of medical schools offered elective courses in CAM or included some CAM information in required courses. The focus and content of the training received by medical practitioners in CAM vary widely and are not universally included in the core curriculum.

Ethical and Legal Issues

Ethical and legal issues in CAM include the practice of CAM modalities by untrained providers (either conventional or CAM providers), client safety, and the interface between CAM and conventional medical practice. Ethical practice requires training and some form of licensure or credentialing. If practitioners do not have a working knowledge and experience in the use of CAM modalities, they cannot obtain informed consent from patients, and therefore client safety cannot be guaranteed. A large proportion of individuals who use CAM do not inform their biomedical practitioners; therefore, physicians have a responsibility to support and protect their patients by inquiring about their use of CAM practices, educating patients about the questions to ask of CAM practitioners, and providing information about known interactions among therapeutic interventions. Although referral to licensed and accredited CAM practitioners is not likely to create liability for malpractice claims, biomedical physicians using CAM interventions themselves may incur legal or insurance repercussions for deviation from standard practice. All CAM providers are, however, ethically bound to practice within their scope of training and to obtain informed consent for treatments.

Toward an Integrative Approach

As the U.S. population continues to change, Americans become increasingly dissatisfied with conventional health care alone, and the cost of health care continues to skyrocket, a model that integrates CAM, and conventional medicine should be considered. An integrative model of health care suggests an evolving population-based approach, with conventional medicine as one approach to health care. Integration is occurring at a number of levels and to various degrees. The consumer level is driving the demand for CAM, followed by practitioners who are responding to the demand. Health care institutions are responding to practitioners' use of CAM, although to a lesser degree. For example, some integrative medicine clinics are being implemented and evaluated, while a significant number of schools of medicine are beginning to offer courses and content in CAM. Regulatory bodies are only beginning to explore the role of CAM in credentialing. Finally, there is little or no integration at the policy level.

Although an integrative approach may be ideal, there continue to be barriers, including lack of research and information about the efficacy and cost-effectiveness of CAM treatments. A truly integrated

system would be holistic in its treatment approach, blend conventional health and CAM into a seamless continuum, employ a collaborative team, and result in effective and cost-efficient care.

—Darcell P. Scharff and Margret O'Neall

See also Cultural Sensitivity; Governmental Role in Public Health; Publication Bias

Further Readings

Astin, J. A. (1998). Why patients use alternative medicine: Results of a national study. *Journal of the American Medical Association*, *279*, 1548–1553.

Astin, J. A., Marie, A., Pelletier, K. R., Hansen, E., & Haskell, W. L. (1998). A review of the incorporation of complementary and alternative medicine by mainstream physicians. *Archives of Internal Medicine*, *158*, 2303–2310.

Eisenberg, D. M., Cohen, M. H., Hrbek, A., Grayzel, J., Van Rompay, M. I., & Cooper, R. A. (2002). Credentialing complementary and alternative medical providers. *Annals of Internal Medicine*, *137*, 965–983.

Eisenberg, D. M., Davis, R. B., Ettner, S. L., Appel, S. M., Wilkey, S. A., Van Rompay, M. I., et al. (1998). Trends in alternative medicine use in the United States, 1990–1997: Results of a follow-up national survey. *Journal of the American Medical Association*, *280*, 1569–1575.

Eisenberg, D. M., Kessler, R. C., Foster, C., Norlock, F. E., Calkins, D. R., & Delbanco, T. L. (1993). Unconventional medicine in the United States: Prevalence, costs, and patterns of use. *New England Journal of Medicine*, *328*, 246–252.

Eisenberg, D. M., Kessler, R. C., Van Rompay, M. I., Kaptchuk, T. J., Wilkey, S. A., Appel, S., et al. (2001). Perceptions about complementary therapies relative to conventional therapies among adults who use both: Results from a national survey. *Annals of Internal Medicine*, *135*, 344–351.

Ezzo, J., Berman, B. M., Vickers, A. J., & Klaus, L. (1998). Complementary medicine and the Cochrane Collaboration. *Journal of the American Medical Association*, *280*, 1628–1630.

Micozzi, M. S. (2006). *Fundamentals of complementary and integrative medicine*. St. Louis, MO: Saunders Elsevier.

Pelletier, K. R., & Astin, J. A. (2002). Integration and reimbursement of complementary and alternative medicine by managed care and insurance providers: 2000 update and cohort analysis. *Alternative Therapies*, *8*, 38–47.

Pelletier, K. R., Marie, A., Krasner, M., & Haskell, W. L. (1997). Current trends in the integration and reimbursement of complementary and alternative medicine by managed care, insurance carriers, and hospital providers. *American Journal of Health Promotion*, *12*, 112–122.

Silenzio, V. M. (2002). What is the role of complementary and alternative medicine in public health? *American Journal of Public Health*, *92*, 1562–1564.

Wetzel, M. S., Eisenberg, D. M., & Kaptchuk, T. J. (1998). Courses involving complementary and alternative medicine at U.S. medical schools. *Journal of the American Medical Association*, *280*, 784–787.

Confidence Interval

Most statistical analysis is conducted to reach some conclusion or decision about one or more parameters associated with a population of interest (statistical inference). There are two types of estimators used to assist in reaching a conclusion: point estimators and interval estimators. If the sampling distribution of the point estimator is known, then a likely range of values can be computed for the parameter being estimated. This range is called an interval estimator or confidence interval.

Before proceeding to more detail, here are some terms specific to confidence intervals. *Confidence limits* or *confidence bounds* are the upper and lower values of the range for the parameter given by the confidence interval. These limits are obtained from a sample and are random variables. The *pivotal quantity* is the point estimate used to estimate the population parameter and is the center of the confidence interval. The *half-width* of the confidence interval is the distance from the pivotal quantity to one of the limits and is a positive value. The desired significance level for the confidence interval is denoted by α (alpha), and is given by a number between 0 and 1. Typical values for α are between 0.01 and 0.05. The *confidence coefficient* is calculated as $(1 - \alpha)$, and is sometimes called the *coverage probability* or *inclusion probability* (although these terms are misleading, as explained later). Typical values for the confidence coefficient are between 0.95 and 0.99. The *confidence level* or *degree of confidence* is the confidence coefficient expressed as a percentage, $100(1 - \alpha)\%$. Typical values for the confidence level are between 95% and 99%. The *standard error* or *standard deviation* of the sampling distribution is used in calculating the confidence interval and depends on the particular point estimator used as well as the sampling method used.

Since point estimates do not provide a means for assessing the reliability or confidence placed on them, confidence intervals are preferred. They require no

additional information about reliability since it is provided by the degree of confidence and the half-width of the interval. The goodness of an interval estimation procedure can be determined by examining the fraction of times in repeated sampling that interval estimates would contain the parameter to be estimated. This fraction is called the confidence coefficient. The confidence coefficient is specified by the researcher, and it expresses how much assurance he or she wishes to have concerning whether the interval estimate encompasses or "covers" the parameter of interest. If 0.95 is the confidence coefficient associated with using a particular confidence interval formula, 95% of the time in repeated sampling, confidence intervals calculated using that formula will contain the true value of the population parameter. The flexibility to change confidence coefficients is another benefit of interval estimation compared with point estimation. Increasing the degree of confidence will widen a confidence interval; decreasing the degree of confidence will narrow a confidence interval.

There are three very common erroneous statements concerning confidence intervals. Here are these misstatements within the realm of using a sample mean, $\overline{X}$, to estimate a population mean, μ, with 95% confidence.

1. *There is a 95% chance that the confidence interval contains* μ. The parameter value is a fixed quantity, although it is unknown. Once a confidence interval is computed, the population value is either in it or not; there is no probabilistic statement to be made. For this reason, the terms *inclusion probability* and *coverage probability* are misleading when referring to the confidence coefficient.
2. *There is a 95% chance that the confidence interval contains* $\overline{X}$. Since the sample mean is the pivotal quantity that is the center of the confidence interval, there better be 100% certainty that the confidence interval contains $\overline{X}$.
3. *Ninety-five percent of the data are contained in the interval.* Regardless of whether the statement is referring to the sample data or the population data, the actual amount of data contained in the interval for the mean is irrelevant. However, the half-width of the 95% interval for individual values of x is approximately $\sqrt{n}$ times larger than the half-width of the 95% interval for the mean, where n is the sample size.

To help avoid misconceptions about confidence coefficients, some statisticians have suggested thinking of the sample value as fixed (the pivotal quantity) and then asking what parameter values (the confidence interval) make that sample value the most plausible. The chances are in the sampling procedure, not in the parameter. The confidence coefficient refers to the long-run proportion of intervals that include the population parameter based on using the same sampling procedure and the same sample size and the same interval estimator. In this sense, a researcher is $100(1-\alpha)\%$ confident that a computed confidence interval covers the true value of the parameter.

Cautious researchers will state their confidence levels as approximate levels. There are two reasons for this caveat being used: (1) The standard errors have been estimated from the data; (2) often an assumption of normality has been made about the population of data values. A confidence interval may be expressed in several different forms. Commonly used and equivalent expressions of a 95% confidence interval for a population mean are given by, "With approximately 95% confidence,"... (1) $3 < \mu < 10$; (2) μ lies between 3 and 10 units; (3) μ lies within 6.5 ± 3.5 units; and (4) μ is contained in the interval [3, 10].

Formulas for Calculations

The general form of a confidence interval is: $\hat{\theta} \pm cv_{\alpha/2}\sigma_{\hat{\theta}}$, where $\hat{\theta}$ is the pivotal quantity, $cv_{\alpha/2}$ is a critical value from an appropriate distribution (also called the $\alpha/2$ quantile of the distribution), and $\sigma_{\hat{\theta}}$ is the standard deviation of the sampling distribution of $\hat{\theta}$. Each different estimator $\hat{\theta}$ will have a different standard error formula, which depends on the sampling method used as well as the sample size. The formulas given here assume that a simple random sample was taken from the target population. If a different type of sample was used, then these formulas may not apply. Other formulas and additional information about confidence interval computation may be found in most elementary statistics texts.

- *Mean.* $100(1-\alpha)\%$ confidence interval for μ, where data are normally distributed and σ^2 is known: $\overline{X} \pm z_{\alpha/2}(\sigma/\sqrt{n})$. The z values come from the standard normal distribution.
- *Mean.* $100(1-\alpha)\%$ confidence interval for μ, where σ^2 is unknown and either n is large or data are normally distributed: $\overline{X} \pm t_{\alpha/2,\nu}(s/\sqrt{n})$. The t values come from the Student's t distribution with $\nu = n - 1$ *df*.

- *Mean difference, paired data.* $100(1-\alpha)\%$ confidence interval for $d=\mu_1-\mu_2$, where n is large and σ_d^2 is unknown: $\bar{d}\pm t_{\alpha/2,\nu}(s_d/\sqrt{n})$. The t values come from the Student's t distribution with $\nu=n-1$ *df*.
- *Proportion.* $100(1-\alpha)\%$ confidence interval for p or π, where n is large enough to have at least five in each category: $\hat{p}\pm z_{\alpha/2}\sqrt{[\hat{p}(1-\hat{p})]/n}$. The z value comes from the standard normal distribution.
- *Variance.* $100(1-\alpha)\%$ confidence interval for σ^2, where n is large:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,\nu}}\le\sigma^2\le\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2),\nu}}.$$

  The χ^2 values come from the χ^2 distribution (chi-square).
- *Odds Ratio (OR).* $100(1-\alpha)\%$ confidence interval for *OR*, where n is large,

$$\widehat{OR}=\frac{ad}{bc}\text{ and }\sigma_{\ln(\widehat{OR})}=\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}},$$

  where the standard 2×2 table presentation is

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix},$$

  which is

$$e^{\left[\ln(\widehat{OR})\pm z_{\alpha/2}\hat{\sigma}_{\ln(\widehat{OR})}\right]}\le OR\le e^{\left[\ln(\widehat{OR})\pm z_{\alpha/2}\hat{\sigma}_{\ln(\widehat{OR})}\right]}.$$

  The z value comes from the standard normal distribution.

—Stacie Ezelle Taylor

See also Critical Value; Inferential and Descriptive Statistics; Normal Distribution; Probability Sample; Random Variable; Sampling Distribution

Further Readings

Smithson, M. J. (2003). *Confidence intervals.* Thousand Oaks, CA: Sage.

Confounding

The word *confounding* has been used to refer to at least three distinct concepts. In the oldest and most widespread usage, confounding is a source of bias in estimating causal effects. This bias is sometimes informally described as a mixing of effects of extraneous factors (called confounders) with the effect of interest. This usage predominates in nonexperimental research, especially in epidemiology and sociology. In a second and more recent usage originating in statistics, confounding is a synonym for a change in an effect measure on stratification or adjustment for extraneous factors (a phenomenon called *noncollapsibility* or *Simpson's paradox*). In a third usage, originating in the experimental-design literature, confounding refers to inseparability of main effects and interactions under a particular design. The three concepts are closely related and are not always distinguished from one another. In particular, the concepts of confounding as a bias in effect estimation and as noncollapsibility are often treated as equivalent, even though they are not.

Confounding as a Bias in Effect Estimation

A classic discussion of confounding in which explicit reference was made to "confounded effects" is in Chapter 10 of John Stuart Mill's 1843 edition of *A System of Logic, Ratiocinative and Inductive.* In Chapter 3, Mill lays out the primary issues and acknowledges Francis Bacon as a forerunner in dealing with them. In Chapter 10, Mill lists a requirement for an experiment intended to determine causal relations: "None of the circumstances [of the experiment] that we do know shall have effects susceptible of being *confounded* [italics added] with those of the agents whose properties we wish to study."

In Mill's time, the word "experiment" referred to an observation in which some circumstances were under the control of the observer, as it still is used in ordinary English, rather than to the notion of a comparative trial. Nonetheless, Mill's requirement suggests that a comparison is to be made between the outcome of our "experiment" (which is, essentially, an uncontrolled trial) and what we would expect the outcome to be if the agents we wish to study had been absent. If the outcome is not as one would expect in the absence of the study agents, then Mill's requirement ensures that the unexpected outcome was not brought about by extraneous "circumstances" (factors). If, however, these circumstances do bring about the unexpected outcome and that outcome is mistakenly attributed to effects of the study agents, then the

mistake is one of confounding (or confusion) of the extraneous effects with the agent effects.

Much of the modern literature follows the same informal conceptualization given by Mill. Terminology is now more specific, with "treatment" used to refer to an agent administered by the investigator and "exposure" often used to denote an unmanipulated agent. The chief development beyond Mill is that the expectation for the outcome in the absence of the study exposure is now almost always explicitly derived from observation of a control group that is untreated or unexposed. Confounding typically occurs when natural or social forces or personal preferences affect whether a person ends up in the treated or control group, and these forces or preferences also affect the outcome variable. While such confounding is common in observational studies, it can also occur in randomized experiments when there are systematic improprieties in treatment allocation, administration, and compliance. A further and somewhat controversial point is that confounding (as per Mill's original definition) can also occur in perfectly randomized trials due to *random* differences between comparison groups; this problem will be discussed further below.

The Potential-Outcome Model of Confounding

Various models of confounding have been proposed for use in statistical analyses. Perhaps the one closest to Mill's concept is based on the *potential-outcome* or counterfactual model for causal effects. Suppose we wish to consider how a health-status (outcome) measure of a population would change in response to an intervention (population treatment). More precisely, suppose our objective is to determine the effect that applying a treatment x_1 had or would have on an outcome measure μ relative to applying treatment x_0 to a specific target population A. For example, cohort A could be a cohort of breast cancer patients, treatment x_1 could be a new hormone therapy, x_0 could be a placebo therapy, and the measure μ could be the 5-year survival probability. The treatment x_1 is sometimes called the *index* treatment and x_0 is sometimes called the *control* or *reference* treatment (which is often a standard or placebo treatment).

The potential-outcome model posits that in population A, μ will equal μ_{A1}if x_1 is applied, μ_{A0} if x_0 is applied; the causal effect of x_1 relative to x_0 is defined as the change from μ_{A0} to μ_{A1}, which might be measured as $\mu_{A1} - \mu_{A0}$ or μ_{A1}/μ_{A0}. If A is given treatment x_1, then μ will equal μ_{A1} and μ_{A1} will be observable, but μ_{A0} will be unobserved. Suppose, however, we expect μ_{A0} to equal μ_{B0}, where μ_{B0} is the value of the outcome μ observed or estimated for a population B that was administered treatment x_0. The latter population is sometimes called the control or reference population. *Confounding* is said to be present if $\mu_{A0} \neq \mu_{B0}$, for then there must be some difference between populations A and B (other than treatment) that affects μ.

If confounding is present, a naive (crude) association measure obtained by substituting μ_{B0} for μ_{A0} in an effect measure will not equal the effect measure, and the association measure is said to be *confounded.* For example, if $\mu_{A0} \neq \mu_{B0}$, then $\mu_{A1} - \mu_{B0}$, which measures the association of treatments with outcomes *across* the populations, is confounded for $\mu_{A1} - \mu_{A0}$, which measures the effect of treatment x_1 on population A. Thus, to say an association measure $\mu_{A1} - \mu_{B0}$ is confounded for an effect measure $\mu_{A1} - \mu_{A0}$ is to say these two measures are not equal. A noteworthy aspect of this view is that confounding depends on the outcome measure. For example, suppose populations A and B have a different 5-year survival probability μ under placebo treatment x_0; that is, suppose $\mu_{B0} \neq \mu_{A0}$, so that $\mu_{A1} - \mu_{B0}$ is confounded for the actual effect $\mu_{A1} - \mu_{A0}$ of treatment on 5-year survival. It is then still possible that 10-year survival, υ, under the placebo would be identical in both populations; that is, υ_{A0} could still equal υ_{B0}, so that $\upsilon_{A1} - \upsilon_{B0}$ is not confounded for the actual effect of treatment on 10-year survival. (We should generally expect no confounding for 200-year survival, since no treatment is likely to raise the 200-year survival probability of human patients above zero.)

A second noteworthy point is that confounding depends on the target population of inference. The preceding example, with A as the target, had different 5-year survivals μ_{A0} and μ_{B0} for A and B under placebo therapy, and hence $\mu_{A1} - \mu_{B0}$ was confounded for the effect $\mu_{A1} - \mu_{A0}$ of treatment on population A. A lawyer or ethicist may also be interested in what effect the hormone treatment would have had on population B. Writing μ_{B1} for the (unobserved) outcome under treatment, this effect on B may be measured by $\mu_{B1} - \mu_{B0}$. Substituting μ_{A1} for the unobserved μ_{B1} yields $\mu_{A1} - \mu_{B0}$. This measure of association is confounded for $\mu_{B1} - \mu_{B0}$ (the effect of treatment x_1 on 5-year survival in population B) if

and only if $\mu_{A1} \neq \mu_{B1}$. Thus, the same measure of association, $\mu_{A1} - \mu_{B0}$, may be confounded for the effect of treatment on neither, one, or both of populations A and B, and may or may not be confounded for the effect of treatment on other targets.

Confounders (Confounding Factors)

A third noteworthy aspect of the potential-outcome model is that it invokes no explicit differences (imbalances) between populations A and B with respect to circumstances or covariates that might influence μ. Clearly, if μ_{A0} and μ_{B0} differ, then A and B must differ with respect to factors that influence μ. This observation has led some authors to define confounding as the presence of such covariate differences between the compared populations. Nonetheless, confounding is only a consequence of these covariate differences. In fact, A and B may differ profoundly with respect to covariates that influence μ, and yet confounding may be absent. In other words, a covariate difference between A and B is a necessary but not sufficient condition for confounding, as can be seen when the impact of covariate differences may balance each other out, leaving no confounding.

Suppose now that populations A and B differ with respect to certain covariates and that these differences have led to confounding of an association measure for the effect measure of interest. The responsible covariates are then termed *confounders* of the association measure. In the above example, with $\mu_{A1} - \mu_{B0}$ confounded for the effect $\mu_{A1} - \mu_{A0}$, the factors responsible for the confounding (i.e., the factors that led to $\mu_{A0} \neq \mu_{B0}$) are the confounders. It can be deduced that a variable cannot be a confounder unless it can affect the outcome parameter μ within treatment groups and it is distributed differently among the compared populations. These two necessary conditions are sometimes offered together as a definition of a confounder. Nonetheless, counterexamples show that the two conditions are not sufficient for a variable with more than two levels to be a confounder. Note that the condition of affecting the outcome parameter is a causal assertion and thus relies on background knowledge for justification.

Prevention of Confounding

An obvious way to avoid confounding is estimating $\mu_{A1} - \mu_{A0}$ to obtain a reference population B for which μ_{B0} is known to equal μ_{A0}. Such a population is sometimes said to be *comparable* to or *exchangeable* with A with respect to the outcome under the reference treatment. In practice, such a population may be difficult or impossible to find. Thus, an investigator may attempt to construct such a population, or to construct exchangeable index and reference populations. These constructions may be viewed as design-based methods for the control of confounding.

Perhaps no approach is more effective for preventing confounding by a known factor than *restriction.* For example, gender imbalances cannot confound a study restricted to women. However, there are several drawbacks: Restriction on enough factors can reduce the number of available subjects to unacceptably low levels and may greatly reduce the generalizability of results as well. Matching the treatment populations on confounders overcomes these drawbacks and, if successful, can be as effective as restriction. For example, gender imbalances cannot confound a study in which the compared groups have identical proportions of women.

Unfortunately, differential losses to observation may undo the initial covariate balances produced by matching. Neither restriction nor matching prevents (although it may diminish) imbalances on unrestricted, unmatched, or unmeasured covariates. In contrast, *randomization* offers a means of dealing with confounding by covariates not accounted for by the design. It must be emphasized, however, that this solution is only probabilistic and subject to severe constraints in practice.

Randomization is not always feasible or ethical, and many practical problems, such as differential loss (dropout) and noncompliance, can lead to confounding in comparisons of the groups actually receiving treatments x_1 and x_0. One somewhat controversial solution to dropout and noncompliance problems is *intent-to-treat analysis*, which defines the comparison groups A and B by treatment assigned rather than treatment received. Confounding may, however, affect even intent-to-treat analyses, and (contrary to widespread misperceptions) the bias in those analyses can exaggerate the apparent treatment effect. For example, the assignments may not always be random, as when blinding is insufficient to prevent the treatment providers from protocol violations. And, purely by bad luck, randomization may itself produce allocations with severe covariate imbalances between the groups (and consequent confounding), especially if

the study size is small. *Blocked* (matched) randomization can help ensure that random imbalances on the blocking factors will not occur, but it does not guarantee balance of unblocked factors.

Adjustment for Confounding

Design-based methods are often infeasible or insufficient to prevent confounding. Thus, there has been an enormous amount of work devoted to analytic adjustments for confounding. With a few exceptions, these methods are based on observed covariate distributions in the compared populations. Such methods can successfully control confounding only to the extent that enough confounders are adequately measured. Then, too, many methods employ parametric models at some stage, and their success may thus depend on the faithfulness of the model to reality.

The simplest and most widely trusted methods of adjustment begin with *stratification* on confounders. A covariate cannot be responsible for confounding within internally homogeneous strata of the covariate. For example, gender imbalances cannot confound observations within a stratum composed solely of women. More generally, comparisons within strata cannot be confounded by a covariate that is unassociated with treatment within strata. This is so, whether the covariate was used to define the strata or not. Thus, one need not stratify on all confounders to control confounding. Furthermore, if one has accurate background information on relations among the confounders, one may use this information to identify sets of covariates statistically sufficient for adjustment. Nonetheless, if the stratification on the confounders is too coarse (e.g., because categories are too broadly defined), stratification may fail to adjust for much of the confounding by the adjustment variables.

One of the most common adjustment approaches today is to enter suspected confounders into a model for the outcome parameter μ. For example, let μ be the mean (expectation) of an outcome variable of interest Y, let X be the treatment variable of interest, and let Z be a suspected confounder of the $X - Y$ relation. Adjustment for Z is often made by fitting a generalized-linear model $g(\mu) = g(\alpha + \beta x + \gamma z)$ or some variant, where $g(\mu)$ is a strictly increasing function such as the natural logarithm $\ln(\mu)$, as in log-linear modeling, or the logit function $\ln\{\mu/(1-\mu)\}$, as in logistic regression. The estimate of the exposure coefficient β that results is then taken as the Z-adjusted estimate of the X effect on $g(\mu)$.

An oft-cited advantage of model-based adjustment methods is that they allow adjustment for more variables and in finer detail than stratification. If, however, the form of the fitted model cannot adapt well to the true dependence of Y on X and Z, such model-based adjustments may fail to adjust for confounding by Z. For example, suppose Z is symmetrically distributed around zero within X levels, and the true dependence is $g(\mu) = g(\alpha + \beta x + \gamma z^2)$; then using the model $g(\mu) = g(\alpha + \beta x + \gamma z)$ will produce little or no adjustment for Z. Similar failures can arise in adjustments based on models for treatment probability (often called "propensity scores"). Such failures can be minimized or avoided by using reasonably flexible models, by carefully checking each fitted model against the data, and by combining treatment-probability and outcome models to produce more robust effect estimators.

Finally, if (as is often done) a variable used for adjustment is not a confounder, bias may be introduced by the adjustment. The form of this bias often parallels *selection bias* familiar to epidemiologists and tends to be especially severe if the variable is affected by both the treatment and the outcome under study, as in classic *Berksonian bias* (e.g., in the use of hospital-based controls when hospitalization is related to exposure). In some cases, the resulting bias is a form of confounding within strata of the covariate; adjustment for covariates affected by treatment can produce such confounding, even in randomized trials.

Confounded Mechanisms Versus Confounded Assignments

If the mechanism by which the observational units come to have a particular treatment is independent of the potential outcomes of the units, the mechanism is sometimes described as *unconfounded* or *unbiased* for μ; otherwise, the mechanism is confounded or biased. Randomization is the main practical example of such a mechanism. Graphical models provide an elegant algorithm for checking whether the graphed mechanism is unconfounded within strata of covariates. Note, however, that in typical epidemiologic usage, the term *confounded* refers to the result of a single assignment (the study group actually observed), and not the behavior of the mechanism. Thus, an unconfounded mechanism can by chance produce confounded assignments.

The latter fact resolves a controversy about adjustment for baseline (pretreatment) covariates in randomized trials. Although some assert that randomized comparisons are ''unbiased'' without adjustment, this unbiasedness is actually a property of the mechanism. Particular assignments can be confounded in the single-trial sense used in epidemiology. Once the trial is under way and the actual treatment allocation is completed, the unadjusted treatment-effect estimate will be biased conditional on the observed allocation if the baseline covariate is associated with treatment in the allocation, and the covariate affects the outcome; this bias can be removed by adjustment for the covariate.

Confounder Selection

An essential first step in the control of confounding is to identify which variables among those measured satisfied the minimal necessary conditions to be a confounder. This implies among other things that the variables cannot be affected by exposure or outcome; it thus excludes intermediate variables and effects of exposure and disease, whose control could introduce Berksonian bias. This initial screening is primarily a subject-matter decision that requires consideration of the causal ordering of the variables. Relatively safe candidate confounders will be ''pretreatment'' covariates (those occurring before treatment or exposure), which at the very least have the advantage that they cannot be intermediates or effects of exposure and outcome. Exceptions occur in which control of certain pretreatment variables introduce bias, although the bias so introduced may be much less than the confounding removed.

Variables that pass the initial causal screening are sometimes called ''potential confounders.'' Once these are identified, the question arises as to which must be used for adjustment. A common but unjustified strategy is to select confounders to control based on a test (usually a significance test) of each confounder's association with the treatment X (a test of imbalance) or with the outcome Y, for example, using stepwise regression. Suppose Z is a pretreatment covariate (potential confounder). The strategy of testing the Z association with X arises from a confusion of two distinct inferential problems:

1. Do the treated ($X = 1$) evince larger differences from the untreated ($X = 0$) with respect to Z than one should expect from a random (or unconfounded) assignment mechanism?
2. Should we control for Z?

A test of the $X - Z$ association addresses the first question, but not the second. For Question 2, the ''large sample'' answer is that control is advisable, regardless of whether the $X - Z$ association is random. This is because an imbalance produces bias conditional on the observed imbalance, even if the imbalance is random. The mistake of significance testing lies in thinking that one can ignore an imbalance if it is random, which is not so. Random assignment guarantees only valid performance of statistics unconditionally, averaged over all possible treatment assignments. It does not, however, guarantee validity conditional on the observed Z imbalance, even though any such imbalance must be random in a randomized trial. Thus, the $X - Z$ test addresses a real question (one relevant to studying determinants of response/treatment), but is irrelevant to the second question.

The case of testing the Z association with Y devolves in part to whether one trusts prior (subject-matter) knowledge that Z affects Y (or is a proxy for a cause of Y) more than the results of a significance test in one's own limited data. There are many examples in which a well-known risk factor exhibits the expected association with Y in the data, but for no more than chance reasons or sample-size limitations that association fails to reach conventional levels of ''significance.'' In such cases, there is a demonstrable statistical advantage to controlling Z, thus allowing subject-matter knowledge to override nonsignificance.

Another problematic strategy is to select a potential confounder Z for control based on how much the effect estimate changes when Z is controlled. Like the testing methods described above, it also lacks formal justification and can exhibit poor performance in practice. The strategy can also mislead if the treatment affects a high proportion of subjects and one uses a ''noncollapsible'' effect measure (one that changes on stratification even if no confounding is present), such as an odds ratio or rate ratio.

In practice, there may be too many variables to control using conventional methods, so the issue of confounder selection may seem pressing. Nonetheless, hierarchical-Bayesian or other ''shrinkage'' methods may be applied instead. These methods adjust for all the measured confounders by estimating the confounder effects using a prior distribution for those effects. Some of these methods (e.g., the

"Lasso") may drop certain variables entirely, and thus in effect result in confounder selection; unlike significance-testing-based selection, however, this selection has a justification in statistical theory.

—*Sander Greenland*

See also Causal Diagrams; Causation and Causal Inference; Propensity Score; Simpson's Paradox

Further Readings

General Issues in Confounding and Confounder Selection

Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding, *International Journal of Epidemiology*, *15*, 413–419.

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, *14*, 29–46.

Hernán, M., Hernandez-Diaz, S., Werler, M. M., & Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation. *American Journal of Epidemiology*, *155*, 176–184.

Pearl, J. (2000). *Causality* (chap. 6). New York: Cambridge University Press.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed., chaps. 4, 15, 20, 21). Philadelphia: Lippincott.

Confounding and Covariate Control in Randomized Experiments

Cox, D. R. (1958). *Planning of experiments* (chap. 2). New York: Wiley.

Fisher, R. A. (1935). *The design of experiments* (p. 49). Edinburgh: Oliver & Boyd.

Rothman, K. J. (1977). Epidemiologic methods in clinical trials. *Cancer*, *39*, 1771–1775.

Historically Important Discussions of Confounding

Mill, J. S. (1956). *A system of logic, ratiocinative and inductive* [Reprint]. London: Longmans, Green. (Original work published 1843)

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, *2*, 121–134.

Control Group

Many experiments are designed to include a control group and one or more experimental groups; in fact, some scholars reserve the term *experiment* for study designs that include a control group. Ideally, the control group and the experimental groups are identical in every way except that the experimental groups are subjected to treatments or interventions believed to have an effect on the outcome of interest, while the control group is not. The control group is the standard to which comparisons are made in an experiment, and inclusion of a control group greatly strengthens the study's ability to draw conclusions.

A typical use of a control group is in an experiment in which the effect of a treatment is unknown, and comparisons between the control group and the experimental group are used to measure the effect of the treatment. For instance, in a pharmaceutical study to determine the effectiveness of a new drug on the treatment of migraines, the experimental group will be administered the new drug and the control group will be administered a placebo (a drug that is inert, or at least that is assumed to have no effect on migraines). Each group is then given the same questionnaire and asked to rate the effectiveness of the drug in relieving symptoms. If the new drug is effective, the experimental group is expected to have a significantly better response to it than the control group. Another possible design is to include several experimental groups, each of which is given a different dosage of the new drug, plus one control group. In this design, the analyst will compare results from each of the experimental groups to the control group. This type of experiment allows the researcher to determine not only if the drug is effective but also the effectiveness of different dosages. In the absence of a control group, the researcher's ability to draw conclusions about the new drug is greatly weakened, due to the placebo effect and other threats to validity. Comparisons between the experimental groups with different dosages can be made without including a control group, but there is no way to know if any of the dosages of the new drug are more or less effective than the placebo.

It is important that every aspect of the experimental environment be as alike as possible for all subjects in the experiment. If conditions are different for the experimental and control groups, it is impossible to know whether differences between groups are actually due to the difference in treatments or to the difference in environment. For example, in the new migraine drug study, it would be a poor study design to administer the questionnaire to the experimental

group in a hospital setting while asking the control group to complete it at home. Such a study could lead to a misleading conclusion, because differences in responses between the experimental and control groups could have been due to the effect of the drug, or could have been due to the conditions under which the data were collected. For instance, perhaps the experimental group got better instructions or it was more motivated by being in the hospital setting to give accurate responses than was the control group.

A control group study can be managed in two different ways. In a single-blind study, the researcher will know whether a particular subject is in the control group, but the subject will not know. In a double-blind study, neither the subject nor the researcher will know which treatment the subject is receiving. In many cases, a double-blind study is preferable to a single-blind study, since the researcher cannot inadvertently affect the results or their interpretation by treating a control subject differently from an experimental subject.

Only in the presence of a control group can a researcher determine whether a treatment under investigation truly has a significant effect on an experimental group and the possibility of making an erroneous conclusion is reduced.

—Mary Earick Godby

See also Clinical Trials; Hawthorne Effect; Placebo Effect; Quasi Experiments; Study Design

Further Readings

Montgomery, D. C. (1984). *Design and analysis of experiments*. New York: Wiley.

Neter, J., Wasserman, W., & Kutner, M. H. (1985). *Applied linear statistical models*. Homewood, IL: Irwin.

Control Variable

In research design, a control variable is defined as a variable that is known to or expected to influence the dependent variable and might also affect the explanatory or independent variable in an analysis, but is not the focus of interest for the researcher. The influence of a control variable may interfere with the main analysis, for instance, by obscuring between-treatment or between-group differences, or creating apparent relationships between variables of interest. Often variables of this type are used to create blocks in an experimental design or stratify a sample. Occasionally, the examination of the influence of control variables is called "elaboration of the analysis" because they are not the variables of main interest in the analysis.

Within the realm of designed experiments, a control variable may be kept constant or controlled for each test or replication of the experiment. In observational studies, control variables are often used in analysis to divide the study population into smaller, more homogeneous groups, to test for the influence of potentially confounding factors. For example, if a researcher wanted to compare death rates between smokers and nonsmokers, he or she would also need to control for gender (males are the largest proportion of smokers, and males also have higher incidences of heart disease). Therefore, death rate differences between smokers and nonsmokers may actually be due to differences between genders. This could be examined by performing a stratified analysis in which data from males and females were analyzed separately. Control variables may also be included in the analysis; for instance, a continuous control variable may be included in a generalized linear model before the independent variables of interest, to separate the amount of variance accounted for by each.

Control variables have many other names within the literature (see Table 1). Some of the names may

Table 1 Other Names of Control Variables

Alternate Names	*Type of Relationship*
Antecedent variables	Figure 1a
Concomitant variables	Figures 1a to d
Confounding variables	Figures 1a to d
Covariates	Figures 1a to d
Incidental variables	Figure 1c
Indirect variables	Figure 1a
Intervening variables	Figure 1b
Mediating variables	Figure 1b
Stratification variables	Figures 1c and d
Subordinate variables[a]	Figure 1b
Test variables	Figures 1a to d

[a]Usually refers to dependent variables, but sometimes used for control variables.

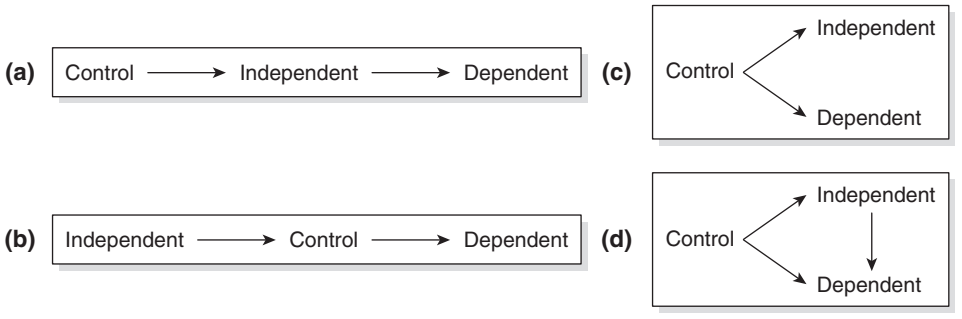


Figure 1 Various Relationships Between Control, Independent, and Dependent Variables

be used interchangeably with "control variable"; while other names indicate the specific relationship between the control variable and the independent and dependent variables. Figures 1a through d illustrate the various types of relationships that may be present between control, independent, and dependent variables.

Antecedent or indirect control variables influence the independent variables within an analysis, as illustrated in Figure 1a, which in turn influence the dependent variable(s).

Intervening, mediating, or subordinate control variables are in the causal chain between the independent and dependent variables, as illustrated in Figure 1b. In this case, the independent variable influences the control variable, which in turn influences the dependent variable, so that the effects of the independent variable on the dependent variable are seen only through the influence of the control variable.

If two statistically related variables become statistically independent when a third variable is included in the analysis, then the control variable may be called an incidental variable, as illustrated in Figure 1c. Usually, this type of relationship between the two variables of interest is referred to as spurious because it is due to the influence of the control variable rather than due to any inherent relationship between the variables of interest. A classic example of a spurious relationship is the relationship between ice cream sales and murder rates. These two variables are significantly positively correlated over the calendar year: When ice cream sales increase, so does the murder rate. However, taking into account the third variable of temperature as a control variable, we find the relationship between sales and murder rates is spurious, because both ice cream sales and murders tend to increase when the temperature increases, as in the hottest days of summer.

Sometimes, introducing a control variable reduces, but does not eliminate, the association between independent and dependent variables. This is illustrated in Figure 1d when the control variable influences both the independent and the dependent variables, but they also have a relationship independent of the control variable. The inclusion of a control variable in an analysis may also reverse the nature of the relationship between the independent and dependent variables.

If a variable may potentially be a confounding factor, it is best to include it as a control variable within the analysis. Potential control variables may be identified from the researcher's experience, from a literature review, from a conceptual model that guides research, or from the researcher's hypothesis. For example, if two blood pressure medications are being compared for their abilities to reduce systolic blood pressure, potential confounding factors include, but are not limited to, starting systolic blood pressure, age of patients, and activity level. By using starting systolic blood pressure and age as continuous covariates and an ordinal (categorical) covariate of activity level within the analysis, these factors become control variables.

In the previously referenced analysis of death rates between smokers and nonsmokers, age would be a confounding factor in addition to gender. Older people tend to smoke more heavily and are, therefore, more at risk for lung cancer. Hence, a comparison of death rates between smokers and nonsmokers should be done separately by gender as well as age. For example, male smokers below the age of 25 would be compared with male nonsmokers below

the age of 25; while female smokers above the age of 65 would be compared with female nonsmokers above the age of 65. These rates could also be compared through standardization of the rates.

—*Stacie Ezelle Taylor*

See also Analysis of Covariance; Confounding; Dependent and Independent Variables; Effect Modification and Interaction; Mediating Variable; Study Design

Further Readings

Analysis Group of PAHO's Special Program for Health Analysis (SHA). (2002, September). Standardization: A classic epidemiological method for the comparison of rates. *Epidemiological Bulletin*, *23*(3). Retrieved January 15, 2007, from http://www.paho.org/english/sha/be_v23n3-standardization.htm.

Hildebrand, D. K. (1986). *Statistical thinking for behavioral scientists.* Boston: Duxbury Press.

Vasu, M. L. (1999). *Day 12: The elaboration model*. PS471 Class notes. Retrieved January 15, 2007, from http://www2.chass.ncsu.edu/mlvasu/ps471/D12.htm.

Convenience Sample

Nonprobability sampling is used by researchers to find members of a population pool when this population cannot be enumerated to allow for a probability sample to be constructed. There are four general categories of nonprobability sampling: convenience, quota, purposive, and snowball sampling.

A convenience sample is a selected group from a particular population that is chosen based on their accessibility to the researcher. Examples of convenience samples include selecting neighborhood residents by inviting participants from those entering a supermarket, selecting grade school students from the local elementary school, and selecting undergraduate students from a freshman-level class.

A quota sample is a convenience sample in which the researcher seeks to control some variability in his or her sample. As in a probability sampling—stratified sampling method—a researcher may preselect the proportions of certain strata that he or she desires in the final sample. For example, a researcher who desires that 50% of the final sample of 100 students be male may recruit subjects from a freshman-level class of undergraduates with a different proportion of male students but restrict participation among males once 50 agree to participate, focusing solely on the inclusion of females until 50 are included in the sample.

A purposive sample is a convenience sample whose population parameters are defined by the purpose of the research. For example, a researcher may want to understand more about undergraduate students who work full-time. To obtain potential participants, the researcher may place fliers throughout the school asking specifically for undergraduates who are currently holding a full-time job.

A snowball sample is a convenience sample in which the researcher identifies a smaller pool of accessible members of the particular population and requests that those members identify other appropriate potential members. This method is similar to a purposive sample selection but is useful if access to the population pool is particularly challenging for the researcher but less so for members of that population. For example, a researcher interested in organ donation may use a known organ donor who is associated with donor support groups to facilitate access to the members of these groups.

Although it is impossible to determine the probability of selection in a nonprobability sample, and thus a researcher cannot statistically measure the representation of the selected population, this does not necessarily mean that the sample is not representative. Whereas nonprobability samples cannot be assessed quantitatively, they can and should be assessed qualitatively with regard to the true sampling frame of the participants in the study. Often, a researcher can use the demographic characteristics of his or her sample to explore how well they compare with the population of interest. Additionally, a researcher should be thoughtful as to why the particular group was accessible and if this in any way may influence the findings of the study.

—*Eve Waltermaurer*

See also Bias; Probability Sample; Sampling Techniques; Study Design

Further Readings

Karney, B., Davila, J., Cohan, C., Sullivan, K., Johnson, M., & Bradbury, T. (1995). An empirical investigation of sampling strategies in marital research. *Journal of Marriage and Family*, *57*, 909–920.

Council of State and Territorial Epidemiologists

The Council of State and Territorial Epidemiologists (CSTE) is a professional organization whose mission is to advance health and guide public health procedures through the appropriate use of epidemiologic data. The organization's activities include the identification of nationally notifiable diseases, development of standardized case definitions and operating procedures, epidemiologic capacity building, and advocacy for the use of epidemiologic information in policy development and decision making.

History

In the United States, during the late 19th century, to prevent the introduction and spread of cholera, smallpox, plague, and yellow fever, Congress charged the U.S. Marine Hospital Service, now the Public Health Service, with the first official collection information from individuals afflicted with disease. For the next 60 years, state and territorial health authorities worked with the Public Health Service to designate additional diseases to be reported. Then, in 1951, under the direction of Alexander Langmuir, the Centers for Disease Control and Prevention's (CDC's) first Epidemiology Division Director, the Association of State and Territorial Health Officers convened the first national meeting of state and territorial epidemiologists. At this meeting, CSTE, initially the Conference of State and Territorial Epidemiologists, was established and, with input from the CDC, given the authority to identify and recommend nationally notifiable diseases for inclusion in the National Public Health Surveillance System list.

Purpose and Select Activities

CSTE aims to improve health and affect public health practice through the appropriate use of epidemiology. To accomplish this, CSTE works with the CDC on a number of projects.

Standardized Definitions and Procedures. To assure that the quality of data being reported by each state and territory is adequate, CSTE and the CDC collaborate to create standardized case definitions and surveillance procedures for notifiable diseases. In addition to communicable diseases, which are typically notifiable, CSTE works with both the National Institute of Occupational Safety and Health and the National Center for Chronic Disease Prevention and Health Promotion to prioritize conditions and risks for national surveillance.

Assessment. CSTE actively seeks to assess and enhance the epidemiologic capabilities of the states and territories. In 2001 and 2004, CSTE surveyed health departments to assess their core epidemiologic capacity and future training needs.

Advice and Expertise. CSTE provides advice and technical support to epidemiologists and also designates surveillance consultants to the CDC and other federal and international agencies.

Membership and Organization

CSTE members include epidemiologists representing 50 states, 8 territories, and Puerto Rico. The organization hosts two types of members. Active members include epidemiologists who currently work for a state, territorial, or a local health department. Associate members include epidemiologists who work in federal health agencies or academia.

—Michelle Kirian

See also Centers for Disease Control and Prevention; Notifiable Disease; Public Health Surveillance

Further Readings

Boulton, M. L., Abellera, J., Lemmings, J., & Robinson, L. (2005). Assessment of epidemiologic capacity in state and territorial health departments: United States, 2004. *Morbidity and Mortality Weekly Report*, *54*(18), 457–459.

Centers for Disease Control and Prevention. (2006, January). *National Notifiable Diseases Surveillance System.* Retrieved November 6, 2006, from http://www.cdc.gov/EPO/DPHSI/nndsshis.htm.

Council of State and Territorial Epidemiologists. (n.d.). Retrieved November 6, 2006, from http://www.cste.org.

Counterfactual Models

See Attributable Fractions

Cox Model

The Cox proportional hazards model is a regression technique that allows modeling of survival times (or hazard functions) as a function of a set of covariates. The model was first introduced by D. R. Cox in 1972. The Cox model makes it possible to estimate the survival distribution while accounting for a number of covariates simultaneously or to compare the hazard functions of two or more groups while adjusting for discrete or continuous covariates. It is widely used in medical research to investigate the association between survival and a set of possible risk factors. In randomized clinical trials, the Cox model is used to determine the efficacy of new treatments or therapies on survival or on the occurrence of an event of interest (e.g., disease remission or recurrence).

The Cox proportional hazards model can be considered a generalization of the Kaplan-Meier (or product limit) estimator of a survival curve that accounts for both discrete and continuous risk factors. A great advantage of the Cox model compared with other regression approaches for survival data is its simplicity and the interpretability of its parameter estimates.

Form of the Model

Let $\mathbf{z} = z_1, z_2, \ldots, z_p$ be a $p \times 1$ vector of covariates or risk factors and let $h(t|\mathbf{z})$ be the hazard function, which depends on the covariates $\mathbf{z}$. The general form of the proportional hazard model is

$$h(t|\mathbf{z}) = \psi(\mathbf{z})h_0(t)$$

where $\psi(\mathbf{z})$ is a function of the covariates, and $h_0(t)$ is the underlying baseline hazard. $h_0(t)$ can be interpreted as the hazard of death in the absence of an effect of the covariates or when the covariates assume a reference value (e.g., absence of an exposure factor or membership to a placebo or a standard therapy group).

The most common form of the function $\psi(\mathbf{z})$ is

$$\psi(\mathbf{z}) = \mathrm{e}^{\beta'\mathbf{z}},$$

so that the hazard, given the covariates, can be expressed as

$$h(t|\mathbf{z}) = h_0(t) \exp\left(z_1\boldsymbol{\beta}_1 + z_2\boldsymbol{\beta}_2 + \cdots + z_p\boldsymbol{\beta}_p\right).$$

In this model, the predictors $(z_1, z_2, \ldots, z_p)$ are assumed to act additively on $\ln h(t|\mathbf{z})$. $\boldsymbol{\beta} = \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p$ is a $p \times 1$ vector of unknown parameters that relate the covariates $\mathbf{z}$ to the hazard.

While in parametric survival models we make an assumption about the distribution of the baseline hazard $h_0(t)$, in the Cox proportional hazard model we make no assumption about the shape of the underlying hazard. The only parametric part of the model is the one that contains information about the predictors, $\psi(\mathbf{z})$. For this reason, the Cox model is called a semi-parametric model. Since in most applications we are interested more in the relationship between the predictors and the survival, rather than the shape of the hazard, the Cox proportional hazard model is suitable for most applied problems. Note that the model does not have an intercept as the baseline hazard $h_0(t)$ accounts for that. The Cox model can be used to estimate the survival function $s(t)$ for fixed values of the covariates.

The Partial Likelihood and Parameter Estimation

As with other regression models, estimation of the regression parameters for the Cox model proceed by maximizing the likelihood function.

Although the full likelihood for the survival data with covariates can be expressed as the product of two terms,

$$L(\boldsymbol{\beta}, h_{0(t)}) = L_1(\boldsymbol{\beta}) + L_2(\boldsymbol{\beta}, h_{0(t)}),$$

the first one involving the model parameters $\boldsymbol{\beta}$, and the second one involving both $\boldsymbol{\beta}$ and the baseline hazard $h_0(t)$, valid inference about the regression parameters can be derived by using $L_1(\boldsymbol{\beta})$ only. The function $L_1(\boldsymbol{\beta})$ is called a partial likelihood and was first introduced by Cox in 1975. The rationale for using the partial likelihood rather than the full likelihood is that it is our interest to make inference about the regression parameters, while we are not concerned about the form of the baseline hazard. Since the partial likelihood $L_1(\boldsymbol{\beta})$ contains most of the information about $\boldsymbol{\beta}$, it is sufficient for estimating the model parameters. In fact, it has been shown that the partial likelihood approach is almost as efficient as the full likelihood approach albeit much simpler.

To see how the partial likelihood is constructed, consider a sample of N individuals who are followed up in time prospectively. During the observation period, suppose that K of these individuals die. Also assume that $N - K$ individuals are right-censored,

that is, do not die during the observation period. Let $t_1 \leq t_2 \leq \cdots t_K$ be the ordered failure times for the K individuals who die during the observation period.

For the generic individual $j(j = 1, 2, \ldots, N)$, let

t_j = observed follow-up time
z_j = vector of predictors
$R(t_j)$ = the risk set at time t_j, that is, the number of individuals who are alive and at risk of death at time t_j.

The probability that a generic individual $j(j = 1, K)$ with covariates z_j dies at time t_j, given that $R(t_j)$ individuals are at risk and only one individual dies at t_j, is given by

$$L_j = \frac{e^{\beta z_{(j)}}}{\left(\sum_{l \in R(t_j)} e^{\beta z_{(l)}} \right)}.$$

The partial likelihood is then obtaining as the product of all these probabilities across all failing individuals in the sample.

$$L_p(\beta) = \prod_{j=1}^{k} \frac{e^{\beta z_{(j)}}}{\left(\sum_{l \in R(t_j)} e^{\beta z_{(l)}} \right)}.$$

Thus, the partial likelihood is obtained by comparing the risk for the failing individual at a specific time with the risk of all other individuals at the same time.

It should be noted that censored observations contribute information only in the denominator of the partial likelihood, not the numerator. Since the partial likelihood is the product of terms each contributing a small amount of information about the parameters $\boldsymbol{\beta}$, the goodness of the partial likelihood does not depend on the sample size but on the number of censored observations. The larger the number of censored observations, the less informative the partial likelihood.

The partial likelihood is valid when there are no ties in the data set, that is, there are no two observations with the same survival time. In cases of ties, the Breslow approximation to the partial likelihood should be used.

Estimates of the regression parameters $\boldsymbol{\beta}$ are obtained as the values that maximize the partial likelihood.

Interpretation of the Regression Parameters

From the model

$$h(t|\mathbf{z}) = h_0(t)e^{\boldsymbol{\beta}'\mathbf{z}}$$

we can derive

$$\frac{h(t|\mathbf{z})}{h_0(t)} = e^{\boldsymbol{\beta}'\mathbf{z}},$$

so that

$$\ln\, h(t|\mathbf{z}) - \ln\, h_0(t) = \boldsymbol{\beta}' z.$$

Thus, $\boldsymbol{\beta}$ is the log-relative risk of a subject with covariates vector z compared with a subject with covariate vector $z = 0$.

Consider the following simple example with one discrete covariate. Let z indicate whether or not individuals are exposed to a risk factor (e.g., smoking), so that

$$z = \begin{cases} 0 & \text{if the exposure is absent} \\ 1 & \text{if the exposure is present,} \end{cases}$$

and suppose we are interested in the effect of this exposure on survival.

The Cox model for this simple case is

$$h(t|\mathbf{z}) = h_0(t)e^{\beta z}.$$

From the model above we derive that the hazard for an unexposed individual is

$$h(t|\mathbf{z} = 0) = h_0(t)e^{\beta(0)} = h_0(t),$$

while the hazard for an exposed individual is

$$h(t|\mathbf{z} = 1) = h_0(t)e^{\beta(1)} = h_0(t)e^{\beta}.$$

The ratio of these two quantities is called the hazard ratio, or risk ratio or relative risk, and is expressed as

$$RR(t) = \frac{h(t|z = 1)}{h(t|z = 0)} = \frac{h_0 e^{\beta}}{h_0} = e^{\beta}.$$

Thus, the parameter β indicates the amount (on an exponential scale) by which the risk of death increases or decreases depending on the exposure. The Cox model can be modified to account for time-varying covariates (in this case, the model is called a hazard model, rather than a proportional hazards model).

Note that $RR(t) = e^{\beta}$ is independent of time. Thus, the model states that the relative risk is constant over time, that is, the effect of the predictors is the same at all times. This characteristic of the Cox model is called proportional hazards property.

Hypothesis Testing About the Regression Parameters

Given two samples of exposed and unexposed individuals, testing the null hypothesis

$$H_0 : \beta = 0$$

that the exposure has no effect on survival is equivalent to testing the hypothesis

$$H_0 : S_E = S_U$$

that the survival distribution of the unexposed sample is identical to the survival distribution of the exposed sample. When the covariate of interest is discrete, testing the hypothesis that $\beta = 0$ is equivalent to performing a log-rank test.

The test statistic for the null hypothesis $H_0 : \beta = 0$ takes the form

$$Z = \frac{\hat{\beta}}{se_{\hat{\beta}}}.$$

This test follows a standard normal distribution with mean 0 and variance 1 under the null hypothesis. Thus, values of this test that exceed the critical value for a prespecified Type I error would lead to rejecting the null hypothesis of no effect of the predictor.

—Emilia Bagiella

See also Kaplan-Meier Method; Regression; Survival Analysis

Further Readings

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, *34*, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*, 269–276.

Kalbfleish, D. J., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). Hoboken, NJ: Wiley.

Critical Value

Critical values are used in hypothesis testing to delimitate two regions: one in which the analyst rejects the null hypothesis and another in which the hypothesis cannot be rejected. Therefore, critical values are closely linked to the concept of hypothesis testing. For instance, in a two-sided t test to compare two sample means, if the sample sizes are 21 and 21, so that we have 40 *df*, the critical value for a 0.05 level of significance (i.e., a 95% level of confidence) is 2.01 (this value can be found in tables for the Student's t test). If the computed t statistic is, say, 2.32, we reject the null hypothesis—that the two samples correspond to two populations with the same population mean—at the 0.05 level of significance, because the computed t value exceeds the critical value (and that implies that if the null hypothesis were true, the probability of values like those observed or more extreme is below 0.05). However, we cannot reject the null hypothesis at the 0.01 level of significance because for that level of significance the critical value is 2.42, which is above the computed $t = 2.32$.

This example shows that critical values are arbitrary. They are just values corresponding to particular probabilities arbitrarily chosen (usually 0.05 or 0.01) in theoretical statistical distributions. The presentation of confidence intervals is usually preferred to the indication of the fact that the null hypothesis has or has not been rejected because the test statistics was above or below the critical value.

Since hypothesis testing is usually ignored in the Bayesian approach to statistics, the concept of critical value is irrelevant in that theoretical framework.

—José A. Tapia Granados

See also Bayesian Approach to Statistics; Degrees of Freedom; Hypothesis Testing; Null and Alternative Hypotheses; Statistical and Clinical Significance

Further Readings

Goodman, S. N., & Royall, R. (1988). Evidence and scientific research. *American Journal of Public Health*, *115*(3), 235–249.

Moore, D. S. (2004). *The basic practice of statistics* (3rd ed., chaps. 13, 14, 15). New York: W. H. Freeman.

Ware, J. H., Mosteller, F., & Ingelfinger, J. A. (1986). *p* Values. In J. C. Bailar, III, & F. Mosteller (Eds.), *Medical uses of statistics*. Waltham, MA: NEJM Books.

Cross-Sectional Study

See Study Design

Cultural Sensitivity

Cultural sensitivity refers to the process by which health research and intervention respond to the cultural belief systems, behaviors, experiences, and social context of populations. Culturally sensitive approaches acknowledge that conceptualizing determinants of health behaviors and outcomes, conducting research, and creating interventions to reduce disease risks may not rely on universally applied predictors, methods, or settings. In addition, certain constructs may not have the same meaning or may be experienced differently in diverse groups. In epidemiology, cultural sensitivity has relevance to both study design and health promotion.

When conducting research, a culturally sensitive approach is necessary from the beginning of the study design. To appropriately consider the role culture plays in health, local experience and cultural norms must be built into the study methodology. That is, the research question being asked and the extent to which it is appropriately informed by sociocultural factors are critical in conceptualizing the study. For example, if a researcher seeks to investigate the risk factors for experiencing depression and operationalizes depression as a psychological state involving feelings of sadness or being "down," this very definition will artificially restrict the research findings if the study is conducted in a population that expresses depressive symptomatology primarily in physical terms or through other psychological symptoms, such as boredom.

The measures that are used to assess both health outcomes and risk factors must be carefully considered. This can be particularly true for studies in psychiatric epidemiology, where measurement of mental, as opposed to physical, illness presents unique challenges. For example, conditions such as "dependent personality disorder" are themselves strongly imbued with cultural notions of abnormality and adaptive functioning, and these are likely to be built into the tools that researchers use to assess outcomes. Moreover, many standardized scales with established reliability and validity were initially developed with European American populations, and their validity with culturally diverse groups has not been empirically tested.

However, cultural sensitivity also applies to physical outcomes such as heart disease or other conditions that were once thought to be relatively "culture-free." In this regard, researchers must attend closely to the ways in which risk factors are defined and assessed. For example, food frequency questionnaires are often used to measure dietary intake. If the measure is not culturally sensitive, it may not include a variety of foods that are commonly consumed in the target population. If this is the case, the researcher may need to modify the instrument by providing additional spaces to report commonly eaten foods. Ideally, however, the measure would already have been validated in the relevant population and would include any foods typically consumed by the population being studied. Examining associations between stress and illness is another relevant example. When assessing stress, researchers must ensure that the measures are not based on sources of stress that are likely to be irrelevant to the population in question, and they may need to include culturally specific sources of stress, such as racism, discrimination, acculturative stress, or language barriers.

In the research setting, a culturally sensitive approach would also attend to the interaction between researchers and participants. For example, racial/ethnic matching may be necessary, meaning that researchers who interact with participants would be of the same race or ethnic group. This is particularly true for certain topics such as worldview, stigma, identity, and experiences with racism. In these and other contexts, participants' responses can be significantly affected by the interviewer. Informed consent procedures can also be made more culturally sensitive by paying particular attention to participants' level of familiarity with the research enterprise and exploring possible feelings of mistrust. Additionally, research has shown that most consent forms are written at literacy levels that are too high for many participants and contain many technical terms and jargon. Poor readability of consent forms is compounded for those with limited English proficiency.

With regard to interventions, some researchers have conceived of culturally sensitive health promotion programs as informed by surface structure and deep structure. Surface structure refers to matching interventions to certain characteristics of a population. This might include using printed materials that depict individuals from the same population, delivering messages through particular channels (e.g., a local Spanish newspaper) or in particular settings. For example, to reach African American populations, some researchers have implemented nutrition interventions in churches, and others have created printed health materials that draw on spiritual themes and imagery relevant to African Americans. Deep structure refers to the sociohistorical, cultural, and psychological factors that shape health beliefs and behaviors. For example, health promotion programs often strongly emphasize individual choices and behaviors. If such a program is implemented with a population for whom an orientation toward kinship and communal bonds is salient, it will be discordant with the deep structure of the culture and less likely to succeed.

Infusing epidemiologic research and health interventions with culturally sensitive methodologies is best completed in collaboration with community partners and may require an iterative process whereby study goals and methods are refined over time. It is also critical for researchers to carefully think through how culture may apply to particular health considerations in particular populations, rather than relying on stereotypes or broad generalizations. Similarly, particular behaviors that are significantly determined by social and economic conditions should not be interpreted as uniquely representing culture. For example, frequent consumption of high-calorie foods may simply reflect local food resources, rather than cultural preferences. Finally, ''culture'' does not consist of static patterns of behavior, core values, or worldviews. Sociogeographic and historical context influences the ways in which cultural thought and behavior are expressed.

—Naa Oyo A. Kwate

See also Acculturation; Health Behavior; Health Communication; Targeting and Tailoring

Further Readings

Campbell, M. K., Demark-Wahnefried, W., Symons, M., Kalsbeek, W. D., Dodds, J., Cowan, A., et al. (1999). Fruit and vegetable consumption and prevention of cancer: The Black Churches United for Better Health project. *American Journal of Public Health*, *89*, 1390–1396.

Kreuter, M. W., & McClure, S. (2004). The role of culture in health communication. *Annual Review of Public Health*, *25*(20), 439–455.

Resnicow, K., Baranowski, T., Ahluwalia, J. S., & Braithwaite, R. L. (1999). Cultural sensitivity in public health: Defined and demystified. *Ethnicity & Disease*, *9*, 10–21.

Cumulative Incidence

Cumulative incidence provides an estimate of the risk that a person will experience an event or develop a disease during a specified period of time. It is calculated as the number of new events or cases of disease that develop in a population at risk during a specific time interval. Cumulative incidence (also referred to as cumulative incidence or incidence proportion) allows researchers to predict risk of a disease or event over a year, 10 years, or even a lifetime; this time period must always be specified when the cumulative incidence is reported.

One example of cumulative incidence is the risk of developing flu among seniors vaccinated against influenza. Another would be the proportion of passengers who develop gastroenteritis while vacationing on a commercial cruise ship for a week. A third example would be the proportion of patients who develop postoperative complications within 1 month of surgery. Individuals in each of these examples meet both the following criteria: (1) They are free of the outcome (influenza, gastroenteritis, or postoperative complications) at the beginning of the study period, and (2) they have the potential to develop the outcome of interest during the study time period.

In the influenza example above, seniors in a study are vaccinated at the beginning of flu season, before any influenza cases arise in the region. There are two ways for the investigators to define the flu season: as a time period (e.g., November to April) or by a combination of a time period and observed events. For instance, in the United States, the flu season is the time period between the first influenza case in the area and the last influenza case in the area during one continuous period of time between September and June. However, the study period is defined; it is the same for all participants in the study, and they all have the same opportunity to be identified as

infected with influenza if they should contract the disease.

Cumulative incidence is a useful tool to provide information about the risk of an outcome during a prescribed period of time. In studies where a group is followed for a short period of time, it is possible to compute cumulative incidence directly. For studies where longer follow-up periods are needed, such as in cohort studies of diet and the risk of diabetes, it is not usually possible to estimate cumulative incidence directly. Rather, this question is addressed through the computation of incidence rates. However, rates characterize disease incidence for a group, whereas cumulative incidence characterizes the accumulated risk over time.

From a clinical perspective, cumulative incidence is helpful to public health professionals and clinicians because it can personalize the risk of developing a disease or condition over a period of time that is meaningful to the patient. For instance, a pediatrician might describe an overweight child's likelihood of developing type 2 diabetes in the context of the next 10 years, or by adolescence. While cumulative incidence cannot be computed directly in studies with long follow-up periods due to loss to follow-up, it can be estimated in such studies by first calculating the incidence rate and then estimating the cumulative incidence from the rate. In this case, rates should be constant throughout the course of the study, and if they are not, distinct rates must be calculated for discrete time periods and then aggregated to obtain the best estimate of the cumulative incidence.

If the incidence rate *does not vary* over the time period, cumulative incidence can be estimated as follows:

$$\text{Cumulative incidence} = 1 - \hat{e}\{-(\text{incidence rate}) \times (\text{time})\}.$$

If the incidence rate *does vary* during the study period, and incidence rates for discrete time periods are known, cumulative incidence can be estimated as follows:

$$\text{Cumulative incidence} = 1 - \hat{e}\{-\Sigma(\text{incidence rate})_i \times (\text{time})_i\}.$$

—Allison Krug and Louise-Anne McNutt

See also Incidence; Prevalence; Rate

Further Readings

Hennekens, C. H., & Buring, J. E. (1987). Measures of disease frequency. In S. L. Mayrent (Ed.), *Epidemiology in medicine* (pp. 57–66). Boston: Little, Brown.

Schoenbach, V. J. (2004). *Incidence rate and incidence proportion*. Retrieved March 4, 2007, from http://www.epidemiolog.net/studymat/topics/incidence/IncidenceRateIncidenceProportion.doc. Chapel Hill: University of North Carolina at Chapel Hill.

D

Data Management

Data analysts often say that they spend 80% of their time getting the data ready to analyze, and 20%, or less, actually analyzing them. As this emphasis usually is not reflected in courses in epidemiology and public health that include data analysis, an obvious gap is created that tends to be filled only through practical experience. Yet data are an omnipresent part of life in the 21st century. Electronic data are in our pockets, in small databases residing on our memory sticks, and in cell phones and personal data assistants. They are on our desktops, residing in our computers' personal information managers. They are in spreadsheets and relational databases on our desktops or on a computer server.

Though the specifics of the data manager's task vary depending on the software we use, a handful of basic principles are common to all systems. This entry looks at issues regarding existing data, designing new databases, and formulating questions about data. It also discusses object-oriented database structures, database utilization with Web-enabled remote servers, and data security.

Studying Existing Data

Often, we work with data we did not collect ourselves. Even if we were involved in the data collection process, many other people may have had input into both the design and the content of those files. When we receive data files from other hands, we first need to inspect each file carefully with two principal concerns in mind: Are these data correct and undamaged? And what issues exist within these files?

Years of experience have taught the lesson that nothing can be assumed, not even that the correct data file has been provided. We should determine the dates of both file creation and most recent update, plus the numbers of cases and numbers of variables included in the files, and compare these item for item with the specifications (contained in the Manual of Procedures—the "MOP") provided by the persons who provided the data sets. The absence of a codebook should be taken as a warning to inspect the file with particular care. We can learn a great deal by running boxplots, violin plots, and panel plots to rapidly obtain a visual depiction of how each variable is distributed, plus basic descriptive statistics or frequencies for every variable in the entire set of files, and perhaps cross-tabulations for selected variable pairs. All these approaches help locate file transfer problems (e.g., if a variable was imported in the wrong format or is entirely missing); peculiarities within the data, especially those values that are out of the expected range; and entries that are apparently meaningless (e.g., a code for a categorical variable that doesn't translate correctly).

Strategies for uncovering hidden issues about the data set will vary with each situation. However, the following procedures are nearly always appropriate: checking for duplicate cases, locating extreme or out-of-range data, and studying the amount and patterns of missing data. Very few data sets are entirely complete.

Therefore, to best analyze a data file, we need to determine how many data are missing and why they are missing, and identify the patterns of missing data among variables. For purposes of analysis, we are obliged to think carefully about how we treat missing data, because such decisions can substantially affect our results. Cross-tabulations can reveal patterns of missing data involving pairs of variables, and we can write code that reveals the patterns among larger groups of variables. Knowing the amount and pattern of missing data within a file allows us to decide if we should consider some kind of interpolation or other substitution for missing data values. Note, however, that newer statistical modeling procedures often have vastly improved ways of handling missing values, so the consequences of varying missing value substitution algorithms are best addressed with the aid of a knowledgeable statistician.

This is also the time to closely study those variables constructed with all the characters on the keyboard, letters only, or some combination of letters and numbers, to see how many variants occur within each such variable, and to design methods to parse the alphanumeric strings into meaningful blocks or subgroups if necessary. The task of analyzing text statistically has quite different requirements from analyzing data statistically.

Existing data sets may contain millions of cases, and each case may contain thousands of possible pieces of information. Regardless of the dimensions, these files must be opened by software that recognizes the same electronic format (a fact we need to know unambiguously), or they may be distributed in legacy formats (including ASCII, one of the most basic of electronic renderings of characters from the keyboard) and must be imported into our software to be useful. Even when the process is a simple push of a button, we always need to check on elementary concerns, such as the new files equaling the source files in both numbers of cases and names and numbers of variables (truncations can quietly occur for numerous reasons). Some legacy formats do not support the variable-naming conventions that are regularly employed in other formats. Imports from ASCII, in particular, need to be rigorously reviewed to ensure that every variable is extracted from the correct position within the source file.

All these concerns fall under the rubric of "achieving veridicality"—essentially, ensuring that our version of the data set faithfully and completely represents the original information. The next step is to determine the meaning of each variable and its categories, information that should be included in the codebook. The analyst should not have to guess at the meaning of variables from their names, particularly when the prior users relied on acronyms for variable naming, based on what might be usual parlance in one or another professional environment but is not in everyday public language.

When there are repeated measures for the same respondent, as is exceedingly common in medical data bases, the complexity of the data's underlying structure is correspondingly increased. It is not unusual to find, for example, that the database's designers elected to capture a given measure at a fixed frequency (heartbeats per minute), which was then aggregated (average heartbeats per minute per 5-min interval), then repeated across one fixed interval (average heartbeats per minute per 5-min interval for each of 288 such intervals per day), then repeated again across another fixed interval (average heartbeats per minute per 5-min interval per day for each of several days of intensive-care treatment). Increasingly, data are collected with repeats occurring not at regular intervals but instead as a function of trigger events (such as the onset of seizure activity or specific treatment intervention); as a result, any single case may differ substantially from the next case in the frequency of such repeats. Because statistical analysis packages expect data files to be structured in very specific ways, the analyst must understand the meaning of each variable and the structure of the data file they receive, and be able to restructure the file if necessary to make it compatible with the program's expectations.

Designing a Well-Functioning Database

In many circumstances, we can reduce the number of key considerations regarding building a database to just three: how complex our data will be, how we want to get data into the database repository, and how we want to ask questions about those data. It is becoming common, however, to also incorporate further mission-critical elements needed to meet the inherent goals of particular studies, elements that are discussed later in this entry.

Complexity of Data

How complex are our data? The anticipated data for our study may be conceptually simple—only a small number of distinct variables are gathered from

a modest number of respondents. We use the term *respondent* here synonymously with *case, patient, study participant, monitoring location*, or other such source of information. Any one respondent is a single instance of the target sample. The term *variable* is a placeholder for a test item, a physiological signal, a status indicator, an outcome assessment, and any of numerous other markers associated with the respondent. Any one variable is a single instance of such a marker with a capture process that generally applies to every respondent (though it might not).

This data may be arranged in a rectangular database, in which by convention each respondent is assigned a row, and each variable—plus one—gets a column. This assignment is literal when we use flat-field software like a spreadsheet or a two-dimensional database software product, though it may be merely figurative when we use one of the more advanced database software approaches discussed later. We use the extra column to identify the respondent with a unique name or identification number or code. If the initial assumption of design simplicity is indeed appropriate, this rectilinear framework has an intuitive appeal harking back to making lists with paper and pencil. The simple framework has advantages over other, more complex database structures: It can be brought to reality with a minimum of time and personnel resources. Its computing demands can be met at virtually no cost using even old equipment and inexpensive software. Its result is a single file that most likely can be saved electronically in a legacy format that is widely recognized by other software. Everyone wins because costs and complexities are reduced to a minimum at all turns.

However, this is not always the simplest arrangement in practice. Every research study has a context in which to place each respondent—a specific set of information that identifies who or what the respondent is, where the data were gathered, and associated factors (many of which may be risk factors essential for one's planned analyses). A time stamp marking the moment the piece of data was collected is but one of many additional variables that can be critical to the analyst in sorting out possible relationships across variables. An instrument calibration table for each case, likely a table off on its own, can be essential to ensuring comparability across cases. So, even in the most simple appearing of circumstances, our data set will likely have to grow to accommodate entries in additional variables.

Another challenge to our admirable but potentially troubling simplicity is that, as noted earlier, the data to be collected may be multiple instances of the same measurement for each respondent. For example, there might be multiple instances of the same test items, multiple physiological signals of the same type, or multiple types of such signals, all of which might be acquired over several repeated occasions, which might or might not be the same occasions for each variable cluster. As noted earlier, rectilinear solutions do not readily accommodate such multiplicities without some care.

Yet another challenge is that often the data we seek are not instantly translatable from reality to a single numeric or alphabetic entry. An example of this phenomenon is medications taken by a patient at a particular time; such a list may have no entries for one individual but dozens for another. Our rectilinear framework mandates that each instance of a medication be allotted a column. Though we could enter all possible medications in one continuous entry, this will raise serious issues for later querying and analyzing. So in our simple solution, the number of columns must grow to accommodate the individual with the longest list of medications, whoever that person might be; this solution means that for many individuals, most of those fields will be empty. We can pick a "good enough" number and live with it even if respondents show up with longer lists, or we live instead with the fact that the design of the database will need to be modified while in use. Neither option is especially palatable. Other forms of data that translate poorly from reality to discrete numeric or alphabetic entries are narrative material, such as physicians' or nurses' notes; decision trees or routing information, such as schedules; and photographic evidence, such as medical imaging. Fortunately, useful database solutions exist for these situations, discussed later in this entry.

Relational databases can accommodate complexity by using more than one single data frame—for example, discrete sets of information (often called "tables") may be stored for each patient, one containing demographic information, a second set containing diagnoses and procedures, a medication set, a physiology set, an imaging set, a calibration set, and a physicians' and nurses' note set. Each of these may have very different dimensions and may be populated at different times by input from different individuals. Not every patient is likely to be represented in every data set, since not

every data set may apply to every patient (e.g., some patients need no special procedures and are discharged quickly, while others are extensively medicated and stay a long time). Keeping these variable sets distinct is a major step toward management efficiency and significantly improves the task of quality assurance.

Getting Data Into the Database

How do we best get data into our database? In retrospective studies, we need to identify the existing sources of information, find and gain approvals and passwords from the correct persons or corporate and government entities controlling access to those sources, ensure that all human subjects' requirements are met, and identify the combination of personnel who have suitable data transfer or data entry skills along with the place and time where they will be able to accomplish their work. Even if data are already computerized, there is no guarantee whatsoever that files are accessible or, once accessed, are transferable. Many computer systems with fine warehouses of potentially useful data are nearly impenetrable except by in-house systems staff who invariably are shorthanded and overworked. Over the years, archives of valuable data were often stored on media, such as computer tape, which are now unreadable because the necessary tape drives have long been obsolete, or because the media have physically decayed—oxide particles resting in the bottom of a tape canister are an inelegant clue. And retrospective studies, except for those using publicly available data sets without case identifiers, need appropriate institutional approval.

In prospective studies, the tasks differ but their number is equally large, and they must include approvals ensuring that all federal, state, and institutional safeguards are met for human (or animal) subject participation. The studies must also include unambiguous plans for exactly how participants are to be identified, recruited, and enrolled; how and when such enrolled participants will be measured and by exactly what measurement instruments; and how such measures will be captured over time into a working database. Although these issues are subject to wide variation, there are certain consistencies. When designing the database for a study, we usually include tables within the data set that will allow participant tracking from the moment of first contact, even before data are formally obtained. Second, at the moment the first subject is enrolled, there must be a systematic way to capture data into a secure archive—that is, a database of records with dimensions large enough to support the complete study. Database design should not be postponed under any circumstances because doing so might compromise the study's fundamental integrity. Third, as subjects start becoming enrolled, there must be a plan for immediate data quality assurance, since there will be no opportunity to redo the enrollment process. This means that our data collection instruments must be fully up and running: Paper-and-pencil testing must be valid and reliable, electronic instrumentation must be plugged in and delivering quality signals that are of the correct electronic format, and plans about what is to happen if a glitch occurs (a systems fault, a participant refusal, an unplanned or adverse event) must be clear in advance, not invented on the fly if the need arises. Fourth, we need a regular mechanism for verifying and backing up the data that have been captured to date, for there is truly very little merit in conducting a study only to discover afterward that the data sets are unreadable or lost.

Some studies, either retrospective or prospective, can gain immeasurably by advance attention to the database's front-end appearance, its "look and feel." With modern database software, there is no reason not to have a visually appealing form for each conceptually discrete section of the data set. Data entry can be simplified and can also be made more accurate by good screen design. Among the options available in modern database software are radio buttons for binary data, drop-down menus for discrete-choice variables, and cascading sheets for data that are to be entered only contingent on the values of other data (alternate branching). Data that are conceptually clustered together (e.g., all diagnoses, all imaging results, all monitors of a certain type) should be visually clustered as well. Throughout the data entry form, each data item can be accompanied by a floating explanation (a "text balloon") that appears on-screen when our mouse pauses for a predetermined time over the item. On-screen help files can be generated even for small studies, and they are often invaluable in large efforts. Additionally, each data item can be quality controlled so the program will accept data only within certain ranges. What all this costs in setup time is amply rewarded at the time of data entry and data quality review. The ultimate aim is to ensure that mistakes in the data can be easily and rapidly found and corrected if they occur at all.

Though once common due to extreme limits on computer memory and storage, it is now seen as essential to avoid unneeded translation codes in the data entry process (e.g., "1" for male, "2" for female). The reason is that such codes rarely, if ever, add any inherent value, while they certainly obscure the actual meaning of the entry and make it harder to conduct complete and thorough reviews for quality assurance. It is also essential to avoid missing-value indicators that could be genuine values in their own right (e.g., "99" for age). Some missing-value codes can be generated automatically by database software, for instance, to differentiate between data missing because they are not present for that respondent (and will not be forthcoming at any future time); missing because they are contingent on an earlier data element and thus not logically appropriate to complete; or missing because they have yet to be identified, although they are expected to be forthcoming as the respondent completes participation in the study.

Asking Questions

Whether we have acquired an existing database of information or have designed and populated a database of our own, we can think about how best to ask questions of it. Conceptually, querying is the way the data in hand are culled into cohesive and manageable chunks suitable for statistical analyses. It is also a way to readily explore relations across discrete tables within the data set. It becomes a powerful tool, driven by the linking of tables using the respondent identifier as the term in common—a single value that allows a subset of data in one table to be joined with a subset of data from another table within the data set. Properly thought out, the linking of subsets in such queries can be one of the handiest ways to assess well-constructed research hypotheses.

Often, a set of conventions known as SQL (Structured Query Language) is used to query databases; variants of the SQL code are included with many types of software, including relational databases, such as Access, and statistical analysis packages, such as SAS (Statistical Analysis Systems). The elemental query statement in SQL is built from a statement containing three clauses:

SELECT < using a list of attributes >

FROM < a data table or list of tables >

WHERE < one or more conditions must be satisfied > .

The results of a query will be anything inside the data set that satisfies every facet of that query statement. Boolean logic applies to these queries, so AND and OR often are used in specifying the WHERE conditions. Queries can be used to answer questions, such as "how many female subjects have been recruited to date?" and "what were our total lab expenses last month?"

One good way to work with large data sets, especially those that are works in progress, with more respondents or patients or study participants still to supply their data, is to set aside a copy of the data set from some start date to some intermediate checkpoint and to explore it in its own "sandbox"—that is, where we can try testing and querying based on the knowledge we have right now and experiment without interfering with the original data even if we make mistakes. A little time devoted to identifying data patterns in this subset and working out the forms of various queries can pay off handsomely when the full data set is available.

Object-Oriented Databases

Not every data set can be readily envisioned as a series of tables. One illustration often cited is the task of understanding educational systems. Each student will appear in a variety of classes, and each class will have a number of different students. Each teacher will have numerous students and each student will have several teachers, but not every teacher is solely responsible for one class. Each group of students belongs to both a grade and a school, though grades occur across schools and different schools may contain only some of the grades. Another illustration is found in "whole-part" relationships, where we need a database to track parts that used to create end products, where a given product can be made up of many different parts and subassemblies, and where the same part or subassembly can be used in the making of many products. In both instances, many of our measures will have more than one attribute, a situation that is not at all easy to convey inside a simple rectilinear table structure. The relationships between various "respondents" who have complex class affiliations are generally captured poorly, the resulting tables are likely to store data quite inefficiently, and querying them successfully may be difficult.

An object data model allows multivalued attributes to be incorporated into the data structure. Though they

do not lend themselves well to conventional querying, some query languages are built to handle situations that involve such complex and irregular data. If we find ourselves with a data design in mind that relies on complicated data structures, we first investigate how we would try to handle the incoming data using traditional relational methods and then explore the many possible programming and performance solutions that can be achieved by object-oriented (OO) technologies.

The hospital patient record is an example of a data design for which OO technologies might be well suited because of their flexibility and possible ease of use. Seldom, if ever, is the patient record a simple information structure, because an enormous variety of patients, diagnoses, specialists, treatments, and so forth must be accounted for. The OO approach, under the right circumstances, can handle these data in ways that support both data acquisition and querying that are optimized for the task. However, this step should be taken only by those who are well versed in data management with conventional tabular data and who have a distinct need for high performance when faced with complex data. These sentiments apply as well to the set of techniques called data mining, based on technology from advanced statistics, artificial intelligence, and machine learning. Data mining focuses on isolating and confirming relationships deep inside the data set that might not have been the principal focus of the original data design team. While these techniques can be practically applied across a range of data management software settings, they should be approached with caution.

Remote Servers

Highly versatile data management solutions can be developed using the World Wide Web for an interface, adding a level of computing power that would have been unthinkable only a few years ago. With careful design, large databases can be assembled that are responsive to any Web-enabled source (with appropriate security precautions), whether the source is an investigator sitting at his or her desk or someone standing anywhere in the world with a Web-enabled cell phone or wireless handheld device.

What is gained is a uniform look-and-feel for accessing data, making inputs and changes, and developing queries. Whether the data are numeric, alphabetic, or images makes no real conceptual difference, only practical differences in how the screens are best displayed. Also gained from such an implementation is the ability to have multiple research teams that are widely separated from one another be able to use and also update the same data sets in an identical manner. Since all data are stored on one main server, it becomes a central source for data checking and for resolving quality concerns wherever they arise.

Implementations of this concept have been resounding successes. One powerful hospital-based solution uses a single integrated data system, to which its designers simply add various functionalities by adding modules. In many hospital-based studies, the required modules are available off the shelf. In addition, wireless devices can be used to gather data from patients and deliver it to the server in real time. A physician at an international conference can call up an entire set of both records and current monitoring for patients on another continent; examine the most immediate physiological signals, medications, scans, and other professionals' notes as if they were at the bedside; and deliver orders to the hospital staff. In the research context, these capabilities mean that multisite studies can be linked together seamlessly even if separated by long distances.

Security

Good data security is predicated on two different principles that should be followed simultaneously. First, data should be kept in locations that cannot ever be accessed by unauthorized people; and second, if security is breached, it should be impossible to derive any personal information of any kind from the data set. This means it is a bad idea to include personal names or government-issued identification numbers (Social Security numbers, driver's license numbers) within a data set. Indeed, it is difficult to see how Social Security numbers can play an appropriate role in research under any reasonable circumstances. In some studies, medical record numbers may be internally essential for purposes of ongoing data acquisition, but the following three precautions should be implemented at the earliest possible time: a plan for data encryption should be enforced; some other identifier, such as an arbitrary study identification number, should replace the medical record number for purposes of linking; and a thorough expunging of any data values on the formal list of protected health information (PHI) should be conducted before the data are archived and released.

PHI is formally defined as any individually identifiable health information in any form, including verbal communications that is collected or created as a consequence of the provision of health care by a covered entity (e.g., a practitioner, clinic, hospital, health system). All researchers whose studies require one or more pieces of such information from study participants must obtain approval by the appropriate institutional review board, even for retrospective studies. Except for purposes of treatment, investigators are restricted to the minimum PHI considered reasonably necessary to conduct the research. In addition to names and Social Security numbers, the identifiers at issue include dates of birth, death, admission, and discharge (except year); postal addresses, including city, state, and zip code; telephone and fax numbers; medical record numbers; health plan ID numbers; account numbers; certificate or license numbers; vehicle identifiers; device identifiers and serial numbers; URLs, e-mail addresses, and IP addresses; biometric identifiers, including fingerprints; full-face photos and other comparable images; and any other unique identifying number, characteristic, or code.

Additional security burdens are increasingly being placed by funding agencies on data managers to ensure "data set durability." That is, once a transaction has been committed to the database, a complete record of its action, including what it replaced, must be fully preserved even if software or hardware fails partway through the transaction. For instance, if a transaction is interrupted by a system crash, then when the system is restarted, the database will both recover itself in its entirety and conclude the incomplete transaction. A complete transaction history, including who made any given change to the database, is stored in a separate file. Data management and information technology specialists are essential when these requirements are in place.

Databases constructed to be Web accessible must use both closely controlled passwords to allow access and high-level signal encryption to ensure security when data are transmitted from one node to another. The fact that a wireless handheld is capable of linking to enormous quantities of clinical or research-related data does not mean this should ever occur without appropriate security measures. In addition, increasingly portable forms of electronic storage require extra vigilance, because there have been many cases of confidentiality being breached, for instance, because data were taken home and placed on an unsecured computer or stored on a memory stick, which was later stolen or misplaced. It has recently become all too easy to put highly inappropriate information in exceedingly vulnerable places. The woeful plight of individuals who have taken data sets home or committed them to memory sticks for portability and then lost these items to thievery is an urgent reminder of the constant need to consider data security at all levels.

—David L. McArthur and Sarah Boslaugh

See also Biomedical Informatics; Ethics in Human Subjects Research; Missing Data Methods; Relational Database; Spreadsheet

Further Readings

Boslaugh, S. (2005). *An intermediate guide to SPSS programming: Using syntax for data management.* Thousand Oaks, CA: Sage.

Cody, R. (1999). *Cody's data cleaning techniques using SAS software.* Cary, NC: SAS Institute.

Elmasri, R., & Navathe, S. (1994). *Fundamentals of database systems* (2nd ed.). Redwood City, CA: Benjamin/Cummings.

Leiner, F., Gaus, W., Haux, R., & Knaup-Gregori, P. (2003). *Medical data management: A practical guide* (3rd ed.). New York: Springer-Verlag.

Data Transformations

Data transformations modify measured values systematically. Epidemiologists often transform measured values so that they conform more closely to a distribution germane to a statistical method that the epidemiologist would like to apply. For instance, many common statistical techniques assume that a data set consists of approximately normally distributed values, and if this is not the case, a transformation may be applied to the data before analysis. For example, measured data might be used to calculate the heart rate (HR) ratio variate, $HRR = (\text{HR work} - \text{HR rest})/(\text{HR predicted maximum} - \text{HR rest})$, then this data transformed to the new variate $\arcsin(\sqrt{HRR})$. In terms of the matchup between, on the one hand, the statistical methodology applied to study $\arcsin(\sqrt{HRR})$ and, on the other, the assumptions that underlie this methodology, a variate such as $\arcsin(\sqrt{HRR})$ is often a preferred transform of a variate such as *HRR*.

In modern statistical usage, transformations help preprocess raw data prior to the implementation of a general-purpose software package. Were the steps from data input to some display or printing device's output compared with a journey by car through a city, a transformation such as $\arcsin(\sqrt{HRR})$ would play the role of an access road to the software package's freeway on-ramp. Software validity, or, loosely speaking, journey safety, depends on underlying assumptions. Hence, data transformations are usually classified on the basis of types of assumptions. These include a measured variate's *standard* normality, its *general* normality, model linearity, and/or variate homoscedasticity—that is, standard deviation equality regardless of a change of explanatory variates' values. In addition, some useful transformations are not designed to preprocess measurements individually. Instead, once an estimator, such as the sample correlation r, or a test statistic has been computed using raw measurements, these transformations can help enhance the normality of the estimator or test statistic.

Transformations and Simulated Data

Besides the transformation of measured values, transformations are often used in the process of simulating artificial data values. Such procedures make use of notation that is informative when transformation preliminaries to data analyses are illustrated. Of particular importance, by using a pair of uniformly distributed random numbers as input, a Box-Muller Transformation (BMT) generates a pair of independent, *standard* normal variates, in other words, a pair of normal variates with a mean of zero and variance of one. To answer the two questions "Why does the BMT have so many applications?" and "How are transformation components assembled?" it is helpful to call on the following notational conventions. The two Greek letters ϕ and Φ (lower- and uppercase phi, respectively) represent the standard normal density function (i.e., bell curve) and standard normal cumulative distribution function (cdf), respectively. In the same way that $\sin^{-1}$ often designates the arcsine function, Φ^{-1} designates the inverse of Φ.

The function denoted as Phi inverse (Φ^{-1}) (which in older statistical and epidemiological texts is often called the *probit function*) provides a useful notational device. One reason for the inverse of Phi's utility stems from the tendency for transformation and other data analysis steps to be taken in the reverse of the order in which data simulation process components are implemented. For instance, no data analysis text discusses a scale parameter σ (standard deviation) before discussing a location parameter μ (mean). Yet to generate artificial variates, a standard variate—say, for example, Z—is always multiplied by σ first, before μ is added to the resulting product.

Similarly, when a transformation such as $\ln(HRR - \tau)$ is computed, an important type of parameter, a *threshold parameter* τ, is subtracted from *HRR* as a *first* step before computing the logarithm of $(HRR - \tau)$ as a *second* step. However, when such an *HRR* is simulated, the value assigned to τ is added as the *last* step that is needed to create this artificial variate. To create the artificial *HRR* variate's value, a function such as Φ^{-1} is often approximated as a second step. This function is evaluated at the value that the program provides for the random and uniformly distributed, newly conceived, variate Z.

Besides its connection to the log-odds-ratio, the logit function, $\Lambda^{-1}(z) = \ln[z/(1-z)]$, plays an important role in many transformation applications. For example, to transform some newly conceived Z to obtain a logistically distributed variate, all that is needed is to compute $\Lambda^{-1}(Z)$, that is, plug in the program-supplied value of Z for the z of $\ln[z/(1-z)]$. The BMT's unique role stems from the mathematical quirk that, according to the Rosenlicht-Ritt Theorem, unlike the logit function Λ^{-1}, it is impossible to calculate Φ^{-1} as a finite composite of elementary functions, such as x^k, $\exp(x)$, $\sin(x)$, and/or their inverses $\sqrt[k]{y}$, $\ln(y)$, and $\arcsin(y)$. It has therefore been proven that finding some nonapproximate way of obtaining an artificial normal $\Phi^{-1}(Z)$ is, even in our computer-rich era, as futile a quest as trying to square a circle was to an early Greek mathematician whose tools consisted of a straightedge and compass.

While a data analyst has the luxury of working with the well-known elementary function relationship for the standard normal density, $\phi(z) = (1/\sqrt{2\pi})\exp(-z^2/2)$, health data simulation procedures must either make do with approximations of Φ^{-1} or, alternatively, if exact values are preferred, call on the BMT. Nature has placed a stumbling block on what can be thought of as the most straightforward way of obtaining simulated normal variates, the computation of one normal by transforming one Z. However, thanks to the cleverness of the BMT, a pair of normal variates can be obtained by using the pair of elementary functions that were discovered by Box and Muller.

Transformations to Standard Normality

Also based on some approximation of the function Φ^{-1}, in epidemiological studies a transformed *measured* variate's standard normality is often enhanced by the same steps psychologists and educators take to construct a normalized score. A simple example is the rank-based algorithm that underlies Van der Waerden's test. This procedure is implemented by certain packaged computer programs, such as SPSS, and is described in their HELP sections.

To understand how this can provide a transformation of any variate that has a continuous cdf, first consider a curve called the sample cdf $F*$, where this curve is constructed using a size n sample. To the left of all ordered data values $F*$ is zero. Then, at each ordered data value, it raises one step of height $1/(n+1)$ from its previous value. Then, $\Phi^{-1}F*$ (i.e., Φ^{-1} of F^*) is an often-used transformation to standard normality. Specifically, to transform any one of n measurement values, say, x_i, the computer program that implements the transformation first computes $\tilde{z}_i = F*(x_i)$. Then, it evaluates $\Phi^{-1}(\tilde{z}_i)$.

Measurement Transformation to General Normality

Unlike the $\Phi^{-1}F*$ transformation to standard normality, a transformation to general normality is implemented with the help of some choice from among ln, arcsin, square root, or other elementary function, as are other types of transform, such as those of Fisher, Box, and Cox described below. Why the roles of transformations to general normality, on the one hand, and to standard normality, on the other, differ is explained by the conceptualization *marginal*.

Suppose a joint density function $f_1(y, x)$ describes the frequency with which variate Y's value, y, and variate X's value, x, occur together, and conditional density function $f_2(y|x)$ describes the frequency that variate Y's value, y, occurs when X equals x. Then, for $Y =$ HR Work, and $X =$ HR Rest, within the identity $f_1(y,x) = f_2(y|x)f_3(x)$, in other words, the frequency that $Y = y$ (i.e., the variate Y assumes the value y) and $X = x$ occur together, equals the frequency that $Y = y$, given that $X = x$, times the frequency that $X = x$. The function $f_3(x)$ is called the marginal density of X. Similarly, within the identity $f_1(y,x) = f_5(x|y)f_4(y)$, the representation $f_4(y)$ denotes the marginal density of Y.

Many investigations, such as those conducted with the help of analysis of variance (ANOVA), concern the parameters μ and σ within the model of a *marginal* density like the general normal $(1/\sqrt{2\pi}\sigma)\exp(-[(y-\mu)/\sigma]^2/2)$. Other investigations including most, but not all, multivariate analyses target the correlation parameter ρ, or in some other way are concerned with a joint distribution. Transforming to *standard* normality paves over marginal density features characterized by μ and σ that procedures such as ANOVA are designed to study.

The representation F designates a measured variate's cdf and $F(y|x)$ designates the cdf of Y, given that $X = x$. Also used to check putative elementary-function-based transformations, the composite curve $\Phi^{-1}F$ can be estimated by either a $Q-Q$ plot or, alternatively, by a graph of either a sample cdf $F*$ or its cumulative polygon counterpart when graphed on normal probability paper. Given normality, the curve estimated by this graph is a diagonal line, the derivative of which is the horizontal line $y = (1/\sigma)$.

Regression and Multivariate Transformations

Epidemiologists and other scientists often wish to check homoscedasticity, in other words, whether the variability of Y changes as x changes. Homoscedasticity is assumed in common statistical analyses, such as linear regression, based on unweighted least-squares methodology; so if the data do not meet this requirement, it may be transformed as described below. The curve $[\phi\{\Phi^{-1}F(y|x)\}]/f(y|x)$ can help check whether or not homoscedasticity and/or linearity will be better ensured by either use of some log-based transformation or, alternatively, by some other elementary-function-based transformation. It can also help assess whether or not a log or other transformation should be implemented or, alternatively, if no transformation is needed. Specifically, if x_0 and y_0 designate fixed values, that is, constants, when, for any given x_0, the function of y, $[\phi\{\Phi^{-1}F(y|x_0)\}]/\ f(y|x_0)$, is constant valued for all y, it follows that $f(y|x_0)$ must be normal. Given normality, it has also been shown that $[\phi\{\Phi^{-1}F(y_0|x)\}]/f(y_0|x)$'s being constant valued for any y_0 is both a necessary and a sufficient condition for homoscedasticity.

Besides elementary-function-choice correctness, such as between the transformation alternatives log(*HRR*)

and $\arcsin\sqrt{HRR}$, a transformation's value often depends on a less obvious choice. This can be explained by again considering the access road leading to a smooth trip made with the help of a packaged program. An often-encountered pothole is the assumption that the threshold parameter τ within $\ln(HRR - \tau)$ is zero valued. When this assumption is incorrect, a transformed data set will often de-emphasize large values and overemphasize small values. Since small values often are unreliable, this can have disastrous consequences for epidemiological investigations.

Figures 1, 2, and 3 illustrate the consequences of a simplistic assumption that τ is zero valued. Data cases reported to the California Department of Health Services through mid-January 1992 were reviewed. Special attention was paid to transfusion-associated cases for whom, in terms of the 1993 expanded case definition of the acquired immunodeficiency syndrome (AIDS), the date when seroconversion occurred could be determined (Singleton, Tabnak, Kuan, & Rutherford, 1996). A total of 648 men and women who were White, African American, Hispanic, and Asian, all with transfusion-associated AIDS, met the criteria for inclusion in the study. Based on the 293 white males found among the 648 cases, a best-fitting lognormal model is shown in Figure 1, as is a model-free (generalized histogram) frequency curve estimator based on the observed values of the variate, time in days between AIDS-contaminated transfusion and AIDS diagnosis.

By using procedures described in detail by Tarter and Lock (1993, chaps. 6 and 7) and elaborated on by Tarter and Hill (1997), it was estimated that an appropriate value to be added to variate values prior to log transformation was 3,352 days. Figure 2 illustrates use of this value. Only in the right tail is there any appreciable difference between a fitted normal and the generalized histogram. However, Figure 3 illustrates the consequences of failing to add this value prior to log transformation, that is, the curve based on the variate ln(Progression), with τ assumed to equal zero, produces an unwanted bubble over the fitted normal's left tail due to the overly spread short progression time.

In HIV/AIDS research, cancer epidemiology, and many other types of studies, lag-time considerations—such as incubation period, study start-up, and end-phase issues—tend to reduce the validity of small observed values. This sort of background noise contamination, and the way that it can distort log-observed and log-expected frequencies seriously, is illustrated in Tarter (2000, sec. 12.1). For example, data obtained from large census tracts are, overall, more informative than its small census tract counterparts. Suppose a graph of points, each corresponding to an individual census tract, describes log-observed and log-expected cancer frequencies. Then, it would be unfortunate if a transformation procedure tends to increase the deviations from a fitted regression curve over small expected values. This would have the effect of emphasizing tracts that, possibly due to the small size of their at-risk populations, have few expected cases.

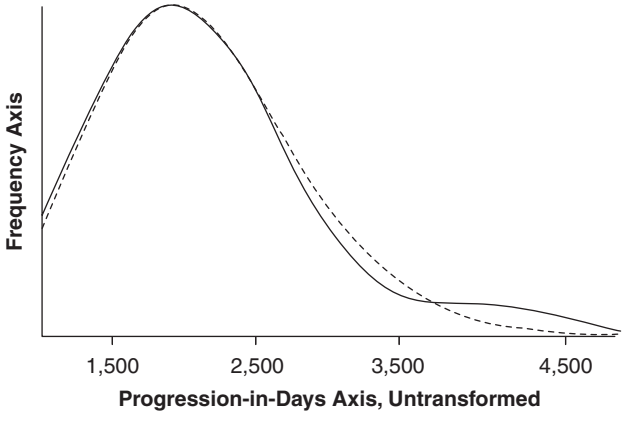


Figure 1 Model-Free Frequency Curve Estimator Based on Pretransformed Progression to AIDS Data Together With Best-Fitting Lognormal Model

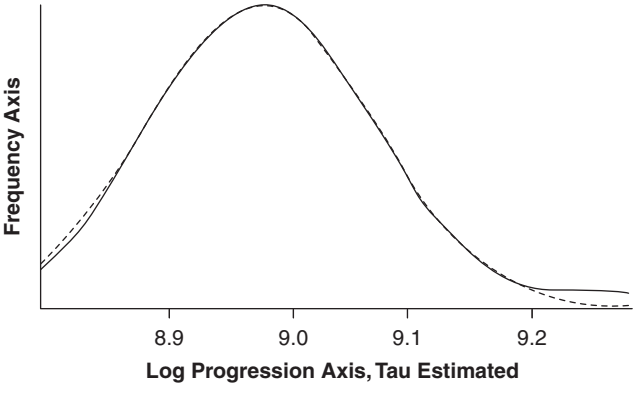


Figure 2 Model-Free Frequency Curve Estimator Based on Transformed Progression to AIDS Data (Based on Appropriate Estimated τ) Together With Best-Fitting Normal Model

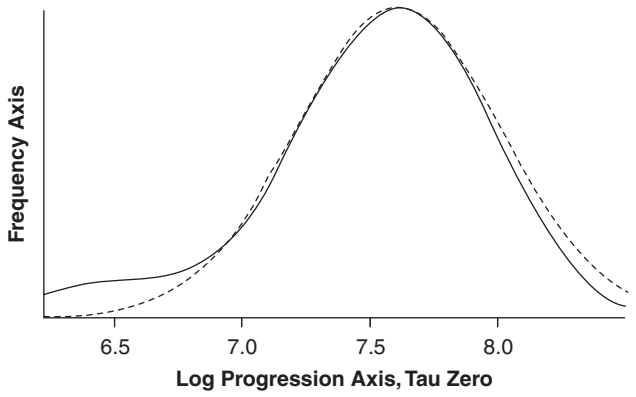


Figure 3 Model-Free Frequency Curve Estimator Based on Transformed Progression to AIDS Data (Based $\tau = 0$) Together With Best-Fitting Normal Model

Background noise versus relevant information limitations are often encountered in animal experiments. In many such studies, early deaths are due to extraneous, nonintervention-related factors. In other environmental as well as in many epidemiological investigations based on observational data, large observed values also tend to contain a noise component. This tends to either be, in effect, a call for attention (such as the need to consider the possibility that outliers are present within a data set) or, alternatively, of little importance in comparison with relevant information. Consequently, because it emphasizes small, not large, observed values, the bubble shown in Figure 3 overwhelms the scientifically informative portion of the observed sample values and has the effect of enhancing the background noise's contribution to any message that Mother Nature has seen fit to send us.

Given that the value assumed by τ is appropriate, the Box-Cox transformation is often useful. It can be applied in parametric regression applications where Y is an observed response variate, such as *HRR*, and the expected value of transformed variate $Y^{(\lambda)}$ is modeled as a linear combination of explanatory variables. For τ less than the smallest observed Y, when $\lambda \neq 0$, the function of Y, $Y^{(\lambda,\tau)}$, is defined to equal $[(Y-\tau)^{\lambda}-1]/\lambda$ and, when $\lambda = 0, t = \ln(Y-\tau)^{\lambda}$.

Fisher's Transformation

For a sample of observed binomial proportions of X outcomes out of N possible outcomes, instead of the variate X/N, a common data transformation is $\arcsin\sqrt{X/(N+1)} + \arcsin\sqrt{(X+1)/(N+1)}$. Correspondingly, for a Poisson variate X whose mean exceeds 0.8, $\sqrt{X} + \sqrt{(X+1)}$ is usually recommended. In more complex applications, such as those involving the sample correlation coefficient r, *Fisher's transformation* is often used. For example, suppose that a Pearson correlation r is computed from $\arcsin(\sqrt{HRR})$ transformed explanatory measurements and corresponding worker productivity measurements. Although it is hoped that both the composite, Fisher's z, as well as its n $\arcsin(\sqrt{HRR})$ components can all be validly assumed to have normal distributions, unlike each of the n $\arcsin(\sqrt{HRR})$ transformed *measurements*, z is called a *transformed estimator* or *transformed test statistic*. Note that a lowercase nonitalicized letter is used here to distinguish *Fisher's* z from a *Z score* such as that provided by the transformation $\Phi^{-1}F*$, or an argument such as Z substituted for the z of $\ln[z/(1-z)]$ to transform Z to a standard logistic variate.

—Michael E. Tarter

See also Analysis of Variance; Histogram; Measures of Association; Normal Distribution; Regression; *Z* Score

Further Readings

Box, G. E. P., & Muller, M. E. (1958). A note on the generation of normal deviates. *Annals of Mathematical Statistics, 28*, 610–611.

Hoyle, M. H. (1973). Transformations: An introduction and a bibliography. *International Statistical Review*, *41*, 203–223.

Rosenlicht, M. (1975). Differential extension fields of exponential type. *Pacific Journal of Mathematics, 57*, 289–300.

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.

Selvin, S. (2004). *Biostatistics: How it works*. Upper Saddle River, NJ: Pearson.

Shimizu, K., & Crow, E. L. (1988). History, genesis, and properties. In E. L. Crow & K. Shimizo (Eds.), *Lognormal distributions* (pp. 1–26). New York: Dekker.

Singleton, J. A., Tabnak, F., Kuan, J., & Rutherford, G. W. (1996). Human immunodeficiency virus disease in California. Effects of the 1993 expanded case definition of the acquired immunodeficiency syndrome. *Western Journal of Medicine*, *164*(2), 122–129.

Tarter, M. E. (2000). Continuous variate adjustment. In *Statistical curves and parameters: Choosing an*

appropriate approach (pp. 175–188). Natick, MA: A K Peters.
Tarter, M. E., & Hill, M. D. (1997). On thresholds and environmental curve tensiometers. *Ecological and Environmental Statistics, 4*, 283–299.
Tarter, M. E., & Lock, M. D. (1993). Curve estimation approaches to model and transformation selection. In *Model-free curve estimation* (pp. 121–136). New York: Chapman & Hall.
Tarter, M. E., & Lock, M. D. (1993). Threshold parameter and transformation applications (pp. 137–166). In *Model-free curve estimation.* New York: Chapman & Hall.

Death Certificate

Most industrialized countries have required certification of deaths for years, and death certificates are a primary source of mortality data in many countries. In such countries, death certificates are the official source of information about deceased persons, the date and location of their death, and the causes of their death. In the United States, mortality data are collected in local jurisdictions and reported using a Standard Certificate of Death and model procedures developed cooperatively by the National Center for Health Statistics (NCHS) and local jurisdictions, the latter being the 50 states, two cities (New York City and Washington, D.C.), and five territories (Puerto Rico, the U.S. Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands). The NCHS compiles and releases aggregated data drawn from death certificates, but requests for a copy of the death certificate for an individual must be directed to the local jurisdiction. Release of death certificate information for an individual may be restricted to relatives and persons with a documented medical need or lawful right to the data.

The Standard Certificate of Death for the United States was most recently revised in 2003. Death certificate information customarily supplied by the medical certifier (typically a physician, medical examiner, or coroner) includes the date and time the person was pronounced dead; the immediate cause of death and up to three conditions leading to the immediate cause; other conditions contributing significantly to death; whether an autopsy was performed and whether the autopsy findings were available to complete the death certificate; if tobacco use contributed to the death; if the manner of death was natural; an accident, a suicide, a homicide, pending investigation or could not be determined, and if accidental, more details about the nature of the accident.

Information on the Standard Certificate of Death typically completed or confirmed by the funeral director includes decedent's name, sex, Social Security number, age, birthplace, residence, marital status, military services, name and address of parents and spouse, place of death, method and place of disposition (burial, cremation, donation, etc.), education, race, ethnicity, usual occupation, and kind of business or industry in which they were employed.

Information from death certificates is widely used in public health reporting, surveillance, and epidemiological research because the data are collected in a standard format and are available, at least theoretically, for every person who dies in the United States. However, death certificate information must be interpreted with care. Some researchers, such as Michael S. Lauer and colleagues, have strongly criticized the usefulness of cause-of-death information drawn from death certificates: They point out that even under ideal conditions, it is often difficult to determine the cause of death in the presence of comorbid illnesses and without information drawn from an autopsy and that the death certificate is often completed by a physician who is unfamiliar with the deceased person's medical history. A more moderate view regards cause-of-death information drawn from death certificates to be accurate for some diseases such as specific cancers but questionable in other cases, so the usefulness of this type of information depends on the particular disease or risk factor being studied. The intended use of the information should also be considered: For instance, death certificate information may be adequate for routine surveillance on many causes of death but completely inadequate for research on many specific causes.

—Sarah Boslaugh

See also Farr, William; Graunt, John; Morbidity and Mortality Weekly Reports; Mortality Rates; National Death Index; National Mortality Followback Survey; Public Health Surveillance

Further Readings

Lauer, M. S., Blackstone, E. H., Young, J. B., & Topol, E. J. (1999). Cause of death in clinical research: Time for

a reassessment? *Journal of the American College of Cardiology, 24*, 618–620.

Moyer, L. A., Boyle, C. A., & Pollock, D. A. (1989). Validity of death certificates for injury-related causes of death. *American Journal of Epidemiology*, *130*(5), 1024–1032.

National Center for Health Statistics. *Mortality data from the national vital statistics system*. Retrieved June 13, 2007, from http://www.cdc.gov/nchs/about/major/dvs/desc.htm.

Decision Analysis

Decision analysis is the art and science of informing difficult decisions. It has a long and varied history, with roots in economics, business, psychology, engineering, and other fields. It is inherently multidisciplinary, especially with regard to analyses that involve the health of individuals or populations. Decision analysis can be used to inform clinical, health funding, or policy decisions.

The basic steps in decision analysis are universal to most rational and systematic decision-making processes. Briefly, a problem is defined, including the decision situation and context. Objectives, based on what the different stakeholders (participants in the decision) value or deem important, are defined and quantitative measures or scales (i.e., ''attributes'') are determined. Alternative choices are defined. The problem is then modeled, using ''expected value'' methods (described later), and the alternatives are ranked in terms of how well they satisfy the objectives. Sensitivity analyses are performed to examine the impact of uncertainties, and the need for further analysis or refinement is determined. A framework that is often cited is Hammond, Keeney, and Raiffa's (1998) PrOACT, which stands for problems, objectives, alternatives, consequences, and trade-offs. An important aspect of decision analysis frameworks is that they are iterative in nature and function as decision support tools rather than the ''final answer.''

In health care and public health, most decisions revolve around improving survival, health state, and/or quality of life. Thus, in decision analyses involving health outcomes, an important consequence measure is typically some measure of health status, such as mortality, morbidity, or a combined measure, such as quality-adjusted life years (QALYs). These measures are assumed to represent utilities or measures of preference; that is, an alternative that improves, say, survival over another alternative (all other things being equal) will be preferred. There is an extensive literature on this subject, including pros and cons of different measures. Good summaries are provided by Brent (2003) and Drummond and McGuire (2001). As health interventions or policies nearly always involve resource limitations, cost per unit utility gained or ''cost-effectiveness'' is often used as a decision criterion. Less common are cost-benefit analyses, in which all consequences, including health status or survival, are measured in monetary terms. Health decision analyses are sometimes incorrectly viewed as synonymous with economic evaluations. Indeed, it is informative to address multiple objectives important to multiple stakeholders in what is termed *a multi-attribute decision analysis*, but to date these applications have been limited in the health care field, although multi-attribute analyses have been applied in the public health field.

The analytical aspects of decision analysis center on estimation of the ''expected value'' of different alternatives. The expected value of an alternative is a function of the probability-weighted consequence(s). This is typically estimated using a decision tree or influence diagram. As an example, a decision may be whether one should choose intervention or Treatment *a*, *b*, or *c* to reduce mortality from disease *X*.

Say that Treatment *a* represents ''watchful waiting'' or even doing nothing. Each ''branch'' of the tree represents the probability that the patient will live or die, given a particular treatment. In the case of the present example, say that a utility of 1 is assigned to life, and a utility of 0 is assigned to death. If Treatment *b* increases the probability of survival by a greater degree than *a* or *c*, then its probability-weighted consequence will be larger, and thus it will be the preferred alternative under the axioms of utility theory.

Whether a decision analysis is performed using single utility measures or a multi-attribute summary function, the result is a set of expected-value estimates for all alternatives evaluated, which allows a ranking; that is, the higher the expected value, the more preferred the alternative. If a ratio measure such as cost-effectiveness is used, the interpretation of a decision analysis is more complex. The concept of ''dominance'' is used in economics to account for the bivariate nature of cost-effectiveness. In the previous example, if Alternatives *a* and *b* are both more effective and less costly than *c*, then *c* would never be preferred; thus, Alternative *c* is dominated by *a* and *b*.

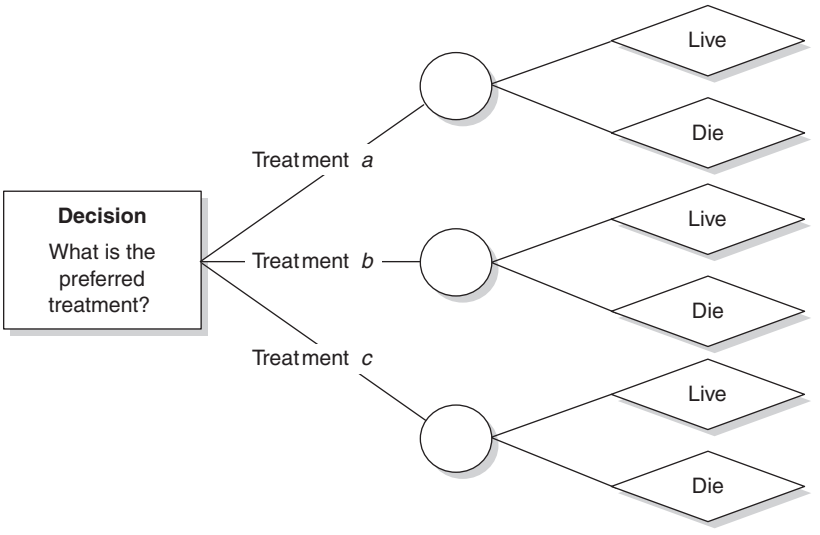


Figure 1 Example of a Simple Decision Tree

Note: In typical representation, a rectangle represents a decision, a circle a probability, and a diamond a consequence.

The appropriate comparator in this case is the marginal or incremental cost-effectiveness of, say, a new drug or procedure compared with others, as cost-effectiveness is a ratio and the absolute value is of little interest.

Most decision analyses are complex, and there are many methodological refinements. In the health field, accounting for variability across individuals in a population, uncertainty in parameter estimates, and temporal changes over time are often important. As in epidemiology, population heterogeneity can be influential on the results of analyses, thus stratification and modeling at an individual level (microsimulation) have proven informative. Health state probabilities change over time (an example is the probability, say, more than 30 years, of developing cancer), thus modeling methods that allow for this, such as Markov methods, are now routinely applied in health decision analysis. Hunink et al. (2001) provide a good summary of these methods.

Although decision analyses in themselves account for uncertainty with regard to expected consequences, they do not address uncertainties in the probability estimates or other inputs themselves. An important adjunct to decision analysis is the ability, known as sensitivity analysis, to examine the impact of parameter uncertainty on decisions. The simplest way to evaluate uncertainty associated with the information used in a decision analysis is to adjust the input values one at a time or simultaneously to examine how the results change. A more powerful way to do this is to use methods such as Monte Carlo simulation to adjust the inputs over ranges or distributions of values and then to use a statistical means, such as rank correlation, to determine the most influential variables. When uncertain variables, such as the predictive value of diagnostic tests, are of interest, a measure known as the expected value of imperfect information is highly informative, because it allows the decision maker to evaluate the impact of using tests to refine judgments. In these analyses, a test that has a higher positive predictive value, for example, will have more "value" in terms of making an informed decision than a test with a lower positive predictive value. Influence diagrams, which are a more efficient way of performing decision analyses than decision trees, are particularly useful in estimating the expected value of information. Clemen and Reilly (2001) provide further explanations of these methods.

Decision analysis intersects with epidemiology in several different ways. In a broad sense, epidemiological studies (both controlled and observational) provide useful information for evaluating the benefits and risks of a drug, device, procedure, or program. One of the more obvious is in the context of randomized clinical trials (RCTs) for drugs, devices, procedures, and

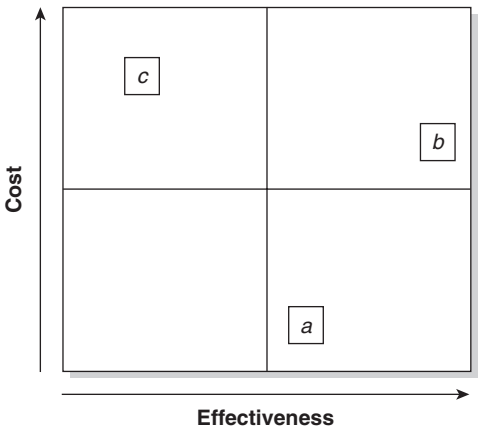


Figure 2 The Results of a Cost-Effectiveness Analysis

Note: Treatment *c* is dominated by Treatments *a* and *b*, and thus *c* would not be a preferred choice. The decision as to whether *a* or *b* is preferred may depend on available resources or other factors not accounted for in this particular analysis.

the like. As many regulatory approval agencies use cost-effectiveness as a decision criterion, a natural extension of an RCT is a decision analysis, in which the new drug or technology is compared with others on the market. Epidemiological studies can also provide useful natural history information for diseases, which can inform detailed decision analyses. For example, Markov decision analyses use transition probabilities between disease states to examine changing health outcomes and costs associated with the disease over time and how these can change with different treatments. This is particularly important in diseases such as cancer, which can last many years.

Decision analysis has many of the same strengths and weaknesses as epidemiology. The strengths relate to robust methodologies that when applied to good information give informative results. Conversely, if the decision analysis is poorly framed and conducted, and applied to biased data, then the results may be worse than meaningless. Decision analysis is fundamentally different from epidemiology because it not only is descriptive but also can be normative; that is, it can tell us both what is and what we should do. A poorly conducted epidemiological study may simply create controversy, but a poorly conducted decision analysis can affect policy on a national scale and indeed affect the lives of the public. For this reason, several organizations publish guidelines for health decision analyses; for example, see Canadian Agency for Drugs and Technologies in Health (2006) and Weinstein et al. (2003).

—Robert C. Lee

See also Economic Evaluation; Ethics in Health Care; Ethics in Public Health

Further Readings

Brent, R. (2003). *Cost-benefit analysis and health care evaluations*. Cheltenham, UK: Edward Elgar.

Canadian Agency for Drugs and Technologies in Health. (2006). *Guidelines for the economic evaluation of health technologies: Canada.* Ottawa, Ontario, Canada: Author. Retrieved from http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf.

Clemen, R., & Reilly, T. (2001). *Making hard decisions*. Pacific Grove, CA: Duxbury Press.

Drummond, M., & McGuire, A. (2001). *Economic evaluation in health care: Merging theory with practice*. Oxford, UK: Oxford University Press.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1998). *Smart choices: A practical guide to making better decisions*. Boston: Harvard Business School Press.

Hunink, M., Glasziou, P., Siegel, J., Weeks, J., Pliskin, J., Elstein, A., et al. (2001). *Decision making in health and medicine: Integrating evidence and values*. Cambridge, UK: Cambridge University Press.

Lee, R. C., Donaldson, C., & Cook, L. S. (2003). The need for evolution in health care decision modeling. *Medical Care*, *41*(9), 1024–1033.

Parmigiani, G. (2002). *Modeling in medical decision making: A Bayesian approach*. Chichester, UK: Wiley.

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty*. Reading, MA: Addison-Wesley.

Weinstein, M., Fineberg, H., Elstein, A., Frazier, H., Neuhauser, D., Neutra, R., et al. (1980). *Clinical decision analysis*. Philadelphia: W. B. Saunders.

Weinstein, M., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C., et al. (2003). Principles of good practice for decision analytic modeling in health-care evaluation. *Value in Health*, *6*(1), 9–17.

Degrees of Freedom

A good understanding of degrees of freedom (*df*) is important in statistics, but most statistics textbooks do not really explain what it means. In most cases, degrees of freedom are thought of as a parameter used to define statistical distributions and conduct

hypothesis tests. For instance, the sampling distribution for the t statistic is a continuous distribution called the t distribution. The shape of the t distribution depends on one parameter, the degrees of freedom. In a sample of size n, the t distribution has $n-1$ *df*.

Degrees of freedom can be thought of in other ways also. The degrees of freedom indicate the number of independent pieces of information that are allowed to vary in a system. A simple example is given by imagining a four-legged table. When three of the legs are free to be any length, the fourth leg must be a specified length if the table is to stand steadily on the floor. Thus, the degrees of freedom for the table legs are three. Another example involves dividing a sample of n observations into k groups. When $k-1$ cell counts are generated, the kth cell count is determined by the total number of observations. Therefore, there are $k-1$ *df* in this design.

Generally, every time a statistic is estimated, 1 *df* is lost. A sample of n observations has n *df*. A statistic calculated from that sample, such as the mean, also has n *df*. The sample variance is given by the following equation:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

where $\bar{x}$ is the sample mean. The degrees of freedom for the sample variance are $n-1$, because the number of independent pieces of information in the system that are allowed to vary is restricted. Since the sample mean is a fixed value—it cannot vary—1 *df* is lost. Another reason there are $n-1$ degrees of freedom is that the sample variance is restricted by the condition that the sum of errors $(\sum_{i=1}^{n}(x_i-\bar{x}))$ is zero. When m linear functions of the sample data are held constant, there are $n-m$ *df*.

We can look at degrees of freedom another way by referring to simple regression for an example. Often, we want to compare results from a regression model (the full model) with another model that includes fewer parameters and, therefore, has fewer degrees of freedom (the reduced model). The difference in degrees of freedom between the full and reduced models is the number of estimated parameters in the full model, $p(f)$, minus the number of estimated parameters in the reduced model, $p(r)$. The full regression model is

$$Y_i=\beta_0+\beta_1X_i+e_i.$$

There are two parameters to be estimated in this model, $p(f)=2$. The reduced model is constructed based on the null hypothesis that β_1 is equal to zero. Therefore, the reduced model is

$$Y_i=\beta_0+e_i.$$

There is only one parameter to be estimated in this model, $p(r)=1$. This means that there is $p(f)-p(r)=1$ piece of information that can be used for estimating the value of the full model over the reduced model. A test statistic used to compare the two models (the F-change statistic, for instance) will have 1 *df*.

Of course, these are very simple illustrations. In some cases, the degrees of freedom must take into account sample size and number of parameters together. For example, in simple regression there are $n-p(f)-1$ *df* for estimating error.

—Mary Earick Godby

See also Inferential and Descriptive Statistics; Regression; Study Design

Further Readings

Galfo, A. J. (1985). Teaching degrees of freedom as a concept in inferential statistics: An elementary approach. *School Science and Mathematics*, *85*(3), 240–247.

Good, I. J. (1973). What are degrees of freedom? *American Statistician, 27*, 227–228.

Rawlings, J. O. (1988). *Applied regression analysis: A research tool*. Pacific Grove, CA: Brooks/Cole.

Toothaker, L. E., & Miller, L. (1996). *Introductory statistics for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole.

Demography

Demography is the study of how populations are structured and change due to the interplay of births, deaths, and migration. In narrow terms, "formal demography" refers to the scientific study of human populations with a focus on their size, structure, distribution, and development. Defined more broadly as "population studies," demography also studies the causes and consequences of population compositions and changes, and draws on neighboring disciplines, such as sociology, economics, anthropology, and epidemiology. Scholars working in this tradition can be designated as social, economic, cultural, or health

demographers, underscoring the field's multidisciplinary nature. There is increasing overlap between the concepts and methodologies of demography and epidemiology. However, separate histories, missions, professional discourses, and cultures have hindered dialogue between the two fields.

Just as epidemiology can be traced to John Snow's account of cholera in 19th-century London, so the origin of formal demography lies in John Graunt's pioneering 1662 analysis of London's "Bills of Mortality." This yielded a crude life table, revealed much about demographic changes occurring within British society, and drew attention to the need for population statistics in public administration. The origins of population studies can be traced to Thomas Malthus, who published his first essay on the "principle of population" in 1798. His thesis, that population growth threatened prosperity because it ultimately outran increases in food supplies, stimulated interest in the relationships between population and resources.

Demography has since built a rich empirical research tradition and a substantial body of conceptual knowledge. In the 20th century, its dominant theoretical preoccupation has been with demographic transition theory. This refers to the movement of death and birth rates in a society from a pretransitional stage, where both are high, to a posttransitional stage, where both are low. In between is the demographic transition itself, a period of rapid and substantial population growth due to births exceeding deaths, as characterized by most developing countries in the latter half of the 20th century. Birth and death rates eventually converge as the transition nears an end, resulting in little growth through natural increase (the excess of births over deaths). Although largely a set of generalizations from observed trends with limited explanatory and predictive power, demographic transition theory has proved a useful framework for describing and comparing population change over time and space. It also gave rise to both epidemiologic and health transition theory.

In the late 1960s, the problem of world population growth dominated demographic thought. The world's population growth rate peaked at 2% in the late 1960s, equivalent to a doubling time of 34 years (i.e., at a 2% rate of growth, the population would double in 34 years). Birth rates have since fallen nearly everywhere, but there remains a clear distinction in both population growth rate and age structure between the industrialized and affluent countries of Europe, North America, Australia, New Zealand, and Japan, and many nonindustrialized countries in Central and South America, Africa, and Asia. The latter continue to exhibit relatively high population growth rates and youthful age structures, meaning a high proportion of their population is less than 18 years of age. In contrast, many industrialized countries have a stable or negative growth rate and aging populations. The United States has a positive growth rate due mostly to its sustained high rates of immigration, with immigrants primarily younger than the native born and more likely to have children. Successive cohorts of small family size eventually lead to population decline because to maintain a stable population size, an appreciable proportion of couples must have families larger than two children to counterbalance those unmarried, married but sterile, voluntarily childless, and one-child families. This is the situation facing much of Europe and East Asia today. In 2005, fertility had reached below the replacement level of 2.1 children per woman in 44 developed countries. In 15 countries fertility had fallen to below 1.3, levels unprecedented in human history.

Demographic analysis is used in both the public and private sectors to help solve planning and management problems. Examples include projecting and analyzing future population composition, delineating legislative districts, marketing new products, and planning new services. Formal demography provides a set of techniques for describing, summarizing, and manipulating data collected in censuses, surveys, and vital registration systems. Techniques such as indirect estimation, intercensal cohort comparisons, and model life tables are applied alongside an array of statistical methods and, increasingly, qualitative methods. Demographers have provided and analyzed much of the factual information about American and other societies, showing how life changes vary by gender, age, socioeconomic status, ethnicity, race, geographic, and family origin. The release of findings from the 2000 U.S. Census through the Internet made demographic information more accessible to the public and policymakers alike.

The main substantive focus of academic demography has shifted, as the dominating concern is no longer world population growth. The toll of the HIV/AIDS pandemic, especially in sub-Saharan Africa and parts of Asia, has reinvigorated interest in studying mortality and morbidity. Demographers study a diverse set of topics, using surveys, such as the U.S. National Survey of Family Growth, to evaluate reproductive

patterns, fertility-regulation trends, or changing family structures. Issues of immigration and aging have assumed a larger role in the discipline.

Health demography has become an important subfield due to the importance of health as both cause and consequence of demographic and socioeconomic changes. More health surveys are being conducted, and more researchers are interested in the close relationship between health and changes in mortality. Increases in the number and proportion of elderly persons, and in per capita spending on their behalf, also receive much attention due to their influence on rising health care costs. Health demographers have extended the concept of conventional life tables to include measures of "active" or "healthy" life expectancy.

Demographers have become increasingly interested in health transitions and trajectories, while epidemiologists have become more concerned with population-level disease dynamics. This is evident in research on aging, health disparities, poverty and health, reproductive epidemiology, sexual risk behavior, sexually transmitted diseases, and other topics. Demography and epidemiology have converged in many ways even if practitioners in the two fields may not be aware of this development.

—Andrzej Kulczycki

See also Fertility, Measures of; Life Expectancy; Life Tables; Mortality Rates

Further Readings

Farley, R., & Haaga, J. (Eds.). (2005). *The American people: Census 2000.* New York: Russell Sage Foundation.

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The methods and materials of demography.* London: Elsevier.

United Nations. (2005). *World population prospects: The 2004 revision.* New York: Author.

Weinstein, M., Hermalin, A. I., & Stoto, M. A. (Eds.). (2002). *Population health and aging: Strengthening the dialogue between epidemiology and demography.* New York: New York Academy of Sciences.

Dependent and Independent Variables

One of the fundamental distinctions in medical and epidemiological research is that between *independent* and *dependent* variables. In the simplest sense, the *dependent* variable is the result or outcome being studied and the *independent* variables are factors that are assumed to exert an influence on it. These basic concepts are very simple but can become confusing in practice, particularly since different researchers use different terms for the same concept. In addition, because some researchers believe that the terms *independent* and *dependent* variable imply a causal relationship, they prefer to use one set of terms for experimental research (where it is possible to hypothesize causality) and another for observational research (where due to the number of uncontrolled influences on the outcome, some researchers prefer to speak of correlations or other observed relationships without labeling them as causal).

In experimental research, if a variable under investigation is to be described in terms of other variables, then this variable is called a *dependent variable* (or *response variable* or *outcome variable*). A variable that is believed to influence a dependent variable is called an *independent variable* (or *explanatory variable*, or *causal variable*, or *input variable*). When data are displayed graphically, the dependent variable is generally represented on the *y*-axis and the independent variable(s) on the *x*-axis.

In a standard regression equation of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots,$$

the Y-variable is the *dependent variable* and the X-variables are the *independent variables.* Studies often include numerous independent variables and may also include multiple dependent variables.

To take an experimental example, a researcher randomly assigns subjects to receive one of five different oral contraceptives to study their effect on high-density lipocholesterol (HDLC), a substance found in blood serum. It is believed that high levels of this substance help delay the onset of certain heart diseases. In this case, the five oral contraceptives are independent variables, and the HDLC is the dependent variable. Other independent variables might be included in the study, including the participants' age, weight, and previous medical history.

The very language of independent and dependent variables implies causality (the value of the dependent variables is assumed to *depend* in some way on the value of the independent variables), and researchers do sometimes refer to independent variables as *causes* of change in the value of the dependent variable and

changes in the dependent variable as *effects* caused by the independent variables. However, particularly in observational studies, causality cannot be assumed simply because a relationship exists between two variables, and for this reason, some researchers prefer to use the terms *predictor* variable for independent variable and *criterion* variable for dependent or outcome variable in nonexperimental research. The reason is that one variable can *predict* the value of another in the sense that there is an observable relationship or association between the two, without implying the *causal* relationship that exists between them. For instance, in the United States, race is *associated* with poorer outcomes on a number of health measurements, but that fact is not taken as proof that race is the *cause* of the poorer outcomes.

—Renjin Tu

See also Causal Diagrams; Causation and Causal Inference; Study Design

Further Readings

Abraham, B., & Ledolter, J. (2006). *Introduction to regression modeling.* Belmont, CA: Thomson Brooks.
Nicholas, P. J., & Nicholas, P. (2004). *Statistics for epidemiology.* Boca Raton, FL: Chapman & Hall.

Descriptive and Analytic Epidemiology

Descriptive and analytic studies are the two main types of research design used in epidemiology for describing the distribution of disease incidence and prevalence, for studying exposure-disease association, and for identifying disease prevention strategies. Generally speaking, descriptive studies deal with the "*what*" questions, for example, describing "what happened" in terms of disease occurrence, while analytic studies ask the "*why*" questions, for example, why some people develop disease and others don't.

Descriptive Studies

Descriptive studies are designed to describe data on health outcomes, such as disease incidence, prevalence, and mortality according to three variables: person, place, and time. *Person* variables describe the people who develop disease in terms of their personal characteristics, such as age, gender, race, marital status, blood type, immune status, occupation, socioeconomic status, and so on, and where and when they were exposed to the agent causing the disease. *Place* variables may include any or all three sites: where the individual was when disease occurred; where the individual was when he or she became infected from the source; and where the source became infected with the etiologic agent. *Time* variables may refer to duration, age, calendar time, birth cohort, or time trend (secular trend, period trend, and seasonal trend). All these belong to *what* questions, because they tell us what happened. Graphical methods and descriptive statistics are commonly employed in descriptive studies.

Descriptive studies can be used to generate hypotheses about the causal association between exposure and outcome. For example, for a given disease, detailed plotting of the duration from exposure to death (time variable) against age and sex (person variable) may prompt generation of the hypothesis of higher mortality being associated with older males. The putative association or causal relationship may then be confirmed or refuted by testing this hypothesis using an analytic study design such as a prospective cohort study. A well-known historical example of a hypothesis-generating descriptive study is the exploratory analysis of the mortality rates from stomach cancer and from colon cancer of ethnic Japanese living in Japan and California. A comparison of cancer mortality rates of Japanese living in these two localities and between first- and second-generation Japanese immigrants in California and Hawaii suggested the hypothesis that diet and lifestyle (environmental factors) were more important risk factors than heredity (genetic factors) for this type of cancer.

Besides the main purpose of generating hypotheses, descriptive studies can also be used to assess the health status of a population and to plan public health programs. Descriptive studies include the following types:

1. *Prevalence surveys* are sample surveys conducted for the purpose of estimating the prevalence of a health condition or outcome or exposure to risk factors in a population, at a particular point in time.

2. In a *cross-sectional study*, both the health condition (or outcome) and exposure to risk factors are

measured on the same subjects at the same time, so that the joint distribution is also available. Cross-sectional studies can be used to calculate the prevalence ratio (referring to the proportion of existing cases), but not the incidence ratio (referring to the rate of occurrences of new cases), and can generate causal hypotheses but not to draw causal inferences.

3. A *case report* describes a single case and often focuses on unusual aspects of a patient's disease/condition or unusual association between the diseases and exposures, while a *case series* is a study of a series of case reports with a common health outcome of interest.

4. *Surveillance studies* refer to the ongoing systematic collection, analysis, and interpretation of outcome-specific data: These studies can be very useful, for instance, in alerting public health officials if many cases of a previously rare disease are occurring in a particular area, which would suggest the need for further investigation.

5. *Analysis of secondary administrative data* is not a true study design but refers to analysis of routinely collected data, such as those from population census, vital registrations, and tumor registries with the usual demographic characteristics of age, sex, race, and so on, and region. The above example of cancer study on Japanese migrants using mortality rates pertains to this type.

Analytic Studies

Analytic studies are the other main type of research design in epidemiology. They are designed to make comparisons and to test statistically hypothesized causal relationships. Ecological (group-level) data may be used for descriptive studies for the purpose of generating hypotheses, but for testing hypotheses, analytic studies are generally employed that require individual-level data. Analytic studies consist of experimental (intervention) studies, quasi-experimental studies, and nonexperimental (observational) studies.

1. In *experimental (intervention) studies*, the investigator manipulates the study factor (exposure variable) and randomly allocates experimental units to different exposure groups so as to reduce bias and increase validity.

2. In *quasi-experimental studies*, the investigator manipulates the study factor only but not randomization of study subjects. Both experimental and quasi-experimental studies must progress longitudinally in time from exposure to outcome. This clear temporal relationship strengthens the researcher's ability to make causal inferences. In addition, random assignment of experimental units to exposure groups in experimental studies improves the researcher's ability to make causal inferences.

3. In *nonexperimental (observational) studies*, the investigator does not randomize the study subjects and also has no control on the study factor (exposure status) that is determined by the experimental units themselves and can only observe, record, and analyze results. An observational study may be either longitudinal or cross-sectional, and when longitudinal, it may be either prospective or retrospective. There are many observational study designs: We mention six of the most common below.

   a. In *prospective cohort studies*, the investigator begins by selecting two groups from the target population to whom the outcome event(s) of interest has not occurred, one exposed and the other not exposed to some risk factor of interest. Both groups are then followed longitudinally to observe the incidence of the outcome event(s) for a fixed period of time, and the investigator notes differences in the incidence of the outcome between the groups at the end of the follow-up.
   b. In *retrospective (historical) cohort studies*, the event of interest has already occurred, and the investigator uses a complete record of secondary data to define which subjects were exposed to which risk factors and traces the subjects from some past time point to the present using the historical record rather than following the participants prospectively over time.
   c. In *case-control studies*, the researcher begins by selecting for study individuals who either have experienced the outcome event (called the *cases)* or have not experienced it (called the *controls*). These two groups are then traced back in time to determine their exposure status. Case-control study design provides weaker causal inference than cohort studies, due mainly to the existence of confounding factors as well as selection and recall biases. However, case-control studies are more efficient for studying rare diseases.
   d. In *ecological studies*, both exposure status and outcome status are measured on groups of people rather than on individuals. Examples of groups as units of observation include classes in school, counties, nations, occupational groups, and socioeconomic groups. Outcome event rates and exposures are

measured in each of a series of groups of people, and their relation is examined to generate hypothesis (descriptive study) or to test etiologic hypothesis at a population level (analytic study). For example, one may plot standardized death rates versus smoking rates for each of many (say, 30) counties to examine the relationship between tobacco smoking and mortality at the county level. Ecological studies are prone to ecological bias when cause-effect relationships at individual level are being inferred from observations at a population level.

e. *A nested case-control study* is a case-control study nested within a cohort study in which only one group without reference to the exposure status is followed for incidence of the outcome event. The incidences that emerge during the study period are entered into the case-control study as cases, with controls sampled from those without outcome event in the cohort during the study period and matched, taking the time of occurrence of the outcome event as the matching variable. Finally, the exposure frequencies are compared between cases and controls as in the classical case-control study to see if there is an association between exposure and outcome. The nested case-control study design is less subject to recall bias than the classical case-control study, and it is more efficient than the classical cohort study.

f. A *case-cohort study* is an unmatched nested case-control study with controls sampled from those without outcome event in the cohort at the beginning of the study period, that is, prior to any occurrence of incidence case, rather than during the study period. Exposure information needs to be ascertained on only those sampled as controls at the beginning of the study period and subsequent cases occurring during the study period. The main advantage of case-cohort study over the nested case-control study is that the same subset of cohorts can be used as a control group for studies of multiple outcomes for the former, while the latter requires different risk sets for each outcome studied. Consequently, nested case-control studies have gradually been replaced by case-cohort studies, particularly in investigation of biomarkers as risk factors for chronic diseases.

—John J. Hsieh, Changyong Feng, and Hongyue Wang

See also Causation and Causal Inference; Clinical Trials; Ecological Fallacy; Public Health Surveillance; Quasi Experiments; Study Design

Further Readings

Ann, A., & George, R. S., III. (2003). *Essentials of epidemiology in public health*. Boston: Jones & Bartlett.

Friis, R. H., & Sellers, T. A. (2004). *Epidemiology for public health practice* (3rd ed.). Boston: Jones & Bartlett.

Kelsey, J. L., Whittemore, A. S., Evans, A. S., & Thompson, W. D. (1996). *Methods in observational epidemiology* (2nd ed.). New York: Oxford University Press.

Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic research: Principles and quantitative methods*. Belmont, CA: Life Learning.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika, 73*, 1–11.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

Determinants of Health Model

The determinants of health model is a hypothetical construct for understanding population health and the multiple range of factors that determine its level. In a broad sense, the determinants of health model is the most recent and comprehensive explanatory attempt to understand causality in epidemiology and its translation into policy intervention, from both the population and societal perspectives; therefore, the model is particularly relevant to public health research and practice. The current determinants of health model is firmly rooted in the eco-epidemiology paradigm, insofar as it recognizes causes of health—and "causes-of-causes," that is, their determinants—at multiple levels of organization and within the historical context of both societies and individuals. Those levels of organization span from the microlevel, downstream, or proximate causes to the macrolevel, upstream, or distal causes of population health and have, as a chief model feature, the multilevel interaction among causes.

Determinants of health are all those factors, whether events, characteristics, or other definable entities that, directly or indirectly, exert an influence on the health of individuals and, by means of their action and interactions among different levels of organization, *determine* the given status of the health of the population. The understanding of the determinants of the population's health and their complex causal patterns has been originally shaped, and continue to be shaped, by contemporary debates—as well as their historical

circumstances—about medical care or curative medicine and health policy, and then by the unremitting need to reexamine and reorient the social debate to better define and put in place the type and scope of reforms that would help improve the health of the population.

Over the course of the second half of the past century, there have been several significant shifts in perspective, each of them associated with correspondingly different *models* of health determinants. A major motivation for this search for a better health determinants model has been the growing need to define a population health approach, vis-à-vis the emergent understanding that the determinants of health are much broader than the medical care system itself. This evolving thinking can be regarded as a consequence of the recognition of, on the one hand, the unsuitability of the dominant biomedical perspective to guide health policy toward the equitable improvement of a democratic nation's health status and, on the other hand, that approaching health from a population perspective commits a nation to understanding and acting on the full array of factors that determine health.

Marc Lalonde's 1974 report on a new perspective on the health of the population in Canada signaled the emergence of the modern era of health promotion by advancing a more comprehensive conceptual framework for health situation analysis and advocating the role of preventive approaches in national health policy. Under this model, health determinants were deemed as pertaining to four dimensions, or so-called health fields: human biology, environment, lifestyle, and health care organization. The report expressed a widely shared view that health determinants go beyond the individual level and, more specifically, that health was not attainable for the majority of the population through a concentration of public health funds on personal services. Fueled by the prominent risk-factor epidemiology paradigm, however, the emphasis was rapidly placed on the role of personal behavioral choices or individual lifestyles in determining health status.

A subsequent shift in perspective took place with the 1986 Ottawa Charter and the expansion of the health promotion movement, calling attention to the role of factors outside the health care sector and, in particular, the social and economic factors as determinants of population health status. Simultaneously, the emergence of ecological activism gave prominence to environmental determinants of health advocating, especially healthy behaviors in the workforce and safety rules in the workplace. Since the early 1990s, evidence has been growing worldwide in support of the contention that society's capacity to generate, accumulate, and/or fairly distribute wealth and prosperity is also of paramount importance to determine the health of its population.

The current and preferred model of determinants of population health was originally proposed by Göran Dahlgren and Margaret Whitehead to guide WHO strategy to promote health equity in Europe and has also guided the Independent Inquiry into Inequalities in Health in the United Kingdom and the Committee on Assuring the Health of the Public in the 21st Century of the Institute of Medicine of the National Academies in the United States. This multilevel model embodies the emergent eco-epidemiology paradigm that postulates an integrated approach to investigating disease in the population and its prevention by subsuming levels of causation, life course trajectories, kinds of causes, and types of diseases, that is, disease causation, pathogenesis, and population health as processes taking place at multiple levels of organization and within the historical context of both societies and individuals.

In addition to the more "downstream" biological and behavioral bases for disease, that is, the proximate health determinants, the Dahlgren and Whitehead model identifies four layers of main influences on health, toward the most "upstream" bases for population health or distal health determinants:

1. the level of individual lifestyles factors and attitudes, such as eating, sleeping, drinking and smoking habits, exercise, sex, consumption patterns, etc.;
2. the level of social, family, and community networks, including social relations, social support, neighborhood contextual influences, social membership, significant others, etc.;
3. the level of material and social conditions in which people live and work, such as housing, water and sanitation, social security, education, employment, work environment, food production and availability, health care services, transportation and other urban-planning determinants, etc.; and
4. the level of broad social, economic, cultural, and environmental conditions, which includes political determinants, socioeconomic inequalities, economic, societal and health systems, environmental

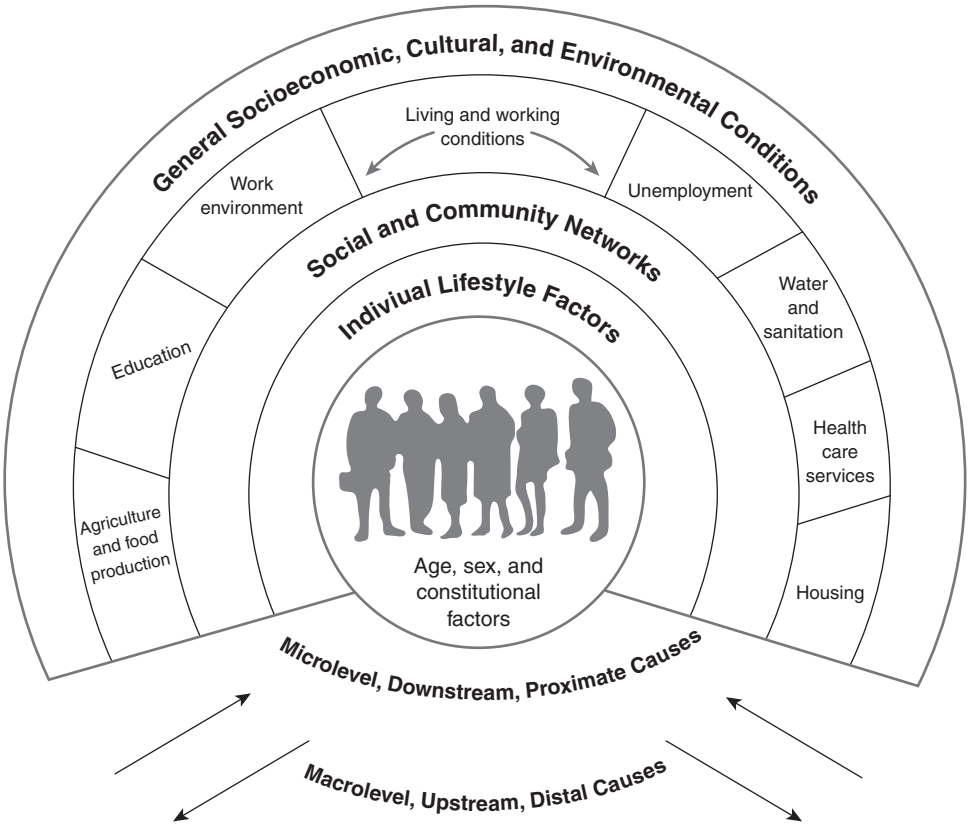


Figure 1 The Dahlgren and Whitehead Model of Health Determinants

Source: Dahlgren and Whitehead (1991). Used with kind permission of the Institute for Futures Studies, Stockholm, Sweden.

preservation, cultural values such as tolerance to diversity, and policies and societal norms at the global, international, national, regional, and local levels.

Figure 1 helps illustrate the multilevel nature of the multiple determinants of population health and the interaction between and among the various levels of health determinants. It may help to see the powerful influences of income maintenance and fairness, educational attainment, public health services, social security and/or air quality, for instance, on a population's health. More important, it may illustrate how macro-, mid-, and microlevel determinants interact along complex and dynamic pathways to produce health at the population level, as well as to realize how historical context changes over time as, for instance, life course changes at the individual level, demographic change at the societal level, and disease itself also changes as agents evolve, adapt, and modify their pathogenicity. The figure may also help show how the health sector must work with other sectors, and how community networks must be taken into account, to generate health policies that improve a population's health. In fact, these levels of hierarchical organization of the population health determinants in the Dahlgren and Whitehead model translate, conversely, into four levels for policy intervention, aimed at (1) influencing individual lifestyles and attitudes, (2) strengthening social and community support, (3) improving living and working conditions, and (4) bringing about long-term structural changes, respectively.

Epidemiology at both edges of the Dahlgren and Whitehead determinants of health model is in frank expansion: Molecular and genetic epidemiology from the side of proximate determinants and social epidemiology from the side of distal determinants are increasing the knowledge base on population health and the potential to create healthy public policy. The development and repercussions of the Human Genome

Project, on the one hand, and the works and efforts of the WHO Commission on Social Determinants of Health, on the other hand, are just two outstanding examples of the burgeoning intellectual activity in which epidemiology is currently engaged to further progress the appropriate conceptualization of the determinants of population health and advance the appropriate public health practice to truly achieve health for all.

—Oscar J. Mujica

See also Causation and Causal Inference; Eco-Epidemiology; Life Course Approach; Multilevel Modeling

Further Readings

Dahlgren, G., & Whitehead, M. (1991). *Policies and strategies to promote social equity in health.* Stockholm, Sweden: Institute for Futures Studies.

Evans, R. G., Barer, M. L., & Marmor, T. R. (Eds.). (1994). *Why are some people healthy and others not? The determinants of health of populations*. New York: Aldine de Gruyter.

Institute of Medicine (U.S.) Committee on Assuring the Health of the Public in the 21st Century. (Ed.). (2003). *The future of the public's health in the 21st century.* Washington, DC: National Academy Press.

Lalonde, M. (1974). *A new perspective on the health of Canadians.* Ottawa, Ontario, Canada: Information Canada.

Smedley, B. G., & Syme, S. L. (Eds.). (2000). *Promoting health: Intervention strategies from social and behavioral research.* Washington, DC: National Academy Press.

Diabetes

Diabetes mellitus is a chronic condition in which the pancreas, a gland near the stomach, fails to make insulin or does not make enough insulin, or in which the body becomes insensitive to insulin. Insulin is a hormone that works to convert what we eat to glucose or sugar. Insulin supplies muscles and other tissues with glucose for growth and energy. Without insulin, the body's blood sugar cannot be regulated, leading to a buildup of sugar in the blood (hyperglycemia) and loss of fuel for the body. Symptoms of diabetes include excessive thirst, frequent urination, unexplained weight loss, extreme hunger, vision changes, tingling in the hands or feet, dry skin, and sores that are slow to heal.

Diabetes is a major public health problem. In 2005, approximately 1.1 million people worldwide died from diabetes. Diabetes is the sixth leading cause of death in the United States. The consequences of diabetes are severe. Diabetes is associated with heart disease, stroke, kidney failure, blindness, nontraumatic amputations, and nerve damage.

History

The earliest written record of diabetes dates back to 1552 BCE in Egypt when a physician, Hesy-Ra, described one of the symptoms of diabetes as frequent urination. It was not until the late 19th century that scientists began to understand the disease. In 1869, a German medical student, Paul Langerhans, described islands of cells in the pancreas. Later, these cells were discovered to be the source of insulin and were named ''Islets of Langerhans.'' In 1889, two European scientists, Oskar Minkowski and Joseph von Mering, discovered that removing the pancreas from dogs resulted in diabetes and therefore recognized that diabetes is a disease of the pancreas. In 1921, Canadian scientists Frederick Banting and Charles Best extended the work of Minkowski and von Mering by giving dogs without a pancreas extracts from the Islets of Langerhans from healthy dogs. Consequently, Banting and Best were able to isolate insulin and to inject insulin from bovine pancreases into humans. Their first patient was an 11-year-old boy, Leonard Thompson, who was suffering from diabetes, at that time an invariably fatal disease. After Banting and Best injected the boy with insulin, his blood sugar levels decreased, and he thrived.

Types of Diabetes

There are three main types of diabetes: type 1, type 2, and gestational diabetes.

Type 1 Diabetes

Type 1 diabetes occurs when the pancreas produces little or no insulin and it is necessary to take insulin daily by injection. Type 1 diabetes is thought to be an autoimmune disease in which the immune system attacks the insulin-producing cells of the pancreas. Approximately 10% of persons with diabetes are type 1 and most new cases occur in children; in fact, diabetes is one of the more common chronic diseases among children. Around 40% of type 1 cases are found in

persons less than 20 years of age at the onset, and the incidence peaks at ages 2, 2 to 6, and 10 to 14.

It is unclear what causes type 1 diabetes, because the exact mechanism for developing the disease is unknown. However, many scientists believe that development of type 1 diabetes follows exposure to an "environmental trigger" that results in an attack against the beta cells of the pancreas in some genetically predisposed people. Risk factors for type 1 diabetes are not well understood, but autoimmune, genetic, and environmental factors are involved in developing this type of diabetes. There has been some implication of a viral connection but that has yet to be proven. Relatives of type 1 diabetics have a 10% to 15% greater risk of developing the disease than those without any family history of the disease. Males and females are at similar risk for type 1 diabetes.

Type 2 Diabetes

Type 2 diabetes occurs when the pancreas produces insulin but either the pancreas produces too little for the body (insulin deficiency) or the body does not respond effectively to the insulin (insulin resistance). Type 2 usually develops in persons over the age of 40, although recently cases have been increasing among children. Many type 2 diabetics go undiagnosed for many years, and therefore it is difficult to determine true incidence and prevalence in the population, but it is estimated that about 90% of persons with diabetes are type 2. About 80% of persons with type 2 diabetes are overweight. Increasingly, young children are being diagnosed with type 2 diabetes, and this increase has been associated with consumption of a high-calorie diet and insufficient exercise. Between 45% and 80% of children diagnosed with type 2 diabetes have at least one parent with diabetes. The risk factors include older age, obesity, family history of diabetes, prior history of gestational diabetes, impaired glucose tolerance, physical inactivity, and race/ethnicity. African Americans, Hispanic/Latino Americans, American Indians, and some Asian Americans and Pacific Islanders are at higher risk for this type of diabetes. Native Americans have the highest prevalence of type 2 diabetes: Prevalence among the Pima Indians of the American Southwest is more than 50% among adults. Unlike type 1 diabetes, type 2 diabetes can be prevented through a number of ways such as maintaining a healthy body weight and increasing the amount of physical activity: For instance, a number of studies have shown that regular physical activity can significantly reduce the risk of developing type 2 diabetes.

Gestational Diabetes

Gestational diabetes is the third type of diabetes. Gestational diabetes occurs when women develop diabetes during pregnancy and usually lasts only during the pregnancy. Gestational diabetes occurs in 2% to 5% of all pregnancies. Offspring of mothers who have had gestational diabetes are at higher risk of developing diabetes during their lifetime. Gestational diabetes occurs more frequently in African Americans, Hispanic/Latino Americans, American Indians, and people with a family history of diabetes. Obesity is also associated with higher risk. Women who have had gestational diabetes are at increased risk of later developing type 2 diabetes. Some studies have found that nearly 40% of women with a history of gestational diabetes developed type 2 diabetes in the future.

Prevalence and Incidence

According to the Centers for Disease Control and Prevention, more than 14 million Americans have diagnosed diabetes and approximately 6 million are undiagnosed. The number of Americans with diabetes is expected to increase to 29 million by 2050.

There are 1.4 million Americans with type 1 diabetes and 10 to 20 million people worldwide. The incidence of type 1 diabetes is increasing each year, by approximately 3% to 5%, and about 50,000 new cases are diagnosed each year worldwide. The highest rates of type 1 can be found in Scandinavia and Sardinia, while Asia has the lowest rates in the world.

The prevalence of type 2 diabetes is expected to double in the next 25 years; in 2000, there were 150 million type 2 diabetics worldwide; in 2025, it is estimated that there will be 300 million.

Treatment

Persons with diabetes need to eat nutritious foods, exercise, and self-monitor blood glucose levels to maintain good control of their diabetes. Some type 2 diabetics need to take oral medication or insulin to regulate their blood glucose levels, and all type 1 diabetics must use insulin to control their blood glucose levels. The HbA1c (glycosylated hemoglobin) is a lab test that is conducted every 3 months to

determine average blood glucose level over the past 3 months. The 10-year Diabetes Control and Complications Trial found that persons with diabetes who had very good blood glucose control (as measured by lower HbA1c) reduced their risk of complications of eyes, kidneys, and nervous systems.

Research

Current research is examining ways to cure diabetes. Transplantation of human islets into patients with type 1 diabetes and pancreas transplantation are approaches that may be more common in the future. However, there is a shortage of donor pancreases, and persons who receive a pancreatic transplantation must remain on immunosuppressant drugs. Other research is being conducted in the use of stem cell therapies to cure diabetes. Medical devices, such as insulin pumps, have not been a cure for diabetes but have enabled patients to significantly improve blood glucose control.

—Britta Neugaard

See also Chronic Disease Epidemiology; Obesity; Physical Activity and Health

Further Readings

Centers for Disease Control and Prevention. *Basics about diabetes*. Retrieved November 17, 2006, from http://www/cdc.gov/diabetes/faq/basics.htm.

Porte, D., Jr., Sherwin, R. S., & Baron, A. (Eds.). (2003). *Ellenberg and Rifkin's diabetes mellitus: Theory and practice* (6th ed.). New York: McGraw-Hill.

World Health Organization. (2006). *Fact sheet #*312 "diabetes." Retrieved November 13, 2006, from http://www.who.int/mediacentre/factsheets/fs312/en/index.html.

Diffusion of Innovations

The *diffusion of innovations* model describes how change takes place within a social system and provides a schema for the systematic study of the adoption of a product, a practice, or an idea by members of a social system. The most commonly used definition of diffusion of innovations is that articulated by Everett Rogers (1995): "Diffusion is the process by which an innovation is communicated through certain channels over time among members of a social system" (p. 5).

The model offers a logical fit with the traditional task of epidemiology—to systematically study changes in the health of populations with a focus on the patterns of disease occurrence and the factors that influence these patterns. The diffusion model offers opportunities to study and analyze changes in products and exposures as well as in behaviors and policies related to health. The diffusion model and the discipline of epidemiology are both concerned with incidence, prevalence, and change as well as with time, place, and persons. The diffusion model offers a valuable framework to guide epidemiologists in the formulation of research questions to study a wide range of health-related exposures, behaviors, as well as policies and their effects on the health of populations.

The diffusion model is based on two underlying premises: (1) that communication is essential for the diffusion and subsequent acceptance or rejection of an innovation and (2) that new products, practices, and ideas can alter the structure and function of a social system. Change is measured by the numbers of people, groups, or institutions adopting the innovation over time. Consequently, the key variables of interest are time (earliness of knowing about innovation), rate (adoption of different innovations in a social system and/or within and among different social groups), and innovativeness (the degree to which groups or organizations adopt new ideas). Analyses can establish the flow of influence, offer charts of the diffusion curve, develop mathematical models of the diffusion process, and test out contributions of key elements and characteristics.

Lessons learned from diffusion studies in anthropology, sociology, education, folklore, communication, marketing, economics, and public health have helped contemporary scholars and practitioners transform the diffusion of innovation model from a descriptive model into a proscriptive one. Diffusion theory is currently used as an analytic framework for understanding and measuring social change and, in practical application, to guide the design and evaluation of products, programs, and communication strategies in public health and health communication.

History

The diffusion model developed through an interest in social transformation and explorations of the

consequences of the development, spread, and adoption or rejection of new products, activities, and ideas. For example, anthropologists studied the introduction of the horse within and among indigenous population groups of North America, the spread and modification of dance ceremonies among Native American groups, and the spread of corn cultivation from America to Europe. Early sociological studies included the examination of social and legal trends, such as the influence of a city on surrounding areas, the diffusion of governing practices, and the use and consequences of technology. Rural sociologists focused on the spread of new ideas among farmers and the subsequent change in agricultural practices. Thousands of studies in myriad fields have added to the diffusion literature over time.

Constructs

Key constructs include the innovation, the communication processes and channels as well as the agents of change, the potential adopters, and the social system. The innovation (the product, practice, or idea) need not be a new invention but must be perceived as new by individuals or other units of adoption. Its characteristics influence the likelihood that it will be adopted. These include relative advantage, compatibility, malleability, and complexity.

Innovations that are perceived to be of lower social or economic costs, that provide a good fit with values and current practices, and that are of low complexity are more likely to be adopted than those carrying high costs, representing a variance with common values, and appearing to be difficult to understand, to communicate, or to use. Those innovations flexible enough to withstand some change to provide a better fit with prevailing practices and cultures hold more appeal. This flexibility, sometimes labeled a reinvention, involves change in some of the characteristics of the innovation to increase compatibility of the innovation with the existing social system. It is also of importance whether or not potential users can observe as well as try the innovation without undue sacrifice or commitment.

Communication channels and the agents of change affect the diffusion process as well. Face-to-face communication, for example, is generally considered stronger than is mass communication. Radio has greater reach than does television in preindustrialized societies. Furthermore, agents of change—those bringing innovations to members of a social system—often have a stronger effect if they are members of or like members of the community rather than if they are perceived as being marginal to or outsiders of the social structure of the community. Overall, the diffusion analysis must consider who talks to whom, who is considered influential and trustworthy, and who has easy access to or is barred from various communication channels. Characteristics of the potential adopters are of critical concern. Overall, factors such as socioeconomic status, culture, gender, race, age, cultural norms, religion, education, social support, and family ties all influence access to and perceptions of the innovation.

A vital aspect of the diffusion model, and one closely linked to an analysis of adopters, is the consideration of time. The population is often divided into groups based on the time it takes for people to adopt the innovation. The groups are innovators, early adopters, early majority, late majority, and late adopters. Innovators, for example, are often viewed as creative but marginal individuals. Early adopters, close to sources of communication, are often highly integrated into the social system. They often carry a high degree of opinion leadership, are respected by their peers, and serve as role models for others. It is only when these respected members of a community consider, discuss, and adopt the innovation that wide diffusion takes place. Those in the early majority generally interact frequently with peers and are exposed to various sources of information. Those in the late majority are people who are further removed from key communication channels or who remain skeptical and adopt only after pressure from their peers or out of economic necessity.

In older articulations of the model, those who are introduced to the innovation late and who adopt late were called laggards. Often, people who fall into the laggard category are distant, disadvantaged, or marginalized. These members of the population were often further removed from key channels of communication than were those who were able to learn about and adopt the innovation early. They are also more likely to lack resources, including time and money, to take chances. Often, they are socially isolated. It is precisely this part of the population, often considered "at risk," that is of greatest concern to public health policymakers and practitioners. Thus, public health program strategies often include a purposive targeted dissemination to those who are further removed from traditional communication sources and who live in relative disadvantage.

Diffusion occurs within a social system. The analysis of the diffusion process considers the members or units of a social system, including individuals, groups, organizations, or subsystems. The social system includes structural, political, economic, as well as geographic characteristics. The structure may be considered the patterned arrangements of the various units, such as the formal hierarchical structure of a bureaucratic organization with formal laws and rules. Norms may be just as powerful in other, less formal groups. Social factors also consider how decisions are made and whether or not people have an array of options or, at minimum, the freedom to adopt or reject the proposed innovation. Decisions may be individually based, be communal (arrived at through consensus), or be mandated (made through the imposition of authority). Those designing public health safety programs, for example, see initial strong adoption of practices such as seat belt use with the passage of laws (an authority-based decision) but diminished use when consensus was not considered and if enforcement is lacking. In general, the structure of a social system can facilitate or impede diffusion of innovations and thereby influence the rate of adoption of the innovation over time.

Finally, consequences of innovations are documented and analyzed. Three classifications of consequences are often considered: direct or indirect consequences, anticipated versus unanticipated consequences, and desirable or undesirable consequences.

—Rima E. Rudd and Vanessa Watts

See also Governmental Role in Public Health; Health Behavior; Health Communication; Health Disparities

Further Readings

Goodenough, W. H. (1963). *Cooperation in change.* New York: Russell Sage.

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

Rogers, E. M., & Shoemaker, F. F. (1971). *Communication of innovations: A cross-cultural approach* (2nd ed.). New York: Free Press.

Direct Standardization

Direct standardization is a method of comparing population disease or death experiences, removing the effects of differences in population characteristics. Direct standardization is an important tool used in comparing the health of populations. Additionally, it can be used to monitor the health of the nation as is done with the Healthy People 2010 objectives. This entry uses an example of deaths from heart disease in the state of New York to illustrate the use of direct standardization.

Do men really have more heart disease deaths than women in the state of New York? In this case, two populations are compared, New York men and New York women. The heart disease death rate for New York men is 144.1 per 100,000 in 2000, and for New York women it is 160.5 per 100,000. Thus, the rate of death due to heart disease for women is higher than that for men. This finding is surprising but true, and is due largely to the fact that women tend to live longer than men, and thus live to an age where heart disease is a very common cause of death. In the terminology of epidemiology, the comparison of heart disease death rates is confounded by age. The death rates for men and women are really composed of two components: different heart disease death rates for each age group and different age distributions.

In this study, age is a confounder, a nuisance variable that affects our ability to clearly understand how gender affects the death rate for heart disease. Both the age distributions of men and women in the United States and the fact that heart disease deaths are more common as people get older are well known, but we need an analytical technique that incorporates these facts into our comparison of heart disease death rates by gender. Direct standardization provides that technique and allows us to calculate age-specific death rates, which compare heart disease death experiences by gender while controlling for age. Table 1 shows that for each age group, men have a higher heart disease mortality rate compared with women. The reason the overall, or crude, mortality rate is higher for women compared with men is because of the older age distribution for women in New York.

Both the crude mortality rates and the standardized mortality rates can be viewed as weighted averages of the age-specific mortality rates. In the crude mortality rate, the weighting reflects the number of men and women in each age category in New York State, while in the standardized rate, a reference or standard population is chosen. In Table 1, the crude rate for men is rewritten as weighted average.

Table 1 Crude Rates

	Men			*Women*		
	Number of Cases	*Number Population*	*Age Specific Rate***	*Number of Cases*	*Number Population*	*Age-Specific Rate***
< 45 years	774	15,304,246	5.1	336	15,959,393	2.1
45–54	1,684	1,226,925	137.3	613	1,326,011	46.2
55–64	2,977	786,187	378.7	1,492	901,800	165.4
65–74	5,540	561,262	987.1	3,772	714,784	527.7
Total	26,357	18,293,496	144.1	31,562	19,659,418	160.5

Note: **Per 100,000 in 2000.

Men

$$\text{Overall rate (crude)} = \frac{(\text{Total cases})}{(\text{Total population})} = \frac{(26,357)}{(18,293,496)} = 144.1 \text{ per } 100,000$$

$$= \frac{\left(\text{Pop}_{<45} \times \frac{(\text{Cases}_{<45})}{(\text{Pop}_{<45})}\right) + \left(\text{Pop}_{45-54} \times \frac{(\text{Cases}_{45-54})}{(\text{Pop}_{45-54})}\right) + \left(\text{Pop}_{55-64} \times \frac{(\text{Cases}_{55-64})}{(\text{Pop}_{55-64})}\right) + \left(\text{Pop}_{65-74} \times \frac{(\text{Cases}_{65-74})}{(\text{Pop}_{65-74})}\right) + \left(\text{Pop}_{65-74} \times \frac{(\text{Cases}_{75+})}{(\text{Pop}_{75+})}\right)}{\text{Pop}_{<45} + \text{Pop}_{45-54} < \text{Pop}_{55-64} + \text{Pop}_{65-74} + \text{Pop}_{75+}}$$

$$= \frac{\left(15,304,246 \times \frac{(774)}{(15,304,246)}\right) + \left(1,226,925 \times \frac{(1684)}{(1,226,925)}\right) + \left(786,187 \times \frac{(2,977)}{(786,187)}\right) + \left(561,262 \times \frac{(5,540)}{(561,262)}\right) + \left(414,876 \times \frac{(15,382)}{(4,414,876)}\right)}{15,304,246 + 1,226,925 + 786,187 + 561,262 + 414,876} = 144.1$$

$$= \frac{(\text{Pop}_{<45} \times \text{Age-specific rate}_{<45}) + (\text{Pop}_{45-54} \times \text{Age-specific rate}_{45-54}) + (\text{Pop}_{55-64} \times \text{Age-specific rate}_{55-64}) + (\text{Pop}_{65-74} \times \text{Age-specific rate}_{65-74}) + (\text{Pop}_{65-74} \times \text{Age-specific rate}_{75+})}{\text{Pop}_{<45} + \text{Pop}_{45-54} + \text{Pop}_{55-64} + \text{Pop}_{65-74} + \text{Pop}_{65-74}}$$

$$= \frac{(15,304,246 \times 5.1) + (1,226,925 \times 137.3) + (786,187 \times 378.7) + (561,262 \times 987.1) + (414,876 \times 3,707.6)}{15,304,246 + 1,226,925 + 786,187 + 561,262 + 414,876} = 144.1$$

$$= \frac{\sum \text{Age-specific pop}_i \times \text{Age-specific rate}_i}{\sum \text{Age-specific pop}_i.}$$

Women

$$\text{Overall rate (crude)} = \frac{(\text{Total cases})}{(\text{Total population})} = \frac{(31,562)}{(19,659,418)} = 160.5 \text{ per } 100,000.$$

Source: New York State Vital Statistics (2000).

In theory, any population can be used as a standard population, even a fictional population. However, a directly standardized rate can be compared only with another directly standardized rate, so if directly standardized rates from this study are to be compared with other studies, then the standard population needs to be the same for both. In the United States, the 2000 census is the standard population used by most

Table 2 Direct Standardization

Men

$$= \frac{\begin{array}{l}(\text{Standard pop}_{<45} \times \text{Age-specific rate}_{<45}) + (\text{Standard pop}_{45-54} \times \text{Age-specific rate}_{45-54}) \\ \quad + (\text{Standard pop}_{55-64} \times \text{Age-specific rate}_{55-64}) + (\text{Standard pop}_{65-74} \times \text{Age-specific rate}_{65-74}) \\ \quad + (\text{Pop}_{65-74} \times \text{Age-specific rate}_{75+})\end{array}}{\text{Pop}_{<45} + \text{Pop}_{45-54} + \text{Pop}_{55-64} + \text{Pop}_{65-74} + \text{Pop}_{75+}}$$

$$= \frac{\begin{array}{l}[(178,932,504 \times 5.1) + (37,030,152 \times 137.3) + (23,961,506 \times 378.7) \\ \quad + (18,135,514 \times 987.1) + (16,573,966 \times 3707.6)]\end{array}}{[178,932,504 + 37,030,152 + 23,961,506 + 18,135,514 + 16,573,966]} = 243.9.$$

Women

$$= \frac{\begin{array}{l}[(178,932,504 \times 2.1) + (37,030,152 \times 46.2) + (23,961,506 \times 165.4) \\ \quad + (18,135,514 \times 527.7) + (16,573,966 \times 3,346.7)]\end{array}}{[178,932,504 + 37,030,152 + 23,961,506 + 18,135,514 + 16,573,966]} = 209.7$$

Source: National Center for Health Statistics (2006).

researchers and the federal government (e.g., Healthy People 2010 goals). Thus, if the 2000 U.S. Census is used in this study, then not only can the standardized rates be compared within this study, but they can also be compared with national directly standardized rates.

Table 2 shows the equations for direct standardization. The formula is the same weighted average formula used for crude rates, but the weights are those taken from the standard population. Thus, the only difference between the formula for men and women now are the age-specific heart disease death rates, and the directly standardized rates are no longer confounded by age. The results show that men have a higher heart disease death rate than women, when directly standardized to the U.S. 2000 population. Note that the raw numbers calculated in direct standardization are meaningless except for purposes of comparison.

The target for heart disease mortality rate is 166 per 100,000, directly standardized to the 2000 U.S. Census, by 2010. Because the rates in this example involving New York are standardized to the same population, comparisons can be made. Clearly, death rates in New York State are much higher than the 2010 target for the nation, and in fact, New York State has one of the highest heart disease mortality rates in the United States.

—Shazia Hussain and Louise-Anne McNutt

See also Healthy People 2010; Indirect Standardization; Public Health Surveillance Rate

Further Readings

Hennekens, C. H., & Burning, J. E. (1987). *Epidemiology in medicine.* Boston: Little, Brown.

National Center for Health Statistics. (2006). *Health, United States, 2006, with chartbook on trends in the health of Americans*. Hyattsville, MD: Author.

Rossignol, A. (2007). *Principles and practice of epidemiology: An engaged approach.* New York: McGraw-Hill.

Rothman, K., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott-Raven.

Timmreck, T. C. (1998). *An introduction to epidemiology* (2nd ed.). Boston: Jones & Bartlett.

Disability Epidemiology

Disability epidemiology constitutes a subdiscipline within epidemiology, similar to areas such as maternal and health epidemiology or nutritional epidemiology, which focus on a particular area and augment the approach and methodology common to many epidemiological endeavors with specific concerns and techniques relevant to the subject matter. This entry describes several key aspects of the definition of

disability epidemiology, provides an overview of disability surveillance, and examines methodological issues in the field.

Definitions

The practice of disability epidemiology includes consideration of the basic elements of descriptive epidemiology (who, what, where), etiological determinants of impairments, and the frequency and predictors of different outcomes of disability. All these questions incorporate the common epidemiology study designs and methods. Adapting from John Last's classic definition of epidemiology, disability epidemiology is, then, the study of the distribution, determinants, correlates, and outcomes of disability and application of this study to maximizing the health and quality of life of people and populations with disability (PWD).

The concepts and content of disability epidemiology derive naturally from a number of other epidemiology disciplines, such as injury, neurological diseases, maternal and child health, aging, and outcomes research. Despite some differences in methods and focus, epidemiology principles do not alter remarkably as one moves from one content area to another.

However, unlike specific disease areas (e.g., neurological diseases), the focus of disability epidemiology is less concerned with diagnoses and specific impairments than with broad functional levels, such as mental, sensory, and mobility. This focus is derived from the definition of disability found in the World Health Organization (WHO) International Classification of Functioning, Health and Disability (ICF), which is based on a social versus a traditional medical model of disability. This paradigm moves from a diagnosis and treatment model of disability to one where people are defined in the broader context of social roles and participation—so that consideration of disability is essentially removed from the constraints of medical care and definitions. In this paradigm, *disability* is an umbrella term, where participation (the key outcome) is a function of aspects of environment, intrinsic personal factors, body function and structure, and activities. The ICF has been proposed as the unifying framework for disability measurement in public health.

In the ICF paradigm, disability must be defined as a state that is largely independent of health and health status. The distinction between the two is difficult to incorporate into traditional epidemiology, however. While epidemiologists are comfortable with the notion of prevention occurring at primary, secondary, and tertiary levels, in disability research, prevention includes both prevention of new primary disabilities or impairments and prevention of the adverse outcomes of disability. This separation of health and disability is accomplished in the Healthy People 2010 chapter on disability. Prominent in this document is the following basic assumption about disability:

> Disability is a demographic descriptor rather than a health outcome. It should be used to monitor disparities in health outcomes and social participation. The Americans with Disabilities Act (ADA) provides an important rationale for universal collection of disability status [in data collection]. (Objectives Report: Disability and Secondary Conditions; Centers for Disease Control and Prevention [CDC], 2001)

The field of disability epidemiology has a small group of professionals who concentrate primarily on disability epidemiology and a number of others whose work touches on that subject matter and approach.

Descriptive Epidemiology: Disability Surveillance

On the basis of the UN disability statistics, the WHO estimates that about 600 million people worldwide live with disability (global statistics and data collection efforts are available on the United Nations Web site listed at the end of this entry). The United States and other Western nations answer public health questions and program plans regarding disability based on ongoing surveillance systems. As a nation, we need to know various characteristics of our population to monitor our progress and allocate resources. We may have a relatively simple question, "How many people are there in the United States with disabilities?" but the answer depends on (1) our definition of disability and (2) the mechanism we choose for collecting the information. Several sources of information about disability prevalence are available from the federal government, the most important of which are the Behavioral Risk Factor Surveillance System (BRFSS) operated jointly by the CDC and state health departments and the U.S. Census Bureau.

The Behavioral Risk Factor Surveillance System

In a report issued in conjunction with the development of Healthy People 2010, the CDC specifically

recommended consistent definitions, universal collection of disability status, and methodological studies to help provide more accurate and useful data. The CDC generated the very broad definition of disability based on the standardized questions below. Two disability-related questions are now a regular feature of the CDC-supported annual core questions that are used by all states (BRFSS, 2001 Survey Questions, p. 21):

1. Are you limited in any way in any activities because of physical, mental, or emotional problems? Yes/No/Don't know/Not sure/Refused
2. Do you now have any health problem that requires you to use special equipment, such as a cane, a wheelchair, a special bed, or a special telephone? Yes/No/Don't know/Not sure/Refused

Additional questions that collect more detailed information about disability and caregiving are offered in some years as optional modules (used by individual states or not, at their discretion).

Publicly available BRFSS data are weighted to be representative of the entire U.S. population and the population of individual states; analysis based on smaller geographic areas is sometimes possible by special arrangement with the CDC or states.

In the 2001 BRFSS, a total of 41,262 respondents said they were either limited, used equipment, or both (19.8%)—people who met definitions based on both questions, however, comprised only 4.8% of the population.

The Census Bureau Resources

The U.S. Bureau of the Census provides data on disability drawn from three primary sources: the decennial census of population, the Survey of Income and Program Participation, and the Current Population Survey. Although the U.S. Census, conducted at 10-year intervals, is not primarily a health surveillance tool, it does provide denominator data for use in population research, policy and political analysis, and resource allocations. In 1990 and 2000, the long form of the census (administered to about one sixth of the U.S. population and weighted to be representative of the entire population) included questions on functional status and disability, permitting better estimation of disability prevalence. The Census 2000 questions addressed instrumental and basic activities of daily living (i.e., self-care, mobility), activity limitations, working, and sensory impairments. These data, available in summary form on the U.S. Census Bureau Web site, show, for example, that 19.3% or 49,746,248 Americans aged 5 years and above were classified as having a disability in 2000, and 56.6% of people aged 21 to 64 years with a disability were employed compared with 77.2% of other Americans. Census data have the advantages of representing the entire U.S. population (as opposed to the BRFSS, which considers only community-dwelling adults above the age of 18 years and has other limitations because it is conducted by telephone) and being representative of and applicable to small geographic areas—as small as neighborhoods. In fact, studies focusing on small local area comparisons may need to rely on census data because of the difficulty of using state (BRFSS) and national data at smaller geographic regions. For example, see the U.S. Census Bureau's county-level report on disability for Florida.

Methodological Issues in Disability Epidemiology

Measurement Issues (Classification and Bias)

One of the benefits of using an ICF classification system is that it allows combining diagnostic groups, which may overcome the problem of small sample size. For example, we might combine specific diagnoses into functional categories, such as mobility, communication, or learning/cognition impairments. This has the advantage of combining conditions that are related in their consequences to the individual as well as increasing the number of cases available in each category and thus the power of the analysis. However, the disadvantage of combining diagnoses in this manner is that heterogeneity of exposures and outcomes bias relationships toward the null, that is, to a finding of no effect. For a study on etiology, combining people with different impairments may obscure the cause of the impairment. For instance, Parkinson's disease, multiple sclerosis, HIV/AIDS, and spinal cord injury all lead to mobility impairments but entail very different etiology. Grouping people with mobility impairments in a study of etiology fails to take into account the very different causal pathways that can produce this outcome. However, for a study on secondary conditions, or outcomes research, this grouping of impairments might be fine as long as proper consideration was paid to other characteristics and risk factors, such as severity of disability.

Ecological Perspectives

Epidemiologists tend to eschew research where associations are made between variables based on grouped data, that is, those that are "ecological" in design. In a purely ecological study, we compare people at an ecological level in terms of exposure and outcome. The problem of inferences about cause-and-effect relationships from grouped data is called the ecological fallacy: Often, conclusions drawn from the analysis of grouped data do not hold when considered at the individual level.

However, some exposures really do operate at an ecological level, including many that have a potential impact on PWD, such as disability awareness media campaigns, laws, severity of winter weather, and wheelchair-accessible public transportation. In cases such as these, there is strong justification for applying a combined ecological and individual-level study design. The Institute of Medicine recommended combined (personal and environmental) models of public health training and thinking as a high priority. Apart from a need for more complex statistical analysis (i.e., mixed variance estimates), there are relatively few barriers to multilevel studies in disability epidemiology. This type of analysis is facilitated by the fact that the ICF includes codes for specific aspects of environment as experienced by individuals. Measures and models that are truly multilevel can provide a very rich area of research to understanding variations in levels of participation of PWD. However, at present, this kind of work is in its infancy, and few disability-relevant ecological-level measures have been defined.

—Elena M. Andresen

See also Health, Definitions of; Health Disparities; Healthy People 2010; Quality of Well-Being Scale (QWB)

Further Readings

Center for Disease Control and Prevention. (1998). *Healthy people 2010 disability objectives: Private sector and consumer perspectives*. Alexander, VA: Author. Retrieved February 5, 2007, from http://www.cdc.gov/ncbddd/dh/publications/hpmarge.htm.

Centers for Disease Control and Prevention. (2001). *Behavioral risk factor surveillance system, 2001 survey*. Retrieved March 9, 2007, from http://www.cdc.gov/brfss/questionnaires/pdf-ques/2001brfss.pdf.

Centers for Disease Control and Prevention, National Center for Birth Defects & Developmental Disability. (2006). *Healthy people 2010 objectives for people with disabilities*. Retrieved February 5, 2007, from http://www.cdc.gov/ncbddd/dh/hp2010.htm.

Centers for Disease Control and Prevention. *Disability and health state chartbook—2006: Profiles of health for adults with disabilities*. Retrieved February 5, 2007, from http://www.cdc.gov/ncbddd/dh/chartbook.

Hough, J. F. (2003). Fundamentals of disability and epidemiology. In *Disability and public health curriculum*. Oregon Office on Disability and Health. Oregon Health and Science University. Retrieved February 5, 2007, from http://cdrc.ohsu.edu/oodh/publications/Module3FundamentalsofDisabilityandEpidemiology.htm.

Last, J. M. (1995). *A dictionary of epidemiology* (3rd ed., pp. 55–56). New York: Oxford University Press.

Lollar, D. J., & Crews, J. E. (2003). Redefining the role of public health in disability. *Annual Review of Public Health, 24*, 195–208.

National Organization on Disability. (2006). *Statistics and surveys*. Retrieved February 25, 2007, from http://www.nod.org.

U.S. Census Bureau. (2006). *County-level report of census disability for Florida*. Retrieved March 9, 2007, from http://ebs.phhp.ufl.edu/FloridaChartbookDisabilityandHealth.pdf.

World Health Organization. (2001). *International classification of functioning, health and disability*. Retrieved February 5, 2007, from http://www3.who.int/icf/icftemplate.cfm.

Web Sites

Behavioral Risk Factor Surveillance System: http://www.cdc.gov/brfss/index.htm.

United Nations Statistical Division, Human Functioning and Disability, provides global statistics and data collection efforts: http://unstats.un.org/unsd/demographic/sconcerns/disability/disab2.asp.

U.S. Census Bureau: http://factfinder.census.gov.

U.S. Census Bureau reports on disability: http://www.census.gov/hhes/www/disability/disability.html.

Disaster Epidemiology

The epidemiologic study of disasters is a relatively new area of research and practice in public health settings. The epidemiologic investigation of disaster events focuses on two approaches. The first is the study of the underlying causes of the disaster. This may focus on the event itself or the mortality and morbidity associated with the event. Learning as much as possible about the reasons for disasters is

important for developing population-based prevention activities in the future. The second approach is to use epidemiologic methods to investigate mechanisms for alleviating the burden of a disaster once it occurs. The most direct applications of epidemiology in this situation are the establishment of surveillance systems to identify injuries and the possible emergence of communicable and mental health diseases, the deployment of rapid needs assessment to identify and prioritize solutions to existing problems, and analytical studies of risk factors and the natural history of health events.

Defining Disasters

What is a disaster? One of the most difficult concepts in the discipline is to arrive at a definition of a disaster. The answer to this question is shaped by many factors. Historically, in many areas, disasters have been viewed from a fatalistic perspective. Disasters were or are accepted as a feature of life. In the opinion of many individuals, there is little that one can do to prevent a disaster—it is an "act of God." In recent years, though, there has been a paradigm shift with the perspective in public health and other settings that disasters are something that one should prepare for to mitigate the circumstances arising from a disaster.

The frequency of an event and the level of magnitude of its impact can also influence whether an event is regarded as a disaster or not. Events with a low frequency in occurrence and a high magnitude of impact (in terms of large economic and human losses) are usually declared disasters by government authorities. Events with a high frequency of occurrence and a low magnitude of impact might be regarded as normal or routine events. The determination of what levels are high and what levels are low, though, can be subjective and may vary by culture, prior history with the event, and the ability to respond to the event. Thus, a disaster of similar characteristics might be viewed differently in different settings. Recent efforts have taken place to begin to standardize our view of disasters. The Center for Research on the Epidemiology of Disasters (CRED) in Brussels, Belgium, and other international agencies now typically use the United Nations Department of Humanitarian Affairs definition of disasters as a situation or event that "overwhelms local capacity, necessitating a request to a national or international level for external assistance" (EM-DAT glossary, unpaginated).

Classifying Disasters

Many different types of events can lead to situations that overwhelm local institutions and require external assistance. As one result, crude classification schemes have evolved in the discipline to classify disasters. Most commonly, disasters are classified as either natural disaster events or man-made disaster events. Natural disasters include situations brought about from extreme climatological, geological, or ecological events. Drought, flood, windstorms, extreme temperature, or extreme rain are the most common climate issues that have led to disasters in the past. Earthquakes, tsunamis, and volcanoes are frequency geological-related disasters. Man-made disasters include industrial accidents and acts of terrorism from nuclear, biological, chemical, or explosive materials. Recent episodes of large population displacement related to war have also been identified as a form of a disaster, largely a man-made disaster.

The Impact of Disasters

Much attention has focused on recent high-magnitude disaster events around the world. A proper assessment of the impact of disasters requires a comprehensive look at the totality of disaster events. The best resource to obtain this comprehensive look is through the databases maintained at the CRED. The main database includes disaster events from 1900 to the present, identified primarily from the information of relief and assistance groups, including the Red Cross/Red Crescent Agency. The data in this resource indicate that there are more than 100,000 deaths each year globally from disasters and about 70,000 to 80,000 injuries. The largest disasters from a human suffering viewpoint are droughts/famines. Significant numbers of deaths and injuries are also attributed to earthquakes, floods, cyclones, hurricanes, and tornadoes.

One recent debate has centered on the question of whether disasters are increasing in frequency compared with earlier years. Comparing data over time can be difficult, as the definition of disasters have varied over time. Generally, though, there is a trend indicating a higher frequency of disasters in these recent years. This trend is primarily due to the influence of several factors that affect the development of a disaster. The risk for disaster occurrence and/or the risk for heightened mortality in the event of disaster are

shaped by population growth, population density, environmental degradation, poor or unplanned urbanization, and poverty. Data indicate that the greatest degree of mortality and morbidity from disasters occur in low- and middle-income countries. The degree of calamity associated with a disaster will also be associated with the population density of the area affected and the level of vulnerability in that area. Events occurring in areas with dense population will result in greater harm (by absolute numbers) than events in less dense areas. Similarly, hazards occurring in areas made vulnerable by poor economic development, environmental degradation, or urban planning will result in greater harm than those occurring in stable areas. Vulnerable areas include river watersheds, undefended coastal plains, and hillsides prone to landslides. Many low-income countries have large populations living on vulnerable ground. The intersection of a hazard, high population, and high vulnerability results in a major catastrophe.

Health and Disasters

Disasters can influence human health in many ways. The largest impact of most disasters on human health lies in the injuries that occur from the event itself. In general, disaster events that involve water (such as floods, storm surges, and tsunamis) are the most significant in terms of mortality. The frequency of mortality in these events exceeds the frequency of nonfatal injury. In contrast, earthquakes and events associated with high winds tend to exhibit more injuries than deaths. Injury patterns related to man-made disasters are much more variable in the ratio of deaths to nonfatal injuries.

Health concerns also lie in the circumstances related to the aftermath of a disaster. In most disaster situations, outbreaks of communicable disease are not the primary concern in the short term. It is the view of many disaster professionals that the risk for outbreaks will not lie immediately after an event, but rather 1 to 2 weeks later and only if substantial population displacement and the disruption of health services occur. What is often the primary concern in the aftermath of a disaster is the impact of population displacement. Natural and man-made disasters will often destroy sizeable amounts of property, including houses and farms. From a health perspective, one is concerned with the effect of having little or no shelter (environmental exposure) foremost and overcrowding in available shelters. In the long term, there may also be a health concern over the ability to feed the population affected adequately. Mental health consequences of a disaster are another important health issue that carries both short-term and long-term implications. Several studies demonstrate higher frequencies of depression and posttraumatic stress disorder following major disasters.

Epidemiologic Response to Disasters

Epidemiology and the related methods of epidemiological practice are increasingly being recognized as valuable components to disaster response and disaster planning. The main goal of epidemiology in a disaster situation is to measure and describe the frequency of health events related to the disaster, to identify the factors contributing to these effects, and to identify potential interventions to alleviate the impact of these issues. Rapid needs assessments and surveillance activities are common practices undertaken in the aftermath of a disaster to address this broad goal. Further epidemiological studies may be conducted to identify risk factors, prioritize health interventions, match resources to needs, or to evaluate an intervention's effectiveness.

Epidemiology can contribute toward the understanding of the management and preparedness for disasters. This contribution can be directed at identifying and assessing factors related to the development of disasters, the public health response to disasters, an examination of the health effects of disasters, and the identification of groups in the population at particular risk for adverse health effects. Disasters are complex events, and practitioners in disaster epidemiology activities face many challenges, including establishing communications with professionals from several disciplines, a changing social environment in the face of destruction, a changing political environment, and constraints on the collection and analysis of data.

The analysis of past disasters provides several clues to the reduction of mortality and morbidity in future events. Unique patterns of death and injury have been noted among different classifications of disasters. Future research in the epidemiology of disasters will likely focus on improving the surveillance of mortality and injuries related to disasters, enhancing our understanding of the long-term chronic health effects of disasters, and approaches to integrating epidemiological information into contingency and mitigation planning.

—Thomas Songer

See also Bioterrorism; Epidemiology in Developing Countries; Governmental Role in Public Health; Injury Epidemiology; War

Further Readings

International Federation of Red Cross and Red Crescent Societies. (1998). *World disasters report*. New York: Oxford University Press.

Noji, E. K. (1997). The use of epidemiologic methods in disasters. In E. K. Noji (Ed.), *The public health consequences of disasters* (pp. 21–35). New York: Oxford University Press.

United Nations Department of Humanitarian Affairs. (1992). *International agreed glossary of basic terms related to disaster management.* Geneva, Switzerland: Author.

United Nations Development Programme. (2004). *Reducing disaster risk: A challenge for development*. New York: Bureau for Crisis Prevention and Recovery, United Nations.

Web Sites

Center for Research on the Epidemiology of Disasters: http://www.cred.be.

EM-DAT Emergency Disasters Data Base—Glossary of definitions from the UNDHR International Agreed Glossary of Basic Terms Related to Disaster Management: http://www.em-dat.net/glossary.htm.

Discriminant Analysis

Discriminant analysis (DA) is a multivariate statistical method used for two purposes: separation of observations into two or more distinct groups and classification of new observations into known groups. In DA, the categorical variables or groups are the dependent variables and the responses are the independent variables, so it is the reverse of a multivariate analysis of variance (MANOVA). Since the two procedures are computationally similar, the same assumptions that apply to MANOVA also apply to DA. Briefly, the assumptions are that the data are normally distributed and the variance/covariance matrices are homogenous across groups. DA is also sensitive to the presence of outliers and multicollinearity among the independent variables.

Separation

DA is used as an exploratory procedure in research to gain a better understanding of reasons for observed differences between groups. For instance, a researcher may be interested in determining which body types are at greater risk for heart disease. The researcher records several measurements (such as height, weight, and cholesterol levels) or response variables on patients who have heart disease and those who do not. The researcher wants to determine how much each measurement contributes to the separation of or discrimination between the groups. The researcher may also wish to classify new patients into risk-level groups given their body measurements.

When DA is used for separation, new variables called canonical discriminant functions (CDF) are created that combine the existing response variables in such a way as to maximize the variation between groups. A CDF is a linear combination of the response variables of the form

$$S_i = l_{i1}x_1 + l_{i2}x_2 + \cdots + l_{ip}x_p,$$

where S_i is the score for ith function, l_{in} is the standardized coefficient for x_n, $n = 1, \ldots, p$. There is one CDF for each independent variable or one for the number of groups minus one, whichever number is the smaller of the two. For example, if there are three groups and five variables, two CDFs are generated. The first CDF explains the greatest percentage of variation between groups. Each successive CDF is independent of the previous function and explains less variance. The characteristic root or *eigenvalue* associated, with the CDF indicates the amount of variance explained by the function.

The magnitude of the CDF coefficients indicates how important each variable is to group discrimination relative to the other variables. However, these coefficients may be misleading in the presence of correlation between the responses, that is, if there is a high degree of collinearity. Just as in regression analysis, addition or deletion of a variable can have a large effect on the magnitude of the other coefficients. Furthermore, when two variables are correlated, their contribution may be split between them or one may have a large weight and the other a small weight.

The canonical structure matrix can also be interpreted. This matrix contains the correlations between the CDF scores and each individual variable. The larger the correlation, the more important the variable is to discrimination.

Classification

There are several ways to use DA for classification. One way is to use Fisher's linear classification functions (LCF) of the form

$$G_j = c_{j0} + c_{j1}x_1 + c_{j2}x_2 + \cdots + c_{jp}x_p,$$

where j indicates the group, c_{jn} is the unstandardized coefficient for the jth group and the nth variable, $n = 1, 2, \ldots, p$, and c_{j0} is the constant for the jth group. There is one LCF for each group, and subjects are classified into the group with the highest discriminant score, G_j.

It is also possible to take a Bayesian approach to DA. This differs from Fisher's method, because it takes into account a priori or prior probabilities of group classification and seeks to minimize the probability of misclassification. Equal prior probability is usually specified when there is no prior knowledge about group proportions. It means that a subject has an equal chance of being assigned to one group or another. For example, without knowing anything about the sample, it may be assumed that a subject has the same chance of having heart disease as he has of not having it. When it is known that the prior probabilities are not equal, they should be specified in the analysis. For instance, it may be known that the probability of a subject having heart disease is 0.4 and the probability that subject does not is 0.6. The classification functions will be more accurate if prior probabilities are proportional to the group sample size.

Subjects are classified into groups using posterior probabilities. Posterior probability is the probability that a particular subject belongs to a certain group given knowledge of the values of the other variables for that subject. It is calculated using the discriminant scores and prior probabilities. A subject is assigned to the group that has highest posterior probability. The maximum likelihood rule for classification is a special case of Bayes's rule where prior probabilities are equal, and therefore, only posterior probabilities are considered when classifying observations.

Classification can also be thought of in terms of decision theory. Suppose that the cost of misclassifying some observations is greater than the cost of misclassifying others. Such a classification rule, therefore, minimizes the expected cost of misclassification. The obvious advantage of such a rule is that varying degrees of error can be taken into account. For example, when classifying individuals as diseased or healthy, it is generally considered more important to correctly classify the diseased individuals than it is to correctly classify the healthy individuals; that is, a greater cost is associated with failing to identify a case of disease where it exists than with incorrectly identifying a healthy person as having a disease.

Once classification functions have been estimated, their performance is assessed by applying the functions to a new set of data, that is, using them to classify different data than was used to estimate the functions in the first place. This is referred to as cross-validation. The accuracy of the classification functions is evaluated by the number of correctly classified subjects. It is not recommended to validate the functions by classifying the same data that were used to estimate the functions for obvious reasons: Doing so capitalizes on chance elements of the original sample that are unlikely to be present in any subsequent samples, thus producing an inflated predicted performance. Classifying the original data can, however, be used to look for outliers and to examine areas where the classification functions do not work well.

—Mary Earick Godby

See also Bayesian Approach to Statistics; Decision Analysis; Multivariate Analysis of Variance; Regression

Further Readings

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice Hall.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. San Diego, CA: Academic Press.

Disease Eradication

Disease eradication is defined as the permanent reduction of disease incidence to zero, globally, through deliberate efforts. On eradication of a disease, no disease-related morbidity or mortality can ever occur again. Disease eradication and disease elimination are not synonymous. In disease elimination, incidence of a disease is reduced to zero within a specific geographic area. Continued preventive measures are required in a state of disease elimination since the disease may still arise (i.e., importation of a communicable

disease across country borders), whereas intervention is no longer required on eradication. Eradication is different from extinction, which occurs when the etiologic agent no longer exists in nature or in a laboratory. Smallpox is the only disease to date that has been eradicated.

Epidemiological Criteria to Achieve Eradication

The biological and technical feasibility of eradication of a particular disease depends on the natural history of the etiologic agent and the disease, population characteristics affecting transmission potential, and the availability of diagnostic and intervention measures. Epidemiological criteria that favor disease eradication include the following.

Lack of Nonhuman Reservoir

The presence of a pathogen in nonhuman reservoirs, such as soil or animals, reduces the likelihood of eradication. For instance, while the rabies virus in humans can be contained through either a preventive vaccine or postexposure treatment, wild animals such as bats or raccoons can introduce the virus into human populations. Mass vaccination of animals and other attempts at controlling all natural reservoirs of rabies are not globally feasible, ensuring that rabies is not a viable candidate for eradication.

Sensitive Surveillance

Effective surveillance to detect the circulation of disease in a population is crucial to eradication efforts. Although largely dependent on the existence of well-functioning health systems infrastructures, surveillance in an eradication context is also promoted by favorable disease characteristics (such as a predictable seasonality of incidence, visible symptoms, or short incubation period) and available technologies (inexpensive, rapid, and accurate serological tests). The World Health Organization (WHO) campaign to eradicate yaws in the 1950s failed in part due to the inability to identify infected persons. Because yaws can exist asymptomatically in a latent state, case finding was hampered and many such carriers relapsed to infectious states even after treatment teams had visited a community. Surveillance can also serve as an indicator of progress in eradication efforts. The first 10 years of the yaws campaign had little, if any, screening, and when serological surveys were finally conducted, the discovery of high prevalence of subclinical infections rendered the eradication campaign essentially futile.

Effective Interventions

Interventions for preventive or curative treatment are necessary to remove potential susceptible or infected persons from a population. In addition, disease-specific means of eliminating vectors are crucial. The WHO malaria eradication campaign from 1955 onward was an uphill battle due to chloroquine resistance by parasites, fostered in part by inconsistent prophylactic treatment at subtherapeutic levels that promoted selection of resistant strains over time. A main component of the eradication strategy targeted the mosquito vector through use of insecticides, relying heavily on DDT. However, growing vector resistance to insecticides reduced the effectiveness of insecticide-based preventive measures and ultimately was a factor in the WHO decision in 1969 to revise the goal of malaria eradication to simply control.

A Safe, Highly Effective Vaccine for Vaccine-Preventable Diseases

Constraints prevent administration of a vaccine to all of the world's population, but principles of herd immunity indicate that for infectious diseases, the entire population does not need to be immune for a disease to be eliminated. The degree of herd immunity required to eliminate a disease depends on both disease and population characteristics and thus will differ between world regions. An effective vaccine exists for measles, but to overcome the high degree of contagion, countries need to achieve correspondingly high vaccination coverage levels to provide sufficient herd immunity. Vaccine characteristics, such as ease of administration, length of protection, side effects, and storage requirements, can determine the candidacy of a disease for eradication. The smallpox vaccination efforts were successful in part because the vaccine was inexpensive, safe even in newborns, heat stable, and required only a single dose. Moreover, smallpox vaccination conferred lifelong immunity and persons who were vaccinated had a recognizable scar.

Candidate Diseases for Eradication

The International Task Force for Disease Eradication (ITFDE), which convened from 1989 to 1992,

considered more than 90 diseases for eradication. After weighing epidemiological vulnerability as well as sociopolitical feasibility, the ITFDE concluded that six diseases were potentially eradicable: dracunculiasis (Guinea worm disease), poliomyelitis, lymphatic filariasis, mumps, rubella, and cysticercosis.

—Brian K. Lee

See also Herd Immunity; Polio; Smallpox; Vaccination

Further Readings

Dowdle, W. R. (1998). The principles of disease elimination and eradication. *Bulletin of the World Health Organization, 76*(Suppl. 2), 22–25.

Henderson, D. A. (1998). Eradication: Lessons from the past. *Bulletin of the World Health Organization, 76*(Suppl. 2), 17–21.

Hinman, A. (1999). Eradication of vaccine-preventable diseases. *Annual Review of Public Health*, *20*, 211–229.

Molyneux, D. H., Hopkins, D. R., & Zagaria, N. (2004). Disease eradication, elimination and control: The need for accurate and consistent usage. *Trends in Parasitology, 20*(8), 347–351.

Recommendations of the International Task Force for Disease Eradication. (1993). *Mobidity and Mortality Weekly Report. Recommendations and Reports, 42*(RR-16), 1–38.

Doll, Richard
(1912–2005)

Richard Doll qualified in medicine in 1937, graduating from St. Thomas Hospital in London, but his main epidemiological work began when he joined the staff of the Medical Research Council in London following service in World War II. More than any other person, he was responsible for establishing smoking as the main cause of lung cancer. His early case-control study of the question with Austin Bradford Hill was not the first such study, but he and Hill followed up some 40,000 British doctors after collecting details of their smoking habits, a cohort study that was unique in its regular updating of the subjects' smoking habits. The strong dose-response relation between lung cancer and smoking, together with the high standard and careful assessment of the study's findings, played a major part in convincing people of a causal relation, and in turn helped to change smoking habits. Thus, in 1950, 80% of the men in Britain smoked, but by 2000, this had declined to less than 30%.

Doll was one of the first epidemiologists to investigate the health effects of irradiation. His follow-up with Court Brown in 1957 of 14,000 ankylosing spondylitis patients treated with radiation brought the first independent confirmation, after the report on atomic bomb survivors, that radiation could cause leukemia. The study has since been a major source of data on the dose-response relation of radiation and cancer. In 1954, long before the relevant advances in molecular biology, Doll and Peter Armitage adduced evidence for the multistage nature of carcinogenesis.

He was also responsible for classic studies of asbestos and of nickel refining, being the first to show a significant excess of lung cancer among asbestos workers. He also collaborated with others on the side effects of oral contraceptives.

In 1969, he was appointed to Britain's premier medical chair, the Regius Professorship of Medicine at Oxford, which brought new attention to the subject of epidemiology. He had a wide influence on the progress of the science and was extensively consulted, for example, in the setting up of the International Agency for Research on Cancer (IARC). He was an early promoter of clinical trials and of cancer registries and was an author of the first compendia of worldwide cancer incidence data, *Cancer Incidence in Five Continents*.

Apart from his contributions to science and health, Richard Doll is commemorated in Oxford by Green College, where he was the first warden, and by the two flourishing units that he helped found at Oxford University, the Cancer Research UK Epidemiology Unit and the Clinical Trial Service Unit.

—Leo Kinlen

See also Asbestos; Cancer Registries; Oral Contraceptives; Radiation; Tobacco

Further Readings

Doll, R. (2002). Proof of causality: Deduction from epidemiological observation. *Perspectives in Biology and Medicine, 46*, 499–515.

Doll, R., & Hill, A. B. (1954). The mortality of doctors in relation to their smoking habits. *British Medical Journal*, *1*, 1451.

Doll, R., Payne, P., & Waterhouse, J. (Eds.). (1966). *Cancer incidence in five continents*. Berlin: UICC (International Union Against Cancer)/Springer-Verlag.

Domestic Violence

See Intimate Partner Violence

Dose-Response Relationship

Dose-response is a term that describes a relationship between an exposure and the risk of an outcome. A dose-response relationship is one in which increasing levels of exposure are associated with either increasing or decreasing risk of the outcome. Demonstration of a dose-response relationship is considered strong evidence for a causal relationship between the exposure and the outcome, although the absence of a dose-response relationship does not eliminate the possibility of a causal relationship.

The increase in the exposure can be in its intensity or its duration. Exposure can be characterized in different ways such as the peak exposure; the duration of exposure at or above a set level; average exposure, which is a time-weighted average of exposure; or cumulative exposure, which is the sum of time-weighted exposures.

The time to response must be considered when examining the relationship of the exposure to the outcome as there may be a latent period between exposure and the outcome. If the effects of exposure are measured too soon after the exposure, no effect will be seen even in the case where the exposure causes the outcome. One example of this is the increased risk of leukemia after exposure to radiation.

Odds ratios or relative risks can be calculated for categories of increasing exposure each compared with a baseline exposure level, as shown in Table 1. The mathematical relationship of exposure to outcome may be linear, log linear, or follow some other pattern. There may be some level of risk even in the absence of exposure, or there may be a threshold dose below which no affect of exposure on risk is seen (Figure 1).

In some cases, the relationship between exposure and outcome may be U-shaped with high risk at both extremes of exposure, but lower with intermediate exposure (see Figure 2). One example of this is the relationship of vitamin A with birth defects. Increased risk of birth defects is seen not only with deficiency in vitamin A but also with excessive doses.

Table 1 Dose–Response Relationship of Increasing Exposure and Increasing Risk of Disease

Exposure Score	*Disease (+)*	*Disease (−)*	*Odds Ratio*
0	5	95	1.0 (Reference category)
1	10	90	2.11
2	7	43	3.09
3	5	20	4.75

A statistical test for trend can be performed to verify that any apparent trend in the data is statistically significant. The Cochran-Armitage Test is one test for a trend in a binary outcome (ill or not ill) and applies to a linear relationship between exposure and outcome. Another is the Mantel-Haenszel Test, an extension of the chi-square test for trend (see Schlesselman, 1982, pp. 203–206, for details).

Inclusion of small numbers in the groups at the extreme ends of the exposure distribution may lead to statistically unstable rates in these groups. This may affect the validity of an apparent trend. Also, these end categories sometimes include extreme values, and the results can be sensitive to these extreme values.

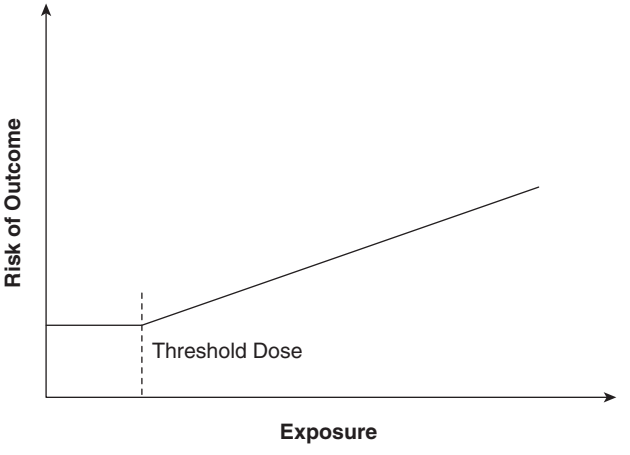


Figure 1 Threshold Dose

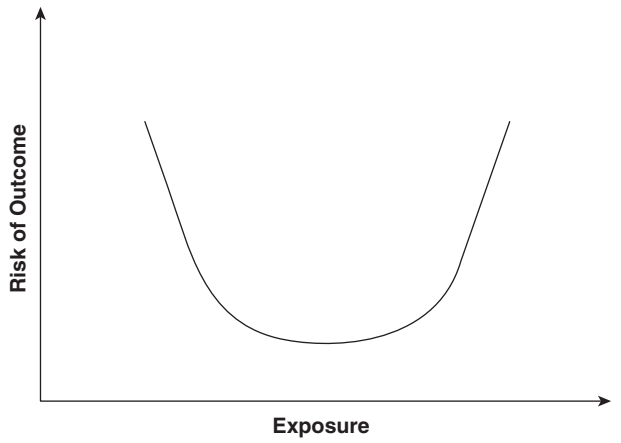


Figure 2 U-Shaped Relationship

For example, very few subjects may be included in the smoking exposure category "More than two packs per day," and this category may include a subject with exposures far in excess of anyone else in the study. For this reason, it is important to examine the effect of extreme values on the results.

—*Sydney Pettygrove*

See also Categorical Data, Analysis of; Causal Diagrams; Causation and Causal Inference

Further Readings

Rothman, K. J. (1986). *Modern epidemiology* (pp. 346–348). Boston: Little, Brown.

Schlesselman, S. (1982). *Case-control studies*. New York: Oxford University Press.

Doubling Time

Doubling time is an important concept in many life sciences and in epidemiology, in particular: It refers to the length of time required for some quantity to double in size. It is commonly used to refer to human population size, but it is also used to describe, for instance, growth of viral counts or antigens, tumors, and the number of cases of a particular disease. Suppose we have a quantity $Q(t)$ growing in time t. The function $Q(t)$ may stand for the population size or the amount of a substance at time t. Doubling time is the time it takes for $Q(t)$ to double in size and is uniquely determined by the growth rate $r(t)$. If $Q(t)$ has an exponential growth, then the doubling time can be exactly calculated from the constant growth rate $r(t) = r$. However, without the assumption of exponential growth or knowledge of the growth rate, one still can closely estimate the doubling time using two data values of $Q(t)$ at two time points. The estimation formulas are derived below:

The growth rate for $Q(t)$, assumed to be a continuous function of t, is defined as the relative change of $Q(t)$ with respect to t:

$$r(t) = \frac{1}{Q(t)}\frac{dQ(t)}{dt} = \frac{d \ln Q(t)}{dt}, \qquad [1]$$

where ln stands for natural logarithm ($\log_e$). Equation 1 indicates that the growth rate $r(t)$ has dimension "per unit time." Equation 1 is a simple linear first-order differential equation, which can be easily integrated to yield

$$Q(t) = Q(0)\exp\left(\int_0^t r(u)du\right) = Q(0)\exp(\bar{r}_t t), \qquad [2]$$

where $Q(0)$ is the quantity at time $t = 0$ and $\bar{r}_t = \int_0^t r(u)du/t$ is the average growth rate over the time interval $[0, t]$. Solving Equation 2 for t we obtain

$$t = \frac{\ln[Q(t)/Q(0)]}{\bar{r}_t}. \qquad [3]$$

The doubling time is the value of t, say, τ, such that $Q(\tau)/Q(0) = 2$, which when substituted into Equation 3 yields the following relation between doubling time and average growth rate:

$$\tau = \ln 2/\bar{r}_\tau = .693147/\bar{r}_\tau, \qquad [4]$$

which is approximately equal to 70% divided by the average growth rate, the so-called rule of seventy. Equation 4 says that the doubling time τ is inversely proportional to the average growth rate. The larger the growth rate, the smaller the doubling time. Doubling the growth rate is equivalent to halving the doubling time. For example, a short tumor-doubling time implies that the tumor is growing rapidly in size (measured either in terms of tumor volume or diameter). In fact, tumor-doubling times are considerably shorter in cancer patients who developed metastases than in those who did not develop metastatic disease.

In real applications, the average growth rate $\bar{r}_\tau$ in Equation 4 is unknown and has to be estimated from data in some time interval [0, T], where $Q(0)$ and $Q(T)$ are both known. From Equation 3, we have

$$\bar{r}_\tau \approx \frac{\ln[Q(T)/Q(0)]}{T}. \quad [5]$$

The closer T is to τ, the better the approximation in Equation 5. When $T = \tau$, the approximation becomes an equality. Substituting Equation 5 into Equation 4 yields an estimate of the doubling time as

$$\tau \approx .693147T/\ln[Q(T)/Q(0)]. \quad [6]$$

If the growth rate $r(t) = r > 0$ is a positive constant, then $\bar{r}_t = r$ and $Q(t)$ has an exponential growth. In this case, strict equality instead of approximation holds in Equations 5 and 6, as stated in the first paragraph. We now illustrate an application of formula (Equation 6) to estimate a doubling time in infectious disease epidemiology. During the outbreak of severe acute respiratory syndrome (SARS) epidemic in the early months of 2003, the health authority in Hong Kong had reported 300 SARS cases as of March 30 and reported 1,425 cases up to April 28. So we take the time interval [0, T] as [3/30/2003, 4/28/2003] so that $T = 29$ days, $Q(0) = 300$, and $Q(T) = 1{,}425$, which when substituted into Equation 6 yields 12.9 days as the doubling time—the period of time required for the number of cases in the epidemic to double—during the early stage of the SARS epidemic. If $Q(t)$ data are available for more than two time points, then the doubling time may be estimated using least-squares regression analysis.

Sometimes the components of the growth rate are available and so the growth rate may be calculated directly by subtracting the decrement rate from the increment rate. For example, in demographic applications, global crude birth and death rates are available, which according to the United Nations estimates are .022 per year and .09 per year, respectively. For the year 2000, yielding a growth rate of .013 per year, which when substituted into Equation 4 yields 53 years or approximately 54 (= 70%/.013) years, according to the *rule of seventy*, as the doubling time—the period of time required for the global population to double its size from 6.1 billion in 2000 to 12.2 billion in 2053—if the growth rate in 2000 were to continue.

—John J. Hsieh

See also Epidemic; Outbreak Investigation; Population Pyramid; Severe Acute Respiratory Syndrome (SARS)

Further Readings

Donnelly, C. A., Ghani, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Riley, S., et al. (2003). Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong [Electronic version]. *Lancet, 361*, 1761–1766.

Staab, H. J., Anderer, F. A., Hornung, A., Stumpf, E., & Fischer, R. (1982). Doubling time of circulating CEA and its relation to survival of patients with recurrent colorectal cancer. *British Journal of Cancer*, *46*, 773–781.

Steen, Bech-Nielsen, Reif, J. S., & Brodey, R. S. (1976). The use of tumor doubling time in veterinary clinical oncology. *Veterinary Radiology & Ultrasound, 17*, 113–116.

United Nations Population Division. (2001). *World population prospects*. New York: United Nations.

Drug Abuse and Dependence, Epidemiology of

Drug dependence is a disorder with neurobiological, social, and psychological influences. It is characterized by a pattern of maladaptive behaviors that develop as a consequence of the reinforcing effects of a drug that promotes continued use. Drug abuse and dependence have both direct and indirect impacts on individual health and also lead to social costs including those associated with drug-related crime. Although dependence is often a chronic, relapsing condition that accounts for a large public health burden, a large proportion of those with drug dependence remit even without treatment. This entry reviews the diagnoses of drug abuse and dependence, examines individual and environmental risk factors, discusses the individual and social costs of abuse and dependence, and considers options for the prevention and treatment of drug abuse and dependence.

Diagnoses of Drug Abuse and Dependence

Objective criteria to diagnose dependence can be found in the *Diagnostic and Statistical Manual of Mental Disorders*, published by the American Psychiatric Association, and the *International Statistical*

Classification of Diseases and Related Health Problems, published by the World Health Organization.

Dependence is diagnosed if at least three of the following problems are present: the user is tolerant to the substance (needs to use more to achieve the same effect or the same amount of substance leads to diminishing effect), experiences withdrawal (exhibits withdrawal symptoms or uses the substance to relieve those symptoms), needs to take larger amounts or for a longer time, wants or unsuccessfully tries to control his or her use, spends most of his or her time getting or using the substance or recovering from its use, gives up important activities to use the substance, or continues to use even with the knowledge that he or she has health problems related to substance use. Maladaptive use that does not meet the criteria for dependence is called abuse. Diagnostic criteria for abuse include failing to fulfill obligations at work, school, or home; recurrent use in physically hazardous situations; recurrent legal problems; or continued use despite serious social or interpersonal problems.

Although abuse and dependence are considered as distinct diagnostic entities, there is considerable overlap between these disorders; abuse of drugs usually involves using a large quantity of one or more substances, and regular use often, though not always, leads to dependence. On the other hand, a minority of those who are dependent do not show symptoms of abuse. In addition, it is more difficult to treat the abuse of those people who are highly dependent than those who are not as dependent. Research suggests that drug dependence changes not only the brain chemicals but also the brain structures, and the more dependent a person is, the more difficult it is to reverse the brain into a nondependent stage.

Distribution of Drug Dependence in the Population

Drug dependence is not distributed randomly in the population. Prevalence of dependence is highest among people aged 18 to 54 years, especially among young adults. The extent to which minorities are affected by dependence varies by drugs, region, and age. For example, in the United States, cocaine dependence is more common among whites who are below 30 years of age and among nonwhites who are above 30 years of age. People in geographic areas with economic deprivation and neighborhood disorganization are also more likely to develop dependence than people in more affluent areas. While treatment is available for drug dependence, only about one in five people with dependence gets treated.

Individual and Environmental Risk Factors

Individual Risk Factors

Many factors contribute to the vulnerability of developing substance abuse or dependence, and these factors stem from three general areas: (1) genetic predisposition, (2) nongenetic individual characteristics (psychological traits and demographics), and (3) environmental factors. The mechanisms through which addiction develops in these areas are (1) neurobiological (i.e., drug dependence is a brain disease), (2) disinhibition and reinforcement, and (3) barriers and protective factors. Results from studies among twins who have similar genetic risk factors but different environmental experiences suggest a genetic liability for dependence, similar to the role of genes in other chronic illnesses. There is likely no single gene responsible, but possibly several genes play a role in the neurobiological pathway to drug use disorders. A history of use of multiple drug types and low educational achievement may put one at risk for drug abuse and dependence. Individual choice, sensation seeking, and strategies of coping with stress may also play a role in drug use and abuse, which ultimately may lead to dependence.

Environmental Risk Factors

Although individual risk factors are important, it is also important to consider the effect of societal, social, and structural conditions, which can play both reinforcing and protective roles. For example, the number of family members, loved ones, and influential peers—an individual's social network members—who use or abuse substances may also influence whether or not an individual will use or abuse substances. Not living alone (e.g., living with parents or a spouse), high taxes on legal substances, and, in some cultures, religion have been found to be protective factors for drug dependence.

In addition, different drugs may be available in different countries, cities, or even neighborhoods within one city, and within these large and small geographic areas, social norms related to use and abuse also

influence an individual's chances and choices. Laws that determine which drugs are legally available and which are illegal are an important environmental factor. For example, alcohol is legal in most countries but is illegal in some, for example, Saudi Arabia and Kuwait; and marijuana is illegal in most countries but is legal in a few, for example, Holland. Furthermore, even within a country, legislation may have changed over the course of history to make certain drugs legal or illegal. For example, alcohol was illegal from 1920 to 1933 in the United States. In the 19th century, opiates were legally and freely available for purchase without a prescription in the United States. In the first decade of the 20th century, however, the production, sale, possession, and use of narcotics became illegal in the United States. In addition, again within a single country and within the same time period, certain drugs may be legal for one segment of the population, while they are illegal for another segment, as it is in the case of underage drinking and smoking. Structural conditions, such as these laws about underage drinking and smoking, are meant to protect a section of the society that is especially vulnerable to drug dependence; those who start using or abusing at an earlier age are more likely to develop dependence than those who start their substance use career at a later age due to their greater vulnerability to the neurobiological effects of substance use.

The proportion of people who use different drugs (the prevalence of drug use) and the prevalence of dependence among those who use these drugs vary by country and by drug. As a result, the potential of different drugs to cause dependence varies not only by drug but also by country or culture. For example, alcohol is readily available in most Western countries, but it is restricted in many Muslim countries. Cannabis is very common all over the world. Opiates, such as opium, heroin, or homemade liquid opiates, are available in most countries of the world, but they are rare in South America. On the other hand, cocaine and other coca derivatives are commonly used in South America and North America, but they are rarely used in Central and Eastern Europe or Africa. The reason for this may be that the world's top producer of heroin is Afghanistan and the leading cocaine producer is Colombia, although there is a growing production of heroin in Latin America. As both heroin and cocaine are illegal all over the world and are trafficked illegally, the availability of these drugs will be limited by geography and by local distribution networks.

Impact of Drug Characteristics and the Effect of Methods of Ingestion

Drugs differ in their potential to make people dependent. Tobacco has the highest likelihood of leading to dependence, followed by heroin, cocaine, and alcohol. Research among drug users in the United States has found that nicotine dependence develops in about one of every three people who smoke tobacco, compared with about one quarter of those who ever tried heroin and about 15% of those who ever tried cocaine or alcohol. Prescription medications, such as pain killers and sleeping pills, that are used initially for therapeutic reasons may also lead to dependence and abuse. The nonmedical use of prescription medications is an increasing public health concern in Western countries. It is estimated that about one in five people in the United States has misused prescription medications in his or her lifetime.

Pharmacokinetic properties of specific drugs and how they are used often affect how quickly one may become dependent. Users may orally ingest drugs in several ways—for example, by swallowing (e.g., pills), drinking (e.g., poppy tea), or chewing (e.g., coca leaves). However, oral ingestion is slow, and the speed with which a drug is metabolized and absorbed varies. Intranasal consumption, when drugs are absorbed through the mucous membranes of the nose, is also common because it provides a more rapid effect; drugs can be vaporized and inhaled in vapor form (e.g., "chasing the dragon"), dissolved and sprayed in the nose (e.g., "shebanging," or using cocaine mixed with water), or snorted in powder form (e.g., cocaine, heroin, and amphetamines). During smoking (e.g., marijuana), the drugs are absorbed into the lungs. Drugs can also be injected into the veins, under the skin, or into the muscle by means of hypodermic syringes and needles. The effect of the drug varies depending on the type of administration. For example, after snorting heroin, the users "get high" in about 10 to 15 min, whereas the effect of heroin can be felt about 7 s after smoking it or injecting it into the vein.

Access to and the cost of drugs may also influence what drugs people use and become dependent on. For example, the type of heroin powder available on the streets on the East Coast of the United States is white, while in Central Europe it is brown heroin. White powder heroin can be used several ways: It can easily be snorted up the nose in powder form, and it can also be dissolved in cold water and injected. Brown

heroin, on the other hand, cannot be snorted, but it can be dissolved in hot water and injected. As a result, there is a large portion of heroin users on the East Coast of the United States who use heroin via noninjecting ways, while almost all heroin users in Central Europe inject heroin. People in certain geographic areas or different socioeconomic groups may have easier access to certain drugs. Prevalence of a less expensive form of cocaine, crack, increased dramatically during the mid-1980s, particularly among urban African Americans in the United States. Although outbreaks of methamphetamine use in the form of "ice smoking" have been occurring in East Asia for several decades, the rapid spread in recent years of use and related dependence among "ice" users in the United States can be partly attributed to the ease in which the drug can be manufactured in local clandestine methamphetamine labs.

The Costs of Drug Dependence

It is estimated that drug abuse and dependence cost about $150 billion alone in the United States. The highest cost of drugs is mortality: The use of tobacco, alcohol, or illicit drugs contributes to one in four deaths in the United States, including deaths resulting from lung cancer and liver cirrhosis, with tobacco smoking being the leading preventable cause of death. In addition, much of the cost results from increased criminal activities committed as part of the lifestyle resulting from drug abuse or the nature of illegal drugs. These costs include, but are not limited to, those associated with law enforcement and incarceration of drug offenders. Low educational levels and unemployment are associated with substance dependence, but this is a two-way relationship. Not only are those with less education and no employment more likely to abuse substances, but substance abuse at an early age may also lead to failure to finish school and obtain a higher education, which in turn contributes to unemployment. Also, substance abuse itself can result in both unemployment and underemployment. In addition, among the employed, substance abuse and dependence, like any other chronic illness, can cause loss of work time and productivity.

Some substances have both direct health effects, such as liver cirrhosis resulting from alcohol dependence, and an indirect impact on health (e.g., "meth mouth," the premature and rapid decay and loss of teeth among methamphetamine smokers), mediated through high-risk behaviors associated with drug use. Drug abuse also speeds up the aging process, so aging-related illnesses are more common among those who abuse drugs. Among childbearing women, drug dependence can result in the birth of underweight, premature babies, who may already be physically dependent on the substance, or whose health is permanently damaged (e.g., fetal alcohol syndrome). Some drug users, both males and females, support their habit by engaging in prostitution, thereby increasing their risk of acquiring and/or transmitting sexually transmitted infections. Injecting drug users are at high risk for bloodborne infections, with HIV (the virus that causes AIDS) being one that is transmitted both sexually and through equipment contaminated with blood. The more dependent substance users are, the more likely they are to forego safe injecting or sex practices, since the severity of their dependence shifts their main focus from safer sex or injecting practices to obtaining and using substances. People who belong to more than one at-risk group are especially vulnerable. For example, males who inject drugs and have sex with other men, particularly those who sell sex to other men and thus have a large number of sex partners, are at the highest risk of getting infected or infecting others with HIV.

Mental illness is also correlated with drug dependence. Over one third of those who abuse alcohol and over half of those who abuse illicit drugs suffer from at least one mental illness, and about a third of those who have a mental illness also abuse alcohol or other drugs. Outcomes among people with dual diagnosis, that is, people who are both dependent on one or more substances and have a mental illness, are more severe than if only one disorder was present. Mental disorders often develop before the onset of drug dependence, but substance use itself can also lead to mental illness. To be able to treat people with dual diagnosis effectively, both psychiatric conditions need to be addressed simultaneously.

Prevention and Treatment of Drug Dependence

Prevention programs target the reduction of onset of drug use and also strive to stop the progression of use before it turns into dependence. Programs often target individual or group characteristics believed to have a causal influence on drug involvement (e.g., aggressive behavior, risk taking) but can also include environmental components (e.g., location of alcohol

outlets) and policies (e.g., taxes) targeting availability. Good prevention programs are research based; they are tested and evaluated similarly to how therapeutic drugs are tested and evaluated in clinical trials for efficiency and effectiveness.

Several treatment alternatives are available to those who want to abstain from drug use. Many dependent individuals can quit on their own; however, depending on the circumstances, professional help is recommended for detoxification, and residential programs can lead to favorable treatment outcomes. Counseling and support groups (e.g., 12-step programs) are widely used forms of treatment for dependence, and behavioral therapies, such as contingency management, have also shown great promise. Medications are also available for some drugs, for example, nicotine patches or chewing gums for tobacco dependence and methadone and buprenorphine for opiate dependence. A combination of behavioral therapy and medication, when available, is the most effective. The more treatment a person receives, the more likely he or she will be to recover. Harm reduction, such as needle-exchange programs to prevent bloodborne infections, and overdose prevention to prevent drug-related deaths are aimed at reducing the harm of substance abuse among those who are still active users. The support and encouragement of family and friends is an invaluable addition to any prevention, treatment, or harm-reduction program.

—V. Anna Gyarmathy, Carla L. Storr, and Howard D. Chilcoat

See also Harm Reduction; Illicit Drug Use, Acquiring Information On

Further Readings

Anthony, J. C., Warner, L. A., & Kessler, R. C. (1994). Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. *Experimental and Clinical Psychopharmacology, 2*(3), 244–268.

Brecher, E. M. (1973). *Licit and illicit drugs: The consumers union report on narcotics, stimulants, depressants, inhalants, hallucinogens, and marijuana—including caffeine*. Boston: Little, Brown.

Gaskell, L. (2000). *Drugs: Dilemmas and choices.* London: Royal College of Psychiatrists.

Heinz, A. (2006, April/May). Staying sober. *Scientific American Mind*, 57–61.

Jaffe, J. H. (1992). Current concepts of addiction. *Research Publication: Association for Research in Nervous and Mental Disease, 70*, 1–21 (Review).

McLellan, A. T., Lewis, D. C., O'Brien, C. P., & Kleber, H. D. (2000). Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *Journal of the American Medical Association*, *284*(13), 1689–1695 (Review).

National Institute on Drug Abuse. (2003). *Preventing drug use among children and adolescents. A research-based guide for parents, educators and community leaders* (2nd ed.). Bethesda, MD: U.S. Department of Health and Human Services.

Nestler, E. J., & Malenka, R. C. (2004). The addicted brain. *Scientific American, 290*(3), 78–85.

Office of National Drug Control Policy. (2001). *The economic costs of drug abuse in the United States, 1992–1998.* Washington, DC: Executive Office of the President (Publication No. NCJ-190636).

Web Sites

National Institute on Drug Abuse (NIDA): http://www.nida.nih.gov.

Drug Use, Illicit

See Illicit Drug Use, Acquiring Information On

Dummy Coding

Dummy coding, also known as indicator coding, provides a means for researchers to represent a categorical variable as a set of independent quantitative variables. The resulting dummy variables take on values of 0 and 1 and can be used as predictors in regression analysis. Given a categorical variable that can take on k values, it is possible to create $k - 1$ dummy variables without any loss of information. Dummy variables are often included in regression models to estimate the effects of categorical variables such as race, marital status, diagnostic group, and treatment setting.

When constructing a set of dummy variables, one level of the original categorical variable is selected as a reference category and is excluded from analysis. Each remaining level becomes a single dummy variable on which observations receive a value of 1 if

they fall into the category and 0 if they do not. For example, given a four-level psychiatric diagnosis variable (bipolar disorder, major depressive disorder, other mood disorder, no mood disorder), the "no mood disorder" group might be selected as the reference category. Three dummy variables would then be constructed: "BIPOLAR" would be scored 1 for persons with this diagnosis and 0 for those in the other three diagnostic groups; similarly, "MDD" would be scored 1 for persons with a diagnosis of major depressive disorder and 0 for those with another diagnosis; and "OTHMOOD" would be scored 1 for persons with a diagnosis of an "other mood disorder" and 0 for those diagnosed with bipolar disorder, major depressive disorder, or no mood disorder. Each dummy variable would have 1 *df*.

In regression analyses, the coefficients for each of the $k-1$ dummy variables quantify the estimated effect on the outcome variable of membership in the group in question versus membership in the reference group. For example, in a logistic regression analysis predicting diagnosis of bloodborne infection, a coefficient (β) for the "BIPOLAR" variable of 1.12 would represent the difference in the log-odds of infection between persons with bipolar disorder and those with no mood disorder. Expressed in terms of an odds ratio ($e^{\beta} = 3.06$), persons with bipolar disorder would have slightly more than three times the odds of infection compared with persons in the reference group. Dummy variables are used in a wide variety of regression analyses, including, but not limited to, ordinary least-squares, logistic, Cox, and Poisson regression.

The choice of reference category should be made based on substantive scientific considerations, and the category should be selected to support meaningful contrasts. In the case of race, this often results in the selection of "White" as the reference category. In experimental studies, the control group is often an appropriate choice. Miscellaneous or "other" categories containing a heterogeneous mix of observations (e.g.,"other race") usually do not provide researchers with the ability to make useful inferences and are therefore rarely suitable as reference categories. To ensure stable estimates, the number of observations in the reference category should not be too small relative to the size of the other categories.

—Scott M. Bilder

See also Categorical Data, Analysis of; Logistic Regression; Regression

Further Readings

Hardy, M. A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage.

Dummy Variable

A dummy, or indicator, variable is any variable in a regression equation that takes on a finite number of values so that different categories of a nominal variable can be identified.

The term *dummy* simply relates to the fact that the values taken on by such variables (usually values such as 0 and 1) indicate no meaningful measurement but rather the categories of interest.

For example,

$$X_1 = \begin{cases} 1 & \text{if Treatment A is used} \\ 0 & \text{otherwise} \end{cases}$$

and

$$X_2 = \begin{cases} 1 & \text{if female} \\ 0 & \text{otherwise} \end{cases}.$$

The variable X_1 indicates a nominal variable describing "treatment group" (either Treatment A or not Treatment A) and X_2 indicates a nominal variable describing "sex."

The following simple rule always is applied to avoid collinearity and the imposition of a monotonic dose-response in the regression model: For an exposure with K distinct levels, one level is first chosen as the baseline or reference group. Refer to that level as Level 0, with other $K-1$ levels referred to as Level 1, Level 2, and so on up to Level $K-1$. Then, define $K-1$ binary exposure variables as follows:

$$X_1 = \begin{cases} 1 & \text{if an individual's exposure is at Level 1} \\ 0 & \text{otherwise} \end{cases},$$

$$X_2 = \begin{cases} 1 & \text{if an individual's exposure is at Level 2} \\ 0 & \text{otherwise} \end{cases},$$

$$\ldots$$

$$X_{K-1} = \begin{cases} 1 & \text{if an individual's} \\ & \quad \text{exposure is at Level } K-1 \\ 0 & \text{otherwise} \end{cases}.$$

For example, in his book *Statistics for Epidemiology*, Nicholas Jewell notes that dummy variables are used for a variety of measures of variables in the Western Collaborative Group Study of risk factors for coronary heart disease in men:

$$X = \begin{cases} 1 & \text{Type A behavior pattern} \\ 0 & \text{Type B behavior pattern} \end{cases}.$$

Let $W_t =$ Body weight (1b), on continuous scale, and choose the baseline for weight $W_t \leq 150$. Then, define the following dummy variables:

$$Z_1 = \begin{cases} 1 & 150 < w_t \leq 160 \\ 0 & \text{otherwise} \end{cases},$$

$$Z_2 = \begin{cases} 1 & 160 < w_t \leq 170 \\ 0 & \text{otherwise} \end{cases},$$

$$Z_3 = \begin{cases} 1 & 170 < w_t \leq 180 \\ 0 & \text{otherwise} \end{cases},$$

and

$$Z_4 = \begin{cases} 1 & 180 < w_t \\ 0 & \text{otherwise} \end{cases}.$$

Another example shows how to use dummy variables to compare two straight-line regression equations: 40 males and 30 females are randomly selected to study the association of systolic blood pressure and age. The data set is presented in Table 1.

Table 1 Systolic Blood Pressure (SBP) by Age and by Sex

Sex	*SBP* (*Y*)	*Age* (*X*)
Male	158	41
Male	185	60
Male	152	41
Male	159	47
Male	176	66
Male	156	47
Male	184	68
Male	138	43
Male	172	68
Male	168	57
Male	176	65
Male	164	57
Male	154	61
Male	124	36
Male	142	44
Male	144	50
Male	149	47
Male	128	19
Male	130	22
Male	138	21
Male	150	38
Male	156	52
Male	134	41

Male	134	18
Male	174	51
Male	174	55
Male	158	65
Male	144	33
Male	139	23
Male	180	70
Male	165	56
Male	172	62
Male	160	51
Male	157	48
Male	170	59
Male	153	40
Male	148	35
Male	140	33
Male	132	26
Male	169	61
Female	144	39
Female	138	45
Female	145	47
Female	162	65
Female	142	46
Female	170	67
Female	124	42
Female	158	67
Female	154	56
Female	162	64
Female	150	56
Female	140	59
Female	110	34
Female	128	42
Female	130	48
Female	135	45
Female	114	17
Female	116	20
Female	124	19

(Continued)

(Continued)

Sex	*SBP* (*Y*)	*Age* (*X*)
Female	136	36
Female	142	50
Female	120	39
Female	120	21
Female	160	44
Female	158	53
Female	144	63
Female	130	29
Female	125	25
Female	175	69

Source: Adapted from Kleinbaum, Kupper, and Muller (1988).

A first-order regression model with an added interaction term for this example is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + E,$$

where

$Y = \text{SBP},$

$X_1 = \text{Age}, \text{and}$

$X_2 = \begin{cases} 1 & \text{Female} \\ 0 & \text{Male} \end{cases}.$

This single multiple regression model yields the following two models for the two values of X_2:

$$\begin{cases} X_2 = 0 : Y_M = \beta_0 + \beta_1 X_1 + E \\ X_2 = 1 : Y_F = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1 + E \end{cases}.$$

For the data set, the least-squares regression equation is

$$\widehat{\text{SBP}} = 110.039 + 0.961\,\text{Age} - 12.961\,\text{Sex} - 0.012\,\text{Age} * \text{Sex}.$$

For males, the least-squares regression equation is

$$\widehat{\text{SBP}} = 110.039 + 0.961 X_1.$$

For females, the least-squares regression equation is

$$\widehat{\text{SBP}} = (110.039 - 12.961) + (0.961 - 0.012) X_1 = 97.078 + 0.949 X_1.$$

Further statistical hypotheses may be generated from this model. For instance, we may like to test the null hypothesis that the two regression lines are parallel, which is equivalent to $H_0 : \beta_3 = 0$. If $\beta_3 = 0$, the slope for females equals to the slope for males. The decision from an F test is that there is no statistical basis for believing that two lines are not parallel.

—*Renjin Tu*

See also Hypothesis Testing; Regression

Further Readings

Jewell, N. P. (2004). *Statistics for epidemiology*. Boca Raton, FL: CRC Press.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1998). *Applied regression analysis and other multivariable methods* (2nd ed., p. 264). Belmont, CA: Duxbury Press.

E

Eating Disorders

Eating disorders comprise a complex, multidimensional group of increasingly common yet poorly understood illnesses. These conditions are characterized by serious and possibly life-threatening disturbances in eating patterns, including both significant levels of food restriction and intense overeating that often involve food bingeing and purging. In addition to a set of extreme, ritualized eating behaviors, sufferers commonly exhibit depression, anxiety disorders such as obsessive-compulsiveness and body dysmorphic disorder, perfectionism, distorted body image, body-checking behavior, and constant weighing. In addition, drug abuse has been found to be prevalent among people with eating disorders. Denial of illness is also characteristic of sufferers as is the valuing of symptoms. Disruption in eating patterns may begin with casual dieting, which, when successful, causes the sufferer to fixate on powerful feelings of accomplishment and personal control leading eventually to an unhealthy relationship with food and one's body.

Typology of Eating Disorders

Three primary types of eating disorders have been described: anorexia nervosa, bulimia nervosa, and binge eating disorder.

Anorexia nervosa is defined by significant weight loss, refusal to maintain a healthy weight for one's height, an obsession with food, acute anxiety about gaining weight, irregular or absent menstrual periods in females, and a distorted body image. In addition to restricting consumption, anorexics often focus their diets on a few food items that they eat in limited quantities. Sufferers report self-medicating their anxiety and relishing the feeling of control that accompanies food restriction.

Anorexia is one of the most deadly mental illnesses, with a morbidity rate as high as 10% of sufferers. It is estimated that 2.5 million Americans experience anorexia, including 0.5% to 3.7% of girls and women. Onset is most frequent during the early or middle years of adolescence, a period of rapid physical and emotional change. While anorexics restrict food intake, they report being preoccupied with it, frequently thinking, dreaming, and talking about it, while obsessively watching others eat or preparing food that they themselves avoid.

Bulimia nervosa is estimated to affect 1.1% to 4.2% of females in the United States. The disease is defined by episodes of binge eating followed by purging. Binges are characterized by the consumption of double or triple the amount of daily needed calories but can involve consuming as many as 20,000 calories at a time. Purging usually involves regurgitation but can include excessive exercise, enemas, and laxative and diuretic abuse. It is used by sufferers as a symbolic emotional cleansing or outpouring of feelings that otherwise remain silenced. While some sufferers purge to rid themselves of excess calories, the cycle is almost always triggered by emotional stress. The binge-purge cycle becomes fixed as an emotional coping method, albeit an unhealthy one, to which most of a bulimic's day is devoted.

Binge eating disorder, although not a psychiatrically affirmed illness, has been widely described and is characterized by a continued pattern of intensive eating episodes over which the sufferer experiences little control. Eating during a binge tends to be rapid, occurs when the sufferer is not physically hungry, and leads to feeling painfully full. Sufferers, who commonly feel self-distain for their behavior, often hide obsessive eating patterns, but unlike bulimics do not engage in purging. Studies have shown that 2% to 5% of the U.S. population suffers from this condition.

The term *bulimarexia* is used to denote a mixed type of eating disorder characterized by the binge-and-purge cycle of bulimia and the serious weight loss typical of anorexia. Rather than reflecting a set of fixed categories that differentiate sufferers, the eating disorder typology represents areas of symptom concentration in what is actually a more dynamic and varied reality. In addition to symptom mixing, anorexia can develop into bulimia because of an inability of anorexics to hide their restrictive tendencies in front of friends and family.

Causes of Eating Disorders

Eating disorders are the product of multiple-interacting factors, including cultural, family, psychological, and genetic factors.

Cultural Factors

Middle-class American society is characterized by a culture of abundance in which food is readily available and the ability to wield control over intake is deemed an admirable quality. As contrasted with societies characterized by food scarcity, in which a plump figure is a sign of beauty, in America, a slender, androgynous body has become a valued female physique. Expectations of thinness are reinforced by the advertisement industry. Studies of magazine content show high numbers of ads and articles about food, including recipes and accounts of the newest chic dishes. At the same time, magazines frequently give voice to social angst about obesity while exhibiting images of perfectly sculpted bodies attained through the use of exercise machines, workout routines, and an endless array of miracle diets. The cultural celebration of thinness and stigmatization of obesity are critical factors in the emergence of epidemic levels of food-related disorders in the U.S. society.

Within the United States, eating disorders traditionally have been most concentrated among upper- to middle-class, white, teenage girls. Having first come to medical attention as a disease of this social group, eating disorders have been called the modern cultural equivalent of 19th-century hysteria, as well as a type of symbolic bodily corseting that parallels the mechanical corseting of an earlier era.

Family Factors

Changing family patterns, including the loss of extended family support, and increasingly blurry family role expectations are thought to contribute to the emergence of eating disorders. Issues of control in parent-child relations are also common contributing factors. Thus, it has been argued that middle-class girls use food restriction as a means of gaining control over lives on which others, especially parents, have had a dominating influence.

Psychological Factors

Sufferers of food disorders are often described as lacking a fully developed sense of self, as self-blaming, and as people inclined toward pleasing others more than themselves. Moreover, because society has not fully accepted eating disorders as bona fide mental illnesses, sufferers are often stigmatized and become objects of derision, further amplifying self-deprecation and denial of symptoms.

Genetic Factors

Research on anorexia among female twins suggests the importance of genetic factors in eating disorders. Anorexia has been found, for example, to be twice as common among both members of identical twin sets compared with fraternal twins.

Globalization and Eating Disorders

Although eating pathologies have been a peculiarly Western and middle-class phenomenon, in recent years, they have begun appearing in non-Western countries, as well as across socioeconomic classes, ethnicities, age groups (including among the elderly), and even genders. Recent analysis has focused on the role of globalization and the diffusion of American or other Western images of the "ideal" body type as

critical elements in the changing international face of eating disorders.

In recent U.S. research, the frequency of a range of eating pathology, including binge eating, restrictive eating, vomiting, and amenorrhea, has not been found to differ by ethnicity, although some symptoms, such as eating until uncomfortably full, is more common among whites and African Americas than among Latinas and Asian Americans. In addition to changes in the ethnic expression of eating disorders, the age of first symptom expression is declining rapidly. While the first appearance of anorexia in the past did not occur before age 13, today it is not uncommon among 9-year-olds.

Health Implications

There are multiple physical effects of eating disorders. Sufferers are often lightheaded and dizzy and can black out. They are commonly sad, irritable, moody, and experience memory loss and fainting. The heart of an eating disorder patient, which is starving along with the rest of the body, may have trouble pumping. This leaves the patient feeling weak and deathly cold. Their limbs often fall asleep due to poor circulation. Electrolyte shortages can cause heart palpitations and the patient may suffer from low blood pressure and a low heart rate. Losing drastic amounts of weight, especially in young people, slows the sex-hormone production rate, delaying puberty. Females who suffer from a food-related disease for prolonged periods of time can become infertile. Similarly, the muscles of food disorder sufferers in time begin to atrophy. Moreover, the bones of sufferers lack nutrients and calcium, leaving anorexics at a high risk for fracture.

Lack of fluids or dehydration can lead to organ failure, while death by starvation or related causes occurs in 10% to 15% of cases.

Treatment and Prognosis

Despite having a poor prognosis, eating disorders are treatable. Treatment includes a combination of methods to address the diverse aspects of the disease. Treatment usually involves a two-pronged approach that focuses both on weight restoration through adherence to a prescribed (and monitored) meal plan as well as on the patient's emotional turmoil with psychotherapy. Depending on the severity of the eating disorder, the patient can be treated in an inpatient clinic or through outpatient therapy. An extensive variety of therapeutic techniques and methods are employed in the treatment of individuals with eating disorders although best practices are in debate. The time period between the initiation of treatment and a return to ideal body weight and healthier eating patterns is approximately 1 year.

—Merrill Singer and Elyse Singer

See also Child and Adolescent Health; Health Disparities; Nutritional Epidemiology; Psychiatric Epidemiology; Women's Health Issues

Further Readings

Bordo, S. (1993). *Unbearable weight: Feminism, Western culture, and the body*. Berkeley: University of California Press.

Chernin, K. (1981). *The obsession: Reflections on the tyranny of slenderness*. New York: Harper & Row.

Claude-Pierre, P. (1999). *The secret language of eating disorders*. New York: Vintage Books.

Gremillion, H. (2003). *Feeding anorexia: Gender and power at a treatment center*. Durham, NC: Duke University Press.

National Institute on Mental Health. (2001). *Eating disorders: Facts about eating disorders and the search for solutions*. Bethesda, MD: Author.

Sacker, I., & Zimmer, M. (1987). *Dying to be thin*. New York: Warner Books.

Eco-Epidemiology

Eco-epidemiology is an emergent area in the evolution of modern epidemiology, rooted in a new public health scientific paradigm that postulates an integrated approach to investigating disease and its prevention by subsuming levels of causation, life course trajectories, kinds of causes, and types of diseases. By incorporating this way of thinking about causes at multiple levels of organization and within the historical context of both societies and individuals, eco-epidemiology advances the adoption of a unified framework to the different domains of the discipline, emphasizing the ties that bind epidemiology to public health and implying a major shift in what qualifies as rational public health practice.

The emergent era of eco-epidemiology arises from the escalating recognition of constraints and prevailing

criticisms of the current era of chronic disease epidemiology, with its dominant risk-factor paradigm. It also emerges from the growing strength of molecular epidemiology on the one side and of social epidemiology on the other side of the determinants of health model. A fundamental charge against the current paradigm of the present era is its general neglect of the social environment in which disease occurs: It conceives risk for disease as residing largely within individuals and their personal behavior. Under the single-level, risk-factor paradigm, questions about macrolevel social and physical environments or microlevel mediators and antecedents are difficult to frame: Inattention to context leads to a limited and precarious knowledge base for public health action. In addition, description of risk-factor/disease associations—increasingly related to small effects detection, particularly vulnerable to indeterminacy from confounding and bias—are afforded priority over the explanation of causal processes and linkages between them; hence the black box analogy.

Molecular epidemiology focuses on biological mechanisms of disease and social epidemiology on societal determinants of disease. Caught between biology and society, risk-factor epidemiology deals with the middle ground of behaviors and exposures. Eco-epidemiology, also known as *multilevel epidemiology*, recognizes these three levels of organization—the micro, the macro, and the individual—as equally fundamental to the purview of public health epidemiology. In other words, eco-epidemiology explicitly reminds that molecular, lifestyle, and societal explanations of disease are interconnected and reciprocally reinforcing, not mutually exclusive, competing alternatives to understand disease causation and to advance the cause of public health. More specifically, eco-epidemiology not only advocates this paradigm in the interpretation of findings from epidemiological research, but also by including into epidemiological study designs direct measures representing the disease process at each level.

Eco-epidemiology contends that fruitful theories of disease causation and pathogenesis can, in principle, be conceptualized at all levels of organization. Since detectable causes differ across levels, theories at different levels may each point to distinct understandings of disease and prevention. Eco-epidemiological study designs that incorporate individual-level exposures, group-level exposures, individual-level health outcomes, and group-level health outcomes that are explicitly defined and related to each other are notoriously complex—a reflection of a relentlessly multilevel, multicausal, multivariable world. It has been argued that this likely complexity may go against the epidemiologist's justified desire for parsimony. Conversely, in its unifying effort, it adds to the coherence of epidemiology. It also calls for a broader multidisciplinary thought collective and a greater methodological pluralism.

The paradigm shift that is shaping the emergent eco-epidemiology era has four basic premises:

1. Causes of disease occur at all levels of organization.
2. Causes of disease can be distinct at different levels of organization.
3. Different levels interrelate among each other in ways that can mutually influence the play of causes of diseases at each level.
4. At any given time, patterns of disease and health states are the result of dynamic antecedent processes; that is, causes of disease are historically contingent.

Consistent with the concept of emergent group properties, at each ascending level of organization, distinctive characteristics confined to that level emerge. The prevalence of an infectious disease in the population and the magnitude of income inequality are two classical examples of unique group-level attributes from the fields of infectious disease epidemiology and social epidemiology, respectively. Both prevalence of infection and income inequality are inherently unique group attributes and can be studied only at the group level, using individual-level characteristics to control for the effect of other influences. If these group-level variables are not included in the model, it will be impossible to estimate the contextual effects of infectious disease or social determinants of health in a valid way.

Proponents of eco-epidemiology, thus, recognize that epidemiology is in transition from a science that identifies risk factors for disease to one that analyzes the systems that generate patterns of disease in populations, by considering multiple levels of causation, investigating the interplay between genetic and environmental factors, examining the trajectory of health and illness over the life course, and proposing a broader and more unified framework to understand

health production. Eco-epidemiology addresses the interdependence of individuals and their connection with the biological, physical, social, and historical contexts in which they live. To do so, it encompasses the changeable contributions and effects on the individual level of both macrolevels (i.e., societal) and microlevels (i.e., molecular) of organization. Firmly rooted in the concerns of public health, the aim is to study multiple relationships across levels that would contribute to the expansion of the understanding of disease processes. It is expected that under this emergent paradigm shift, epidemiology will be rooted in the investigation of the pathways by which biological and social experiences generate health and disease, and will be equipped to identify the impact of biological and social changes on the health of the populations.

—Oscar J. Mujica

See also Causation and Causal Inference; Determinants of Health Model; Life Course Approach; Multilevel Modeling

Further Readings

Schwartz, S., Susser, E., & Susser, M. (1999). A future for epidemiology? *Annual Review of Public Health, 20*, 15–33.

Susser, E. (2004). Eco-epidemiology: Thinking outside the black box. *Epidemiology, 15*, 519–520.

Susser, M. (1998). Does risk factor epidemiology put epidemiology at risk? Peering into the future. *Journal of Epidemiology and Community Health, 52*, 608–611.

Susser, M., & Susser, E. (1996). Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *American Journal of Public Health, 86*, 674–677.

Ecological Fallacy

Relationships observed for groups do not necessarily hold for individuals, and vice versa. The *ecological fallacy* is a fallacy in ecological studies that may arise when an investigator makes an inference about an individual based on aggregate data for a group. In ecological studies, we assess the relation between exposure rates and event rates at group level because we know only marginal distributions of exposure (risk factor) and outcome event and not their joint distributions. Researchers have made unwarranted inferences from the association between exposure to risk factor and outcome event among groups (ecological data) to association among individuals within each group without accounting for the possible ecological bias. Aggregating data loses information. The ecological fallacy may arise because the process of aggregating data may conceal the variations that are not visible at the larger aggregate level (see explanation in Example 2 below). Statistically, a correlation tends to be larger when an association is assessed at the group level than when it is assessed at the individual level.

Example 1: Nativity and Literacy

Robinson calculated the correlation coefficient between nativity (represented by the percentage of the population who are foreign-born) and literacy (represented by the percentage of the population who are literate) for the 48 states in the United States of 1930 to be .53, a positive correlation. This is an ecological correlation because the unit of observation and analysis is the state. But when computed at the individual level, the correlation coefficient turns out to be −.11, a negative coefficient! The fallacy arises because the foreign-born tend to live in states where the native-born are more literate.

Example 2: Blood Pressure and Stroke Mortality

In the Seven Countries Study, Menotti et al. (1997) found that the mean entry-level blood pressures and stroke mortality rates were highly inversely correlated for 16 cohorts of men aged 45 to 59 with 25-year follow-up. This is contrary to the expectation. So the analyses were repeated at the individual level within cohorts, the association between blood pressure and stroke mortality was then found to be strongly positive among most of the cohorts, and hence the correlation for all individuals should have been positive. The explanation of this paradox is that within each cohort, individuals who had had and had died from stroke tend to have had high blood pressure, but when the individual values in each cohort were averaged and the 16 pairs of average values were used to calculate the correlation, the cohorts with higher average blood pressures may have turned out to have smaller mortality rates simply because of the heterogeneity of correlations among the cohorts.

Example 3: Breast Cancer and Fat Consumption

Carroll found that death rates from breast cancer were significantly higher in countries in which fat consumption was high than in those in which fat consumption was low. This is an association for aggregate data, for the unit of observation is country. When inference is made to individual-level association, saying that if countries with more fat in the diet have higher rates of breast cancer, then women who eat fatty foods must be more likely to get breast cancer, an ecological fallacy may be committed because one cannot be certain that the breast cancer cases had high fat intakes. In fact, the problem of ecological fallacy on the link between breast cancer and fat intake was raised by Holmes et al. (1999) when they examined the individual-level data.

The ecological fallacies in the three examples above arise from assuming that all individuals in each ecological group have the same summary measure (the mean value) of the group without accounting for possible confounding by other variables and for the unobserved heterogeneity of individuals in each group. In view of this, when statistically significant association is found between exposure and health outcome at group level (usually aggregate data are easily assessable as they already exist, having been previously collected for other purposes), then individual-level data should be collected to obtain the joint distributions of exposures and outcomes. This would make it possible to test the ecological hypothesis thus generated so as to corroborate or refute the putative ecological association at the individual level. This is because for causal inference, individual data are required to account for population heterogeneity and confounding bias.

More generally, the ecological fallacy may be defined as the fallacy of drawing inferences regarding relationship for units defined at a lower level (such as individuals) based on data collected for units at a higher level (such as groups). In contrast to the ecological fallacy is the less well-known counterpart, the *atomistic fallacy*, which refers to the influence that associations found at the individual level that will necessarily hold at the group level or, more generally, the fallacy of drawing inferences regarding relationship for units defined at a higher level (such as groups) based on data collected for units at a lower level (such as individuals). The atomistic fallacy occurs when characteristics of higher levels are distributed to all lower-level units without taking into account the dependency of the observations among the lower units within each higher-level unit, so that factors that explain variability among lower-level units within higher-level units may differ from those explaining variability across higher-level units.

To avoid both ecological and atomistic fallacies, use may be made of multilevel analysis to separate the total variation in the outcome variable into that part due to variability among lower-level units and that due to variability among higher-level units so as to uncover the true relationship between exposure and outcome.

—John J. Hsieh

See also Causation and Causal Inference; Descriptive and Analytic Epidemiology; Multilevel Modeling; Unit of Analysis

Further Readings

Carroll, K. (1975). Experimental evidence of dietary factors and hormone-dependent cancers. *Cancer Research, 35*, 3374–3383.

Diez, R. A., Schwartz, S., & Susser, E. (2004). Ecological variables, ecological studies, and multilevel studies in public health research. In R. Detels, J. McEwen, R. Beaglehole, & H. Tanaka (Eds.), *Oxford textbook of public health* (4th ed., pp. 493–507). Oxford, UK: Oxford University Press.

Goodman, L. (1959). Some alternatives to ecological correlation. *American Journal of Sociology, 64*, 610–625.

Greenland, S., & Robins, J. (1994). Invited commentary: Ecologic studies—biases, misconceptions, and counterexamples. *American Journal of Epidemiology, 139*, 747–760.

Holmes, M. D., Hunter, D. J., Colditz, G. A., Stampfer, M. J., Hankinson, S. E., Speizer, F. E., et al. (1999). Association of dietary intake of fat and fatty acids with risk of breast cancer. *Journal of the American Medical Association, 281*, 914–920.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Menotti, A., Blackburn, H., Kromhout, D., Nisinen, A., Karvonen, M., Aravanis, C., et al. (1997). The inverse relation of average population blood pressure and stroke mortality rates in the Seven Countries Study: A paradox. *European Journal of Epidemiology, 13*, 379–386.

Morgenstern, H. (1982). Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health, 72*, 1336–1344.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Selvin, H. C. (1958). Durkheim's "suicide" and problems of empirical research. *American Journal of Sociology, 63*, 607–619.
Twisk, Jos W. R. (2006). *Applied multilevel analysis*. New York: Cambridge University Press.

Economic Evaluation

Economic evaluations involve the quantification of changes in health resource utilization due to the introduction of new courses of action. Policymakers are increasingly turning to such analyses to acquire information before making decisions about alternatives in health care. Such analyses are used by insurers to determine which services to pay for, and government policy analysts use technology assessments to shed light on the economics of new interventions and courses of action. Economic evaluations are used by policymakers and analysts to make systematic decisions concerning the allocation of resources in the market.

There is a growing literature on economic evaluation in health care. Although the studies vary in quality, several good introductions such as the ones by Drummond (1981) provide a basic interpretation of the nature of economic evaluation and an appreciation of the decision making required at all levels.

The Purpose of Economic Evaluations

Economic evaluations answer the following questions to provide an objective set of criteria for making choices among alternatives given scarce resources:

1. Are health services, and the like, worth doing given limited resources?
2. Are we satisfied with the way health resources are used among the different courses of action chosen?

The purpose of an economic evaluation is to compare alternative courses of action that are solutions to the same problem. Without systematic analysis, it is difficult to clearly identify the alternative uses for resources and the opportunity cost of employing one alternative over another in solving a problem. For example, Messonnier, Corso, Teutsch, Haddix, and Harris (1999) suggest that a health department may need to evaluate the efficiency of a diabetes prevention program or a bicycle helmet initiative in reducing the number of disability days in a population.

Cost

Costs are the value of the resources used for any particular course of action. The type and scope of costs depend on the analysis viewpoint (i.e., society, government, patient, employer, program agency). When in doubt, it is best to go to the broadest or societal viewpoint. The real cost of any alternative is measured not by the budgetary allocations but by the health output that could have been achieved through some other alternative that has been foregone because of the commitment of resources or inputs to the alternative in question. This cost is the opportunity cost of the alternative considered and is compared with the alternative's benefits.

Direct costs are the actual expenses incurred by participating in the alternative. This includes the medical expenses, transportation costs, and other training costs that can be part of implementing the alternative. Indirect costs are the productivity losses associated with the course of action, which reflect the opportunity costs of using one alternative and foregoing another. For example, this could include the waiting time for appointments and the transportation time as part of the participation in a course of action. In the private perspective, transfer costs are also included since they reflect changes in payments for individuals, providers, or organizations. From the societal view, direct and indirect costs are included but not transfer costs, since these are not resources used. Overall, economic costs go beyond simply listing expenditures, since opportunity costs need to be reflected.

Private Versus Societal Viewpoints

There are two general viewpoints to an economic evaluation: private and societal. The private perspective is focused on the individual, an organization, or a set of organizations. A health care organization may be interested in the cost benefit of a palliative care program versus traditional medical protocols. In this case, the firm is not interested in the transfer payments that may result by participation, since these are not paying for resources being used. Instead, the firm is concerned with its own direct and indirect costs of the courses of action and their associated outputs. The societal view includes all persons so that the

opportunity cost of the various courses of action can be taken into account. In terms of a palliative care intervention, this would include all direct and indirect costs of the courses of action and the transfer payments that may be involved as well, since they reflect the opportunity costs of pursuing one course of action versus another for the population as a whole. Economic evaluations link the alternative courses of actions' inputs and outputs and provide a comparative analysis of alternative courses of action in terms of both the value of their inputs and outputs. Without such an analysis, it is difficult to objectively justify the value for the money invested in an alternative.

Types of Economic Analyses

The type of output from the alternative courses of action can vary significantly across the methodologies.

Cost Minimization Analysis

In this case, outputs of the courses of action are assumed to be identical, and costs only are considered. For example, Evans and Robinson (1980) did a comparison of the common output of interest in the number of successful procedures at a day surgery center versus performing the procedures at an outpatient center of a hospital. Here, one may find an identical number of procedures performed but possibly different costs. The principal decision rule is focused on the costs per procedure successfully performed, where the least cost course of action is determined to be the efficient choice.

Cost-Effectiveness Analysis

In this case, the output of the courses of action is common across alternatives, but the alternatives have varying degrees of success in achieving the output. Examples by Gray and Elixhauser are the comparisons of different diabetes prevention programs. The decision rule is based on the cost per unit of output or output per unit of costs. The decision maker selects the course of action that yields the most output per dollar spent or the least cost per output. The latter decision is used when the decision maker is working within a given budget. This implies that there is a single, common affect that is constrained, and the alternatives are within the same range of scale. This analysis can be done considering any courses of action with a common output. The worth of the courses of action is assumed to be positive.

The outputs can be health effects directly or measures that show improvements in health status. For instance, one can compare a prevention program versus a chronic care program in terms of disability days saved per dollar invested in each program, as seen in the work of Hatziandreu, Koplan, Weinstein, Casperson, and Warner (1988), and Tengs, Adams, Pliskin, Safran, Siegel, and Weinstein (1995), and others. In cost-effectiveness analysis, there is a dominant dimension of success that is considered. It is important to be open to the possibility of using more sophisticated analyses, such as cost-benefit analysis, if there is more than one dimension of effectiveness.

In conducting a cost-effectiveness analysis, Drummond, O'Brien, Stoddart, and Torrance (1997) noted that several data issues should be addressed. First, the analyst must assure that there is a random allocation of patients to groups. Second, if the investigation is looking at existing literature, it is important to see how studies relate to provider expertise and patient caseload in question. Third, a sensitivity analysis, discussed later, can eliminate the need for clinical trials (especially in extreme effectiveness issues). However, if a clinical trial is used, the investigator must assure that the analysis of the clinical trial doesn't cause any deviation of normal working practices. Laupacis also notes that it is more meaningful if the results of the cost-effectiveness analysis are compared with some standard for the problem being investigated.

Cost-Utility Analysis

This is often considered a special case of cost-effectiveness analysis, where the output of the courses of action is valued commonly across alternatives, but the alternatives have varying degrees of success in achieving the value of the improvement in the output. In this case, both the output and the worth of the courses of action are measured. An example of such an analysis is the improvement in the quality-adjusted life years (QALYs) due to a diabetes intervention versus usual care. As Torrance and Feeney (1989) noted, this technique is preferred by many economists, since it incorporates the utility of the output, or in other words, the preferences of the patients or the population considered.

Utility is the value or worth of a specific health state and can be measured by the preferences of persons for any set of health states. Utility of the health output is different from the health output itself. It brings in quality-of-life adjustments for treatment output, while providing a common denominator for comparing the costs and outputs of different alternatives. The measure for utility is seen in the measures of healthy days or QALYs. Here, Sintonen (1981) and Williams (1981) note that the length of time of the health state is adjusted through a utility scale (0 to 1, with 0 = *death* and 1 = *perfect health*). The decision rule is to choose the alternative with the lowest cost per healthy year equivalent or QALY. Olsen (1994), O'Brien (1995), and Hirth, Chernew, Miller, Fendrick, and Weissert (2000) noted that willingness to pay for an additional QALY can be determined from community-based surveys. However, these surveys need to follow procedures similar to those for contingent valuation studies.

Cost-Benefit Analysis

In this case, the output of the courses of action may not be a single common effect but may be multiple effects that may or may not be common to the alternatives. For example, one could compare a health promotion program for youths with a chronic care intervention with the elderly on a variety of output dimensions. One could perform a cost-effectiveness analysis on multiple effects to determine a decision rule where an alternative is superior on all or a majority of dimensions or choose a primary effect to base the comparison.

Alternatively, one could develop a method to combine multiple effects into one common valuation. Here, the measure of value is the dollar, translating effects into the dollar value of benefits of life years gained, improved productivity, more convenience, and so on. This comparison of dollar costs to dollar benefits is cost-benefit analysis. This results in a ratio of dollar costs to dollar benefits or the sum of net social benefits, where net social benefits = social benefits − social costs. As Drummond (1981) noted, the decision rule is to choose the course of action that has the greatest net social benefits. Benefits will be large enough so that those who gain could theoretically compensate the losers, and everyone is made better off (i.e., the Pareto Principle). The preferred method is to maximize net benefits rather than benefit/cost ratio, since the ratio can be misleading depending on how benefits and costs are categorized.

The implicit assumption is that the courses of action are compared with a do-nothing alternative. However, in health care, since there are usually costs involved in do-nothing states, this is not usually done in practice. The valuation of the benefits can be done through the human capital method or the contingent valuation framework. The instrument depends on the purpose of the evaluation.

Viscusi's (1995) human capital method places a value on the opportunity cost of lost time, such as lost wages or the value of replacement workers for duties without a wage. For example, if a person is in the labor force and needs to take time off from work due to disability, then the value of the loss of work would be measured in the wage rate. If the person is out of the labor force and has a disability that reduces the level of productivity, then the value of the loss is the cost of the replacement worker who has to complete the tasks no longer completed by the person in question. In many ways, this approach is debatable among economists, since wages underestimate the total loss of time, particularly leisure time. Also, the approach favors the employed rather than those out of the labor market, which leads to inequities in compensation. Contingent valuation is what one would hypothetically pay if one could achieve the benefits from specific interventions.

Since the Pareto Principle is satisfied hypothetically, cost-benefit analysis traditionally doesn't account for income redistribution. Redistribution takes the form of taxes and transfers and can be criticized as inefficient. In practice, such as welfare reform plans, redistributional effects have been explicit, where the most general procedure is to classify the benefits and costs on a person-by-person or group-by-group basis.

—Diane Mary Dewar

See also Disability Epidemiology; Health, Definitions of; Health Economics; Quality of Life, Quantification of

Further Readings

Drummond, M. F. (1981). Welfare economics and cost-benefit analysis in health care. *Scottish Journal of Political Economy*, *28*, 125–145.

Drummond, M. F., O'Brien, B., Stoddart, G. L., & Torrance, G. W. (1997). *Economic evaluations of health care programmes.* New York: Oxford University Press.
Elixhauser, A. (1989). The cost effectiveness of preventive care for diabetes mellitus. *Diabetes Spectrum, 2*, 349–353.
Evans, R. G., & Robinson, G. C. (1980). Surgical day care: Measurements of the economic payoff. *Canadian Medical Association Journal, 123*, 873–880.
Gray, A., Raikou, M., McGuire, A., Fenn, P., Stevens, R., Cull, C., et al. (2000). Cost effectiveness of an intensive blood glucose control policy in patients with type 2 diabetes: Economic analysis alongside randomized controlled trial (UKPDS 41). *British Medical Journal, 320*, 1373–1378.
Hatziandreu, E. I., Koplan, J. P., Weinstein, M. C., Casperson, C. J., & Warner, K. E. (1988). A cost effectiveness analysis of exercise as a health promotion activity. *American Journal of Public Health, 78*, 1417–1421.
Hirth, R. A., Chernew, M. E., Miller, E., Fendrick, M. F., & Weissert, W. G. (2000). Willingness to pay for a QALY: In search of a standard. *Medical Decision Making, 20*, 332–342.
Laupacis, A., Feeny, D., Detsky, A. S., & Tugwell, P. X. (1992). How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Canadian Medical Association Journal, 146*, 473–481.
Messonnier, M. L., Corso, P. S., Teutsch, S. M., Haddix, A. C., & Harris, J. R. (1999). An ounce of prevention: What are the returns? *American Journal of Preventive Medicine, 16*, 248–268.
O'Brien, B. (1995). Principles of economic evaluation for health care programs. *Journal of Rheumatology, 22*, 1399–1402.
Olsen, J. A. (1994). Productivity gains: Should they count in health care evaluations? *Scottish Journal of Political Economy, 41*(1), 69–84.
Sintonen, H. (1981). An approach to measuring and valuing health states. *Social Science in Medicine, 15c*, 55–65.
Tengs, T. O., Adams, M. E., Pliskin, J. S., Safran, D. G., Siegel, J. E., & Weinstein, M. C. (1995). Five-hundred life-saving interventions and their cost-effectiveness. *Risk Analysis, 15*, 369–390.
Torrance, G. W., & Feeny, D. (1989). Utilities and quality adjusted life years. *International Journal of Technology Assessment in Health Care, 5*, 559–575.
Viscusi, W. K. (1995). Discounting health effects for medical decisions. In F. Sloan (Ed.), *Valuing health care: Costs, benefits and effectiveness of pharmaceuticals and medical technologies* (pp. 123–145). New York: Cambridge University Press.
Williams, A. H. (1981). Welfare economics and health status measurement. In J. van der Gaag & M. Perlman (Eds.), *Health, economics and health economics* (pp. 271–281). Amsterdam: North-Holland.

EFFECTIVENESS

Effectiveness is the extent to which a particular health technology (medical device, drug, procedure, health program, or health service, including interventions) does what it is intended to do (i.e., leads to a beneficial health outcome or result) when it is provided under clinical practice conditions or in the field.

The basic conceptual difference between effectiveness and efficacy of a given medical technology lies in the conditions under which it is provided and estimated. Efficacy refers to its benefits when it is deployed under ideal conditions (under the highest possible control of variables) and effectiveness when it is provided under realistic conditions, such as those encountered in clinical practice.

The effectiveness of a particular health technology can be established or estimated by means of different quasi-experimental and observational study designs. The selection of a study type to estimate the effectiveness of a given health technology depends, among other factors, on the objectives of the study, the conditions under which it has to be carried out, and the availability and the type of data that must be observed.

It is important to clearly distinguish among effectiveness, efficacy, and efficiency (the relationship between the cost or the resources used to provide a specific treatment, intervention, program, or procedure and the results obtained), since they are often confused.

In practice, efficacy and effectiveness of the same medical technology differ, and the former should not be necessarily considered an accurate estimate of the latter. For example, the efficacy of a vaccine or a drug, estimated in a randomized controlled trial, tends to be higher than its effectiveness. The degree of control of variables that is usually attained during the administration of a vaccine or a drug in a randomized trial can almost never be reached during their administration in clinics, primary health care centers, or hospitals. Reasons why the health benefits expressed by efficacy are greater than those referred by effectiveness often include the following:

- The cold chain can fail (i.e., the drug may not be stored at the proper temperature).
- Doses of the active principle of the drug contained in the tablets may vary.

- The clinical characteristics of patients may differ from those of the individuals who participated in the randomized trial, or the spectrum of the disease covered can be different.
- Patient compliance to the prescribed treatment regime is difficult to control, and the degree of adherence often changes according to different factors related to the patient's psychological profile or educational level, the mechanisms for providing health care, accessibility of health services, or, in general, the way in which health services are delivered.

Examples of endpoints (primary or secondary events observed in a patient during the course of a treatment or in a health program during its implementation) usually used to estimate effectiveness are as follows: the 1-year survival rate of patients with a particular type of neoplasm treated with chemotherapy in an outpatient service at a university hospital, the percentage of hypertensive patients with systolic and diastolic pressure adequately controlled after 1 year of treatment in a primary health care center, the 5-year death rate of women enrolled in a breast cancer early detection program, and the postoperative infection rate in surgical patients enrolled in an antibiotic prophylaxis for surgical infections program.

—Carlos Campillo

See also Clinical Trials; Economic Evaluation; Efficacy; Program Evaluation

Further Readings

Alan Dever, G. E. (1984). *Epidemiology in health services management.* Gaithersburg, MD: Aspen.

Gordis, L. (1996). *Epidemiology.* Philadelphia: W. B. Saunders.

Meinert, C. L. (1986). *Clinical trials: Design, conduct, and analysis.* New York: Oxford University Press.

Effect Modification and Interaction

The term *effect modification* has been applied to two distinct phenomena. For the first phenomenon, effect modification simply means that some chosen measure of effect varies across levels of background variables. This phenomenon is thus more precisely termed *effect-measure modification*, and in the statistics literature it is more often termed *heterogeneity* or *interaction*. Referring to the second phenomenon, effect modification means that the mechanism of effect differs with background variables, which is known in the biomedical literature as *dependent action* or (again) "interaction." The two phenomena are often confused, as reflected by the use of the same terms (effect modification and interaction) for both. In fact, they have only limited points of contact.

Effect-Measure Modification (Heterogeneity)

To make the concepts and distinctions precise, suppose we are studying the effects that changes in a variable X will have on a subsequent variable Y, in the presence of a background variable Z that precedes X and Y. For example, X might be treatment level such as dose or treatment arm, Y might be a health outcome variable such as life expectancy following treatment, and Z might be sex ($1 = female$, $0 = male$). To measure effects, write Y_x for the outcome one would have if administered treatment level x of X; for example, if $X = 1$ for active treatment, $X = 0$ for placebo, then Y_1 is the outcome a subject will have if $X = 1$ is administered, and Y_0 is the outcome a subject will have if $X = 0$ is administered. The Y_x are often called *potential outcomes*.

One measure of the effect of changing X from 0 to 1 on the outcome is the difference $Y_1 - Y_0$; for example, if Y were life expectancy, $Y_1 - Y_0$ would be the change in life expectancy. If this difference varied with sex in a systematic fashion, one could say that the difference was modified by sex, or that there was heterogeneity of the difference across sex. Another common measure of effect is the ratio Y_1/Y_0; if this ratio varied with sex in a systematic fashion, one could say that the ratio was modified by sex.

For purely algebraic reasons, two measures may be modified in very different ways by the same variable. Furthermore, if both X and Z affect Y, absence of modification of the difference implies modification of the ratio, and vice versa. As a simple example, suppose for the subjects under study $Y_1 = 20$ and $Y_0 = 10$ for all the males, but $Y_1 = 30$ and $Y_0 = 15$ for all the females. Then $Y_1 - Y_0 = 10$ for males but $Y_1 - Y_0 = 15$ for females, so there is a 5-year modification of the

difference measure by sex. But suppose we measured the effects by expectancy ratios Y_1/Y_0, instead of differences. Then $Y_1/Y_0 = 20/10 = 2$ for males and $Y_1/Y_0 = 30/15 = 2$ for females as well, so there is no modification of the ratio measure by sex.

Consider next an example in which $Y_1 = 20$ and $Y_0 = 10$ for all the males, and $Y_1 = 30$ and $Y_0 = 20$ for all the females. Then $Y_1 - Y_0 = 10$ for both males and females, so there is no modification of the difference by sex. But $Y_1/Y_0 = 20/10 = 2$ for males and $Y_1/Y_0 = 30/20 = 1.5$ for females, so there is modification of the ratio by sex.

Biologic Interaction

The preceding examples show that one should not in general equate the presence or absence of effect-measure modification to the presence or absence of interactions in the biologic (mechanistic) sense, because effect-measure modification entirely depends on what measure one chooses to examine, whereas the mechanism is the same regardless of that choice. Nonetheless, it is possible to formulate mechanisms of action that imply homogeneity (no modification) of a particular measure. For such a mechanism, the observation of heterogeneity in that measure can be taken as evidence against the mechanism (assuming, of course, that the observations are valid). It would be fallacious, however, to infer that the mechanism is correct if homogeneity was observed, for the usual reason that many other mechanisms (some unimagined) would imply the observation.

A classic example is the simple "independent action" model for the effect of X and Z on Y, in which subjects affected by changes in X are disjoint from subjects affected by changes in Z. This model implies homogeneity (*absence* of modification by Z) of the average X effect on Y when that effect is measured by the difference in Y (if Y is a disease indicator, the average of Y is the risk, and the average Y difference is the risk difference). If X and Z both have effects, this homogeneity of the difference forces ratio measures of the effect of X on Y to be heterogeneous across Z. When additional factors are present in the model (such as confounders), homogeneity of the risk differences can also lead to heterogeneity of the excess risk ratios.

Biologic models for the mechanism of X and Z interactions can lead to other patterns; for example, certain multistage models in which X and Z act at completely separate stages of a multistage mechanism can lead to homogeneity of ratios rather than differences, as well as particular dose-response patterns. Special caution is needed in interpreting observed patterns, however, because converse relations do not hold: Many different plausible biologic models will imply identical patterns in the effect measures.

Taking the independent-action model as a baseline, one may offer the following *dependent-action* definitions for an outcome indicator Y as a function of the causal antecedents X and Z. *Synergism* of $X = 1$ and $Z = 1$ in causing $Y = 1$ is defined as necessity and sufficiency of $X = 1$ and $Z = 1$ for causing $Y = 1$; that is, $Y = 1$ if and only if $X = 1$ and $Z = 1$. We also may say that $Y = 1$ in a given individual would be a synergistic response to $X = 1$ and $Z = 1$ if $Y = 0$ would have occurred instead if either $X = 0$ or $Z = 0$. In potential-outcome notation where Y_{xz} is the outcome when $X = x$ and $Z = z$, this definition says synergistic responders have $Y_{11} = 1$ and $Y_{10} = Y_{01} = Y_{00} = 0$. *Antagonism* of $X = 1$ by $Z = 1$ in causing $Y = 1$ is defined as necessity and sufficiency of $Z = 0$ for $X = 1$ to cause $Y = 1$. This definition says synergistic responders have $Y_{10} = 1$ or $Y_{01} = 1$ or both, and $Y_{11} = Y_{00} = 0$.

With these definitions, synergism and antagonism are not logically distinct concepts, but depend on the coding of X and Z. For example, switching the labels of "exposed" and "unexposed" for one factor can change apparent synergy to apparent antagonism, and vice versa. The only label-invariant property is whether the effect of X on a given person is altered by the level of Z (action of X depends on Z); if so, by definition we have biologic interaction. Absence of any synergistic or antagonistic interaction among levels of X and Z implies homogeneity (*absence* of modification by Z) of the average X effect across levels of Z when the X effect is measured by the differences in Y across levels of X. The converse is, however, false: Homogeneity of the difference measures (e.g., lack of modification of the risk difference) does not imply absence of synergy or antagonism, because such homogeneity can arise through other means (e.g., averaging out of the synergistic and antagonistic effects across the population being examined).

A more restrictive set of definitions of interactions is based on the sufficient-component cause model of causation. Here, *synergism* of the indicators X and Z is defined as the presence of $X = 1$ and $Z = 1$ in the

same sufficient cause of $Y = 1$, that is, the sufficient cause cannot act without both $X = 1$ and $Z = 1$. Similarly, *antagonism* of $X = 1$ by $Z = 1$ is defined as the presence of $X = 1$ and $Z = 0$ in the same sufficient cause of $Y = 1$. These definitions are also coding dependent.

The use of indicators in the above definitions may appear restrictive but is not. For example, to subsume a continuous outcome T such as death time, we may define Y_t as the indicator for $T \leq t$ and apply the above definitions to each Y_t. Similar devices can be applied to incorporate continuous exposure variables. The resulting set of indicators is, of course, unwieldy and in application has to be simplified by modeling constraints (e.g., proportional hazards for T).

Noncausal (Statistical) "Interaction"

Both the preceding usages of "effect modification" and "interaction" refer to causal phenomena. In the statistics literature, "interaction" is often used without explicit reference to causality. For example, in the context of regression modeling, an "interaction term" is usually nothing more than a term involving the product of two or more variables. Consider a logistic regression to predict a man's actual sexual preference A ($A = 1$ for men, 0 for women) from his self-reported preference R and the interviewer's gender G ($G = 1$ for male, 0 for female),

$$Pr(A = 1|R,G) = \text{expit}(\alpha + \beta R + \gamma I + \delta RI),$$

where $\text{expit}(x) = e^x/(1 + e^x)$ is the logistic function. Such a model can be useful in correcting for misreporting. The term δRI (or sometimes just RI or just δ) is often called an *interaction term*. It is, however, more accurately called a *product term*, for presumably neither self-report nor interviewer status has any causal effect on actual preference and, thus, cannot interact causally or modify each other's effect (since there is no effect to modify).

If $\delta \neq 0$, the product term implies that the regression of A on R depends on I: For male interviewers the regression of A on R is

$$\begin{aligned} Pr(A = 1|R,G = 1) &= \text{expit}(\alpha + \beta R + \gamma 1 + \delta R1) \\ &= \text{expit}(\alpha + \gamma + (\beta + \delta)R), \end{aligned}$$

whereas for female interviewers the regression of A on R is

$$\begin{aligned} Pr(A = 1(|R,G = 0) &= \text{expit}(\alpha + \beta R + \gamma 0 + \delta R0) \\ &= \text{expit}(\alpha + \beta R). \end{aligned}$$

Thus, we can say that the gender of the interviewer affects or modifies the logistic regression of actual preference on self-report. Nonetheless, since neither interviewer gender nor self-report has an effect on actual preference (biologically or otherwise), they have no biologic interaction.

When both the factors in the regression do causally affect the outcome, it is common to take the presence of a product term in a model as implying biologic interaction and, conversely, to take absence of a product term as implying no biologic interaction. Neither inference is correct: The size and even direction of the product term can change with choice regression model (e.g., linear vs. logistic), whereas biologic interaction is a natural phenomenon oblivious to our choice of model for analysis. Assuming no bias is present, however, a product term in a linear statistical model for a causal dependency can arise only from the presence of biologic interaction in the dependent-action sense.

—*Sander Greenland*

See also Causal Diagrams; Causation and Causal Inference

Further Readings

General

Greenland, S., Lash, T. L., & Rothman, K. J. (2008). Concepts of interaction. In K. J. Rothman, S. Greenland, & T. L. Lash (Eds.). Modern epidemiology (3rd ed., chap. 5). Philadelphia: Lippincott.

Models of Interaction and Synergism

Greenland, S. (1993). Additive-risk versus additive relative-risk models. *Epidemiology, 4*, 32–36.

Greenland, S., & Poole, C. (1988). Invariants and noninvariants in the concept of interdependent effects. *Scandinavian Journal of Work, Environment, and Health, 14*, 125–129.

Moolgavkar, S. (1986). Carcinogenesis modeling: From molecular biology to epidemiology. *Annual Review of Public Health, 7*, 151–169.

VanderWeele, T. J., & Robins, J. M. (2007). The identification of synergism in the sufficient-component cause framework. *Epidemiology, 18*(3), 329–339.

Weinberg, C. R. (1986). Applicability of simple independent-action model to epidemiologic studies

involving two factors and a dichotomous outcome. *American Journal of Epidemiology, 123*, 162–173.

Efficacy

Efficacy is the extent to which a particular health technology (medical device, drug, procedure, health program, or health service, including interventions) does what it is intended to do (i.e., a beneficial health outcome or result) under ideal conditions. For instance, clinical trials, which are conducted with a selected population and during which subjects may be monitored more closely than they would be in a clinical practice, can establish the efficacy of a health technology.

In fact, these ideal conditions are similar to those sought in conducting experimental research. Therefore, efficacy can be determined or estimated on the basis of the design, the analysis, the conduct, and the results of randomized controlled trials. The characteristics of properly designed and conducted randomized controlled trials make it possible to attain the highest possible control of variables, and to minimize biases as well as other threads to their internal validity. For these reasons, this type of study design can reproduce the closest conditions to the ideal. Of course, the results obtained under the artificial conditions of a clinical trial may not be replicated in ordinary clinical practice, which explains why the efficacy of a treatment is often higher than its efficiency.

Efficacy is calculated using defined endpoints that are primary or secondary events observed in a patient during the course of a treatment or in a health program during its implementation. Examples of endpoints used to estimate efficacy in randomized trials include 1-, 3-, or 5-year survival after the administration of chemotherapy; the number of people immunized after the administration of a vaccine to a group of individuals at risk of becoming infected; the death rate at the end of a randomized trial in patients receiving a drug for lowering hypertension (compared with the death rate of those receiving placebo or other drugs in the control arm of the trial); reduction of the mortality rate from myocardial infarction at the end of the Multiple Risk Factor Intervention Trial; the viral load and the adherence rate among nonadherent HIV-infected patient groups receiving different prescribed regimes of antiretrovirals; and the 24-hr pain relief response to sumatriptan and naproxen in a double-blind, two-arm controlled trial to determine the efficacy of sumatriptan for the acute treatment of migraine.

It is important to clearly distinguish among efficacy, effectiveness (results obtained under realistic conditions, such as in regular medical practice), and efficiency (the relationship between the cost or the resources used to provide a specific treatment, intervention, program, or procedure and the end results obtained), since they are often confused. The basic difference between efficacy and effectiveness of a given medical technology lies in the conditions under which it is provided and estimated: Efficacy refers to benefits when it is deployed under ideal conditions and effectiveness when it is provided under field conditions—that is, those encountered in clinical practice.

—Carlos Campillo

See also Clinical Trials; Economic Evaluation; Effectiveness; Program Evaluation

Further Readings

Alan Dever, G. E. (1984). *Epidemiology in health services management*. Gaithersburg, MD: Aspen.

Gordis, L. (1996). *Epidemiology*. Philadelphia: W. B. Saunders.

Meinert, C. L. (1986). *Clinical trials: Design, conduct, and analysis*. New York: Oxford University Press.

Ehrlich, Paul (1854–1915)

Paul Ehrlich was immensely important in the fields of chemistry, immunology, and histology. Perhaps his greatest contribution to public health was his discovery of Salvarsan, also known as 606, the first organic antisyphilitic. Martha Marquardt, his secretary for 13 years, describes his life as "one long fight for the promotion of medical science in the service of mankind" (Marquardt, 1951, p. vi).

Born on March 14, 1854, in Strehlen, Germany, to prosperous innkeepers Ismar and Rosa, Ehrlich's fascination with science was fostered from an early age by his grandfather, a natural scientist, and his cousin, Karl Weigert, a bacteriologist. Ehrlich excelled in

Latin and the sciences throughout school, earning a Doctor of Medicine degree at Leipzig in 1878.

A fervent researcher, nearly all Ehrlich's waking hours were consumed with planning, executing, or contemplating his latest experiment. Though he spent time—at least early in his career—treating patients, it was always the histology and inquiry into the staining of tissues with aniline dyes (to differentiate cell types) that intrigued him most. His methods led to the development of the Gram stain many years later.

After Robert Koch's discovery of the tuberculosis bacillus in 1882, Ehrlich set out to stain the organism and succeeded quickly. Thus began a partnership between Ehrlich and Koch, as well as the capability for rapid diagnosis of tuberculosis, even in the absence of symptoms. Ehrlich contracted the disease during his research and spent 2 years recovering in Egypt. In 1889, Ehrlich returned to science and began working in Koch's newly established Institute for Infectious Disease in Berlin. It was there in 1892 that colleague Emil von Behring discovered the diphtheria antitoxin. It was Ehrlich, however, who optimized the practical use of the serum when he quantified dosage in terms of a standardized measurement obtained through animal studies. This procedure became standard protocol in vaccine development.

Ehrlich's work led him to the formulation of the side-chain theory, his explanation of antibody production. Ehrlich postulated the existence of cellular receptors that attract foreign chemicals and bind them specifically. Once bound, if the chemical did not kill the cell, Ehrlich believed the cell would produce many more receptors, some of which would separate and float through the bloodstream to become antibodies. Ehrlich also worked to find compounds he termed *magic bullets*, which could target pathogenic organisms and destroy them without harming the surrounding host tissues. Ehrlich recognized antibodies as a natural magic bullet and sought to synthesize similar agents to kill trypanosomes and spirochetes. He called this field "chemotherapy" to distinguish it from the prevailing drug research of the day in which investigators sought to remedy symptoms rather than destroy the causative agent. In 1908, Ehrlich was awarded the Nobel Prize in Medicine for his work in immunity.

It was at the Institute for Experimental Therapy in Frankfurt, of which Ehrlich was director from 1899 to 1915, where he made his most famous discovery, Salvarsan. Ehrlich worked to create derivatives of Atoxyl (arsanilic acid), an effective but toxic drug. He produced and tested hundreds of compounds, including 606, which was initially deemed ineffective by an assistant. Two years later, a Japanese apprentice, Sahachiro Hata, used 606 to treat *syphilis-infected rabbits* and found it to be highly efficacious. After repeated tests in a variety of animal models, Ehrlich and Hata announced the finding to the world, and soon after, they were inundated with requests from physicians to begin clinical testing. The drug proved much safer and more successful than mercury or other common treatments available at that time and was given the name Salvarsan. Though supplanted by penicillin in the 1940s, Ehrlich's discovery greatly reduced the prevalence and negative sequelae of syphilis and yaws worldwide and paved the way for chemotherapeutic investigation as we know it today.

In 1914 as World War I began, Ehrlich's health began to falter and he suffered a minor stroke in December. He died on August 20, 1915, after a second stroke.

—Erin L. DeFries

See also Koch, Robert

Further Readings

Nobel lectures, physiology or medicine 1901–1921. (1967). Amsterdam: Elsevier. Retrieved August 6, 2006, from http://nobelprize.org/nobel_prizes/medicine/laureates/1908/ehrlich-bio.html.

Marquardt, M. (1951). *Paul Ehrlich.* New York: Henry Schuman.

Emerging Infections

Although emerging infections is a relatively new field of study, many of the diseases it researches have beset mankind for centuries and, along with new emerging infections, are the direct causes of more than 15 million deaths worldwide each year. Millions more die as a result of prior infections such as streptococcal rheumatic heart disease or because of the complications associated with chronic infections.

Among the "plagues" or "pestilences" familiar to students of history are the Black Death, which is believed to have killed up to half the population of medieval Europe, and smallpox and measles, which

are known to have decimated indigenous populations of North and South America when imported by European conquerors, causing millions of deaths and destroying entire civilizations. More recently, the Spanish influenza of 1918 to 1919 was responsible for more than 50 million deaths worldwide. Indeed, a soldier fighting in the trenches of World War I was more likely to be killed by influenza than by a bullet. Emerging infections that are presently at the forefront of medical and epidemiological research include severe acute respiratory syndrome (SARS), HIV/AIDS, West Nile virus, and avian (bird) influenza.

Definitions

Emerging infections can be defined as infections that have recently appeared within a population or that may have existed before but are rapidly increasing their geographic range and prevalence. They may be further classified as newly emerging infections, reemerging infections, or deliberately emerging infections.

Newly Emerging Infections

Newly emerging infections are diseases that have not previously appeared in humans. One such disease is HIV/AIDS, which is believed to have made the leap from animals to humans between 60 and 70 years ago, possibly as a result of contact with infected chimpanzees. To date, human immunodeficiency virus (HIV), the virus that causes AIDS, has infected more than 60 million people worldwide, most living in developing countries, and 70% living in sub-Saharan Africa. With as much as 25% to 30% of the adult population infected, life expectancies have fallen dramatically in many of these countries. Since the appearance of the first recognized case in December 1981, more than 25 million have died of the disease, causing socioeconomic devastation in many populations. While antiretroviral drugs have been developed, there is no cure for the disease, and most of its victims have no access to the treatment.

Reemerging Infections

Reemerging infections are ones that have existed in the past but are undergoing a rapid resurgence in incidence or geographical and host range. Among the most deadly is the tuberculosis (TB). Recognized among humans as far back as 4000 BCE, it was an incurable disease that, in the 19th and early 20th centuries, struck especially hard at the urban poor population. It was not until 1906 that an immunizing agent was developed, and only in 1921 was it used successfully in humans. The development of the antibiotic streptomycin in 1946 provided a treatment for those already suffering from the disease, replacing sanatoria and often draconian surgical interventions.

With the availability of vaccine and treatment, many hoped that tuberculosis could finally be eradicated. However, in the 1980s, drug-resistant strains began to appear, aided in part by failure of patients to complete the full course of drugs. The resulting reemergence of the disease has led to the infection of as many as one third of the world's population. Factors such as the prevalence of HIV/AIDS make entire populations more susceptible to infection with TB. Of special concern is the emergence of multiple drug-resistant strains of TB.

Deliberately Emerging Infections

These infections are caused by microbes that have been developed or adapted by man, generally for belligerent uses. Their intentional spread is considered to be bioterrorism. And while terrorists have historically relied primarily on more accessible and conventional weapons such as guns and explosives to further their objectives, biological weapons are becoming increasingly available. They allow terrorists to move from low-casualty, high-visibility attacks to mass-casualty attacks for which biological weapons are especially suited. Bioterrorist agents may include naturally occurring microbes, or bioengineered organisms deliberately designed to cause the greatest possible harm. Only two bioterrorist events have occurred in the United States. The first was on September 9, 1984, when a religious cult, the Rajneeshees, sprayed *Salmonella typhimurium* on salad bars in The Dalles, Oregon, causing 751 cases of food poisoning but no deaths. The second attack occurred in September 2001, when *Bacillus anthracis* spores were distributed through the U.S. Postal Service, leading to 22 cases of anthrax infection and 5 deaths. Whereas they used a chemical agent in their 1995 subway attack in Tokyo, the religious extremist group Aum Shinrikyō had earlier attempted to use anthrax and botulinum toxin.

Evidence of Al Qaeda's interest in bioterrorism raises the specter of uncontrollable global pandemic.

Moreover, because the bioengineering of microbes makes it possible to significantly increase the deadliness of a disease and the efficiency of its delivery, even diseases that previously occurred naturally, such as smallpox, could become untreatable.

Emergence and Transmission

Many factors influence patterns of emergence and transmission in emerging infections. Most pathogens have been present in the environment without causing serious illness until an opportunity arose for them to infect new populations. Thus, a pathogen in animals (a zoonosis) may become transmittable to humans. Such pathogens include West Nile virus and the avian H5N1 influenza virus. Most mutations in virus strains are minor, resulting in "antigenic drift" over an extended time period. Occasionally, however, strains will mutate significantly, resulting in "antigenic shift" that creates a disease to which humans are susceptible.

The likelihood of transmission to a human host is increased by close proximity between infected animals and humans. Thus, recent human cases of avian influenza are believed to have been a result of human contact with chicken feces and secretions among people who live with or handle affected poultry.

Additional factors in the emergence and transmission of infectious diseases include many in which humans are implicated. Sudden changes in the environment caused by agricultural development, deforestation, and overgrazing can tip the balance in favor of diseases such as Rift Valley fever and schistosomiasis, which are caused by damming rivers, or pulmonary hantavirus syndrome, which struck in the southwestern United States in 1993 as a result of drought and human activities.

Human demographic factors are another source of infection emergence and transmission. The dramatic rise of global travel and commerce was instrumental in introducing the SARS virus to Canada in 2003 when people who had been to China and become infected flew into airports carrying the disease. Smuggling of infected birds and poultry between countries has spread avian influenza among birds, while their natural migrations have led to additional spread of the disease. Population upheavals caused by mass migration or war are further contributors to disease emergence and spread.

Behavior that encourages the emergence and spread of infectious diseases among humans includes intravenous drug use, behaviors that cause the spread of sexually transmitted disease, and drug resistance frequently caused by noncompliant patients. This drug resistance can, as in the case of tuberculosis, turn a formerly curable disease into a pandemic. Such resistance can also lead to reemerging infection when the transmitting hosts develop resistance to the treatment or prevention, as in the case of malaria, which had been held in check by the extensive spraying of pesticides to kill mosquitoes that transmit the disease.

Other factors, including high-density urban living, food mass production, and failures of public health measures and infrastructures, play an important part in the emergence and transmission of infectious disease. Frequently, a number of these factors combine to create a situation like the resurgence of cholera in South America and Africa, possibly as a result of water treatment failure, drought, and civil tensions. And, as we noted above, intentional release of disease into populations, such as in the anthrax cases of 2001 in the United States, are causing increasing concern.

Responses to Emerging Infections

Any treatment of emerging infections must depend on comprehensive surveillance, rapid diagnosis of disease, and implementation of containment strategies. Concern over emerging infections such as avian influenza and hepatitis C has led to concerted efforts to develop appropriate response systems. The continuing efforts at developing effective retroviral therapy for emerging infections have resulted in multiple treatments for HIV/AIDS patients, as well as a number of drugs that may be useful in response to pandemic influenza.

Vaccines are the most effective method of preventing disease when they are available. However, millions who have no access to preventive treatment still die of vaccine-preventable diseases every year. Because emerging infections are, by their very nature, new and/or constantly evolving and developing resistance to treatment, ongoing research and development of new vaccines are crucial to the health of the public. Vaccine development and production technologies must keep pace with emerging infections and the needs of the populations that suffer from them.

—Rachel D. Schwartz and R. Gregory Evans

See also Avian Flu; Bioterrorism; Epidemic; Severe Acute Respiratory Syndrome (SARS); Vaccination; Zoonotic Disease

Further Readings

Fauci, A. S. (2001). Infectious diseases: Considerations for the 21st century. *Clinical Infectious Diseases, 32*, 675–685.

Lederberg, J., Shope, R. E., & Oaks, S. C., Jr. (Eds.). (1992). *Emerging infections: Microbial threats to health in the United States.* Washington, DC: National Academy Press.

Morse, S. S. (1995). Factors in the emergence of infectious diseases. *Emerging Infectious Diseases, 1*, 7–15.

Subbaro, K., & Katz, J. (2000). Avian influenza viruses infecting humans. *Cellular and Molecular Life Sciences, 571*, 1770–1784.

Environmental and Occupational Epidemiology

Environmental and occupational epidemiology are subdisciplines of epidemiology and use the same standard epidemiologic research approaches, study designs, and analytical methods in estimating disease occurrence, relative risk, and statistical variability. These two disciplines focus on studying the effects on human health attributable to chemical, particulate, metallic, physical, infectious, and psychosocial agents in the workplace and general environment. Occupational and environmental epidemiology involves a wide variety of methodological techniques that are used to evaluate associations between workplace or community exposures and health outcomes. This entry reviews the study designs used in occupational and environmental research and also examines sources of bias in such studies.

Occupational and environmental epidemiology are closely linked by the nature of the many common exposures of concern. Epidemiologic studies in these fields often consider other factors such as genetics, nutrition, and behavioral patterns, although in the context of their confounding influence on the workplace or environmental-exposure/outcome relationship rather than as primary exposure factors. For example, in a study of radiation exposure and risk of lung cancer, smoking in this context would be considered as a confounder (not a primary risk factor), whereas in a study of health risks due to secondhand smoke, smoking by family, friends, or coworkers would be considered as an environmental or occupational exposure. Several important causal relationships have been identified through occupational and environmental research, such as benzene exposure and acute myeloid leukemia, vinyl chloride and angiosarcoma of the liver, and ultraviolet radiation and melanoma.

Diseases of concern in occupational and environmental epidemiology include virtually the entire spectrum of health events, including cancer and cardiovascular, neurological, respiratory, immunological, and dermal diseases, as well as injuries, reproductive health, and mental health. Infectious disease and nutritional disorders are typically the domain of other epidemiologic subdisciplines, but they are also studied in occupational or environmental epidemiologic research (e.g., the evaluation of health care of food-processing workers, or the evaluation of health impacts of environmental disasters [weather, earthquakes] or sociopolitical conditions [war, political unrest]). The well-known John Snow natural experiment study is a classic example of an environmental epidemiologic study, in which mortality rates of cholera were found to be attributable to different sources of residential water.

The occupational setting offers several advantages for conducting epidemiologic research. Often occupational epidemiologic study results can provide a ''sentinel'' for the potential health effects for a particular agent because exposures are often higher in the workplace and frequently are better characterized there than in the general environment. Also, because the population at risk (a workforce) can be more accurately enumerated, occupational studies may be less vulnerable to potential biases. However, because these studies examine adult working populations, certain limitations exist. Typically, only adult health issues are assessed, and since working populations generally present with a better health profile than the general public, the likelihood for detecting certain health risks may be lower than if the study were conducted in a general population. Environmental epidemiologic studies often examine involuntary exposures and can be useful in providing guidance for public health policy. Epidemiologic methods applied to occupational and environmental settings, particularly in the area of exposure assessment, will continually evolve to keep pace with rapidly changing and more complex environmental exposures.

Study Design and Analysis Issues

Most epidemiologic studies are observational; that is, researchers observe events over time or events that have already happened. Characterizing the distribution of disease is the primary focus of descriptive epidemiology, while analytic epidemiology examines the determinants—that is, potential causes and risk factors of disease. Descriptive studies provide information on patterns of disease by factors such as time, age, and gender and may provide hypotheses for testing possible explanations for the patterns observed. The two primary objectives of analytic epidemiology are to (1) identify associations of disease with possible etiological factors (e.g., occupational, environmental, or lifestyle) and (2) test hypotheses regarding etiology (causation). Analytic studies include a comparison population, thereby enabling the researcher to examine causal relationships between exposure and disease. The two most commonly used types of analytic studies used in occupational epidemiology are cohort studies and case-control studies. These designs are also used in environmental epidemiology, although because of resource limitations and other feasibility considerations, other study designs are often used in environmental epidemiology as well, including cross-sectional and ecological studies. The proportionate mortality ratio or *PMR* study design is a preliminary exploratory design that, historically, was often used in occupational epidemiology but is less relied on in current occupational research.

Cohort Studies

In cohort studies, the investigator selects exposed and nonexposed individuals and follows both groups to compare disease incidence or disease-specific mortality. In a prospective cohort study, the study population is defined contemporaneously and then followed into the future to determine disease occurrence. In a retrospective cohort study, the study population is defined to start at some point in the past and the disease occurrence is determined to the present time. The relative risk is based on a comparison of groups selected within the worker population (known as internal comparison) or by comparison of disease rates from an external population, such as the general population of a country, state, or county. The rates of disease mortality or incidence in the general population or in the internal comparison population are used to calculate a standardized mortality ratio (*SMR*) or standardized incidence ratio (*SIR*). The *SMR* or *SIR* calculated using an external population is a common measure in occupational cohort studies, and it is defined as the number of observed deaths or cases that occur in an exposed population divided by the number of deaths or cases that are *expected* to occur based on the mortality or incidence rates in the general population.

$$SMR = \frac{\text{Observed deaths in exposed group}}{\text{Expected deaths}}.$$

For the *SIR*, observed cases and expected cases are substituted for observed deaths and expected deaths. In calculating an *SMR* or *SIR*, the denominator reflects what the expected number of deaths or cases would be if the group under study had the same age, sex, and race-specific rates as the general population. Because calculations of the *SMR* or *SIR* typically rely on the distribution of age, sex, and race for the specific cohort under study, comparisons of *SMR*s or *SIR*s across different studies and cohorts should be evaluated cautiously given that the distribution of these weighting factors can vary across cohorts, resulting in an alteration of the standardized calculation.

Proportionate Mortality Ratio (PMR) or Proportionate Incidence Ratio (PIR) Studies

Another study design used in occupational epidemiology is the *PMR* study. The PMR is the ratio of the proportion of deaths due to a specific cause in an exposed group to the proportion of deaths due to the same specific cause in an unexposed group (usually state or national mortality data). The advantages of *PMR* studies relate to the relatively low cost, easy access to data, and quick availability of study results. Methodological limitations include the following: (1) exposure information that is restricted to occupation and industry as recorded on the death certificate, (2) the potential for misleading conclusions if populations with different distributions of causes of death unrelated to the exposure under study are compared, and (3) potential biases resulting from the healthy worker effect (see below). Since the direction of bias resulting from these limitations is not usually known, *PMR* studies are difficult to interpret and considered exploratory studies that are generally not relied on for causal inference.

Proportional incidence studies, which are based on disease incidence registries rather than mortality records, have similar characteristics as *PMR* studies.

A less biased approach that is recommended when only numerator data (cases only) are available is a case-control study where controls are selected from other cases of disease not thought to be related to the exposure under study.

Case-Control Studies

Another commonly used study design in occupational and environmental epidemiology is the case-control study. In this design, study participants are divided into two groups based on disease status: Cases are persons who have the disease and controls are persons who do not. Data regarding past exposures in both groups are then ascertained by a variety of methods, including interviews, questionnaires, medical records, work histories, workplace air or biomonitoring levels, industrial hygiene measurements, or job titles. The distributions of exposures between cases and controls are compared. Exposure status should not be known at the time of defining cases and controls.

Nested Case-Control Studies

In occupational epidemiology, a nested case-control study is an efficient design often used within a well-defined source cohort, generally consisting of workers employed within the same industry or occupational setting. This study base is defined a priori and is fully enumerated before or as part of the nested case-control study. Similar to population-based case-control studies, there are four primary control-sampling methods: incidence density, cumulative incidence, cumulative survival, and case base. In occupational studies, incidence density sampling may be the most appropriate control selection procedure, in which controls are selected for each case among workers who were disease free or did not have health outcome of interest at the time of the identification of the case. Control sampling may be stratified or matched according to selected factors, such as age, gender, race, or occupational group. A nested case-control study is then analyzed using standard techniques applied to the analyses of case-control studies. This design is generally more cost-effective and timely than a cohort study and may allow for more extensive and accurate exposure assessment because the ascertainment of exposure and other information may be more feasible for a smaller number of workers than what would be required for an entire cohort study. On the other hand, there may be greater potential for selection bias in nested case-control studies than cohort studies since controls are sampled from the study base. However, there is less potential for this bias than standard population-based case-control studies because eligible controls are sampled from a well-defined cohort.

Cross-Sectional Studies

A frequently used design in environmental and occupational epidemiology is the cross-sectional study. In this design, disease or other health-related characteristics and exposure are determined simultaneously at a given point in time. Generally, the results of such studies have more limitations in assessing causation than cohort or case-control studies for several reasons. Because cross-sectional studies usually do not involve a time component, the dynamic interaction between exposure and disease can be difficult to determine. In addition, cross-sectional studies are often based on prevalent rather than incident cases although incidence measures can be generated if data are collected that determine the date of disease onset. Most often, this design relies on interviews or questionnaire surveys, which can be a source of both strengths and limitations depending on the implementation of the survey. One advantage of the survey approach is the ability to collect information on exposure, disease, and confounders directly from participants. However, self-reported information can be subject to interviewer, reporting, and recall biases. Validity of information drawn from these types of surveys also can be limited due to low participation rates. In occupational research, cross-sectional designs may be useful for evaluating exposure-outcome relationships of short duration, for example, surveys of respiratory symptoms, musculoskeletal disorders, or other symptom surveys.

Ecologic Studies

In contrast to previously discussed study designs, ecologic or community-level studies use populations as units of analysis rather than individuals. These studies may, for example, compare exposure and disease profiles between geographical units such as countries or may evaluate changes in morbidity and mortality in relation to changes in potential exposure at a population level.

Ecologic studies are frequently conducted in environmental epidemiology mainly because of their low costs and reliance on more readily available and accessible data. In some cases, ecologic studies offer advantages because individual-level studies may not be practical because of exposure assessment considerations or study design issues. Also, the research question may be one that is targeted at the ecologic level, for example, what is the impact of a policy that restricts cell phone use while driving on the rate of motor vehicle crashes?

Ecologic designs have been characterized into three basic types: (1) multiple group design (mapping studies), (2) time-trend design, and (3) the combination of location and time (mixed design). Mapping studies essentially involve comparisons of disease rates across different geographic regions (e.g., variation of female breast cancer across different countries or different regions within countries). In an exploratory sense, this evaluation would not necessarily have a particular exposure under consideration but may generate hypotheses after observing certain trends. More specific hypotheses can be developed by characterizing regions by average levels of environmental risk factors (e.g., industrial emissions, air pollution, pesticide use) or nonenvironmental factors (e.g., diet, reproductive history, alcohol consumption). Time-trend analyses examine changes in disease occurrence across calendar time, and mixed designs combine both geographic and time-trend analyses.

Because of their use of aggregate-level data, ecologic studies are subject to the "ecological fallacy," the inappropriate application or interpretation of an aggregate relationship to the level of the individual. The effect estimates of an ecologic study may not reflect the biologic effect at the individual level. These studies can also be affected by in- or outmigration of the study population, which will limit the interpretation of associations from studies that define exposure in terms of time and place. For example, in a hypothetical study of the association between drinking arsenic-contaminated well water and risk of bladder cancer, the results may be biased toward the null if a large proportion of the population moved into the arsenic-endemic area from an area without arsenic-contaminated drinking water shortly before the population was analyzed. In contrast, findings may be biased upward if a population migrated from an arsenic-endemic area to a study area in which arsenic levels in community drinking water were considered low. Because of these and other limitations, results of ecologic studies are used mainly for the generation of hypotheses and are given much less weight in evaluating cause-and-effect relationships. Some of these biases can be reduced by (1) examining smaller geographic areas where exposures may be more homogeneous and residential mobility/stability can be better characterized and (2) by stratifying data into subgroups that have more homogeneous disease risks.

Ecologic studies, however, can be useful for assessing the overall health of a community and in evaluating overall impacts of community-wide intervention programs (e.g., gun control, seat belt laws, minimum-age drinking laws). Community intervention studies represent a type of ecologic study because an intervention is made at the community level and subsequent disease or outcomes of interest are ascertained through time, before and after the community intervention. A classic example is public water system fluoridation (intervention) and the evaluation of adverse dental manifestations such as cavities. Water supply system fluoridations were initiated in the 1940s and occurrences of dental conditions before and after fluoridation were compared. The rates of adverse dental manifestations decreased after fluoridation intervention at the community level. Furthermore, dental conditions remained elevated in areas that did not receive this intervention.

Cluster Studies

Investigations of cancer or disease clusters are commonly requested of public health agencies by citizens concerned about a potential environmental hazard. A cluster is generally defined as an aggregation of relatively uncommon events or diseases in space and/or time in rates or levels that are greater than could be expected by chance. Potential disease clusters are often perceived to exist on the basis of anecdotal evidence, and epidemiologists and biostatisticians may spend considerable efforts in evaluating whether a true cluster exists. The most commonly used analytical approach is to statistically compare, in the time/location of interest, the observed number of cases in the perceived cluster to what would be expected among the population that generated the cluster if cases occurred at the same rate in a relevant comparison geographical area such as a state, province, or country.

A survey of state health departments in 1997 found that the national total for public requests for cancer

clusters investigations was approximately 1,100. "True" clusters that are tied to a specific agent can be common for infectious diseases but are extremely rare for cancer or other diseases. Thus, studies of environmental exposures and cancer clusters may identify research hypotheses, but rarely have they produced useful information about disease causation because of several common limitations:

- Ascertainment bias is a concern because the media or the local community may conduct special searches for cases, or diagnostic criteria are vague, thus resulting in cases that may be too heterogeneous for a valid study.
- The data are subject to gerrymandering—that is, circumscribing the study area around the identified cluster of cases.
- Insufficient disease induction period—that is, the time between exposure to environmental agent and the manifestation or diagnosis of the disease is too short.
- A lack of statistical power due to small sample size characterizes many perceived clusters.

Furthermore, the exposures under investigation are often poorly characterized, heterogeneous, or even too low in concentration to increase the likelihood of disease.

Bias in Occupational and Environmental Epidemiology Studies

Any systematic error in the design of a study, its implementation, or analysis of the data is called bias. Assessment of bias is critical in conducting and evaluating occupational and environmental epidemiology studies because the presence or lack of bias provides a measure of the study validity and helps determine how study results can be interpreted. Limitations in exposure assessment are a common problem for all epidemiologic studies but are especially relevant to occupational and environmental health research. Several types of bias relevant to all types of epidemiologic research, particularly occupational and environmental epidemiology, are given below.

Selection bias refers to systematic differences in the characteristics between those who are selected for study and those who are not. A special case of this threat to the validity of a study is surveillance bias, which refers to differential disease ascertainment (e.g., through medical monitoring) among those who are exposed compared with those who are not exposed. Selection bias may also invalidate conclusions from surveys that would include only volunteers or persons known to have certain medical conditions.

Misclassification is a type of bias that is created by inaccurate assignment of study subjects to "exposed" versus "unexposed" or "case" versus "noncase" status. Misclassification may occur in two forms. In differential misclassification, the rate of misclassification of exposure is different between cases and controls or the rate of misclassification of disease status is different between the exposed and unexposed groups. In nondifferential misclassification, the rate of misclassification of exposure may be the same for both cases and controls or the rate of misclassification of disease may be the same for both exposed and unexposed individuals.

Information bias refers to errors in collecting data that result in differential accuracy of information about comparison groups. Specific examples of information bias include recall bias, interviewer bias, and nonresponse bias. For example, when asking workers about the frequency of lifting events and lifting requirements (e.g., estimated weights lifted) at work, those who have low-back pain may be more likely to recall lifting events than workers in the same occupation without low-back pain. This is an example of recall bias that may lead to differential misclassification of exposure.

Confounding bias is the error introduced into the evaluation of an exposure-disease relationship by a third factor that is an independent cause of the disease and unequally distributed among the exposed and the unexposed subjects. A confounder is a factor that distorts the effect of the risk factor (exposure) under study. For example, an association observed between work as an auto mechanic and lung cancer might not be due to exposures at work but due to confounding by smoking, which is a cause of lung cancer and associated with this group of workers. Thus, smoking confounds the relationship between occupation (auto mechanics) and lung cancer and is referred to as a "confounder" or a "confounding factor."

The *healthy worker effect* (HWE) is a well-recognized potential bias in occupational studies. The HWE is actually the result of several distinct components: (1) selection of healthier people into the workforce or active employment ("healthy hire effect"), (2) selection of unhealthy workers out of the workforce ("healthy survivor effect"), and

(3) the decline in health with time since hire. The impact of HWE varies by disease outcome with larger impacts expected for nonfatal outcomes such as musculoskeletal conditions, asthma, or other symptoms and noncancer endpoints such as heart disease, diabetes, nonmalignant respiratory disease, and nonfatal injuries. The HWE impact on cancer is assumed to be much smaller. Cross-sectional, *PMR*, and cohort studies that use external comparison groups are more affected than other study designs (e.g., case-control and cohort studies using internal comparison groups). Various analytical procedures have been suggested to reduce the bias due to HWE: (1) use of an internal comparison group, (2) analytical control of employment status (active/inactive), (3) restriction of analysis to groups with longer follow-up, and (4) analytical control for time since hire or follow-up duration. However, bias may still result from more subtle and complex scenarios that require other approaches. For *PMR* studies when cancer is the disease endpoint, use of only cancer data may reduce potential bias introduced by HWE.

—Michael A. Kelsh and Dominik D. Alexander

See also Bias; Cox Model; Disaster Epidemiology; Ecological Fallacy; Exposure Assessment; Healthy Worker Effect; Pollution; Study Design

Further Readings

Arrighi, H. M., & Hertz-Picciotto, I. (1994). The evolving concept of the healthy worker survivor effect. *Epidemiology, 5*, 189–196.

Checkoway, H., Pearce, N. E., & Kriebel, D. (2004). *Research methods in occupational epidemiology* (2nd ed.). New York: Oxford University Press.

Hertz-Picciotto, I. (1998). Environmental epidemiology. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott-Raven.

Morgenstern, H. (1998). Ecologic studies. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott-Raven.

Morgenstern, H., & Thomas, D. (1993). Principles of study design in environmental epidemiology. *Environmental Health Perspectives, 101*(Suppl. 4), 23–38.

Neutra, R. R. (1990). Counterpoint from a cluster buster. *American Journal of Epidemiology, 132*(1), 1–8.

Rothman, K. J. (1990). A sobering start for the cluster busters' conference. *American Journal of Epidemiology, 132*(1, Suppl.), S6–S13.

Trumbo, C. W. (2000). Public requests for cancer cluster investigations: A survey of state health departments. *American Journal of Public Health, 90*(8), 1300–1302.

Epidemic

An epidemic is a marked increase in the number of cases of a disease relative to the expected number of cases. Epidemic disease is sometimes contrasted with endemic disease, which is the expected or usual incidence of disease in a location. While the term *endemic* is typically confined to infectious diseases, the term *epidemic* is more widely used. Endemic can refer to either the usually observed rate of disease or simply the fact that a disease is present in a locale. For example, hantavirus is endemic to many parts of the United States. A rate of disease that is endemic on one country would constitute an epidemic if it occurred in a country where the disease is ordinarily less common.

The terms *outbreak* and *epidemic* are both used to describe sudden increases in disease occurrence, although outbreak is usually reserved for a localized occurrence of a disease that is typically not present in the population (such as an outbreak illness due to *Escherichia coli*), while epidemic is reserved for more widespread conditions. Both outbreaks and epidemics may be described as being *common source* or *propagated.* In a common source outbreak, there is a single source of disease to which the population is exposed, such as cryptosporidium in a city water supply or the interstate distribution of contaminated packaged spinach. In a propagated epidemic, the pathogen is spread from person to person, such as HIV. Sometimes, an epidemic may have both types of transmission: initial point source exposure that is then passed from person to person.

A *pandemic* is an epidemic that has spread beyond national boundaries. Most often, the term *pandemic* is used to describe influenza, and the most notable pandemic in recent history was the Influenza Pandemic of 1918. However, other diseases such as cholera, bubonic plague, and HIV have similarly demonstrated rapid spread across continents at specific historical periods.

The progress of an epidemic or outbreak is often depicted with an epidemic curve, a graph where time is on the *x*-axis and the number of cases is on the *y*-axis. Figure 1 displays the epidemic curves of

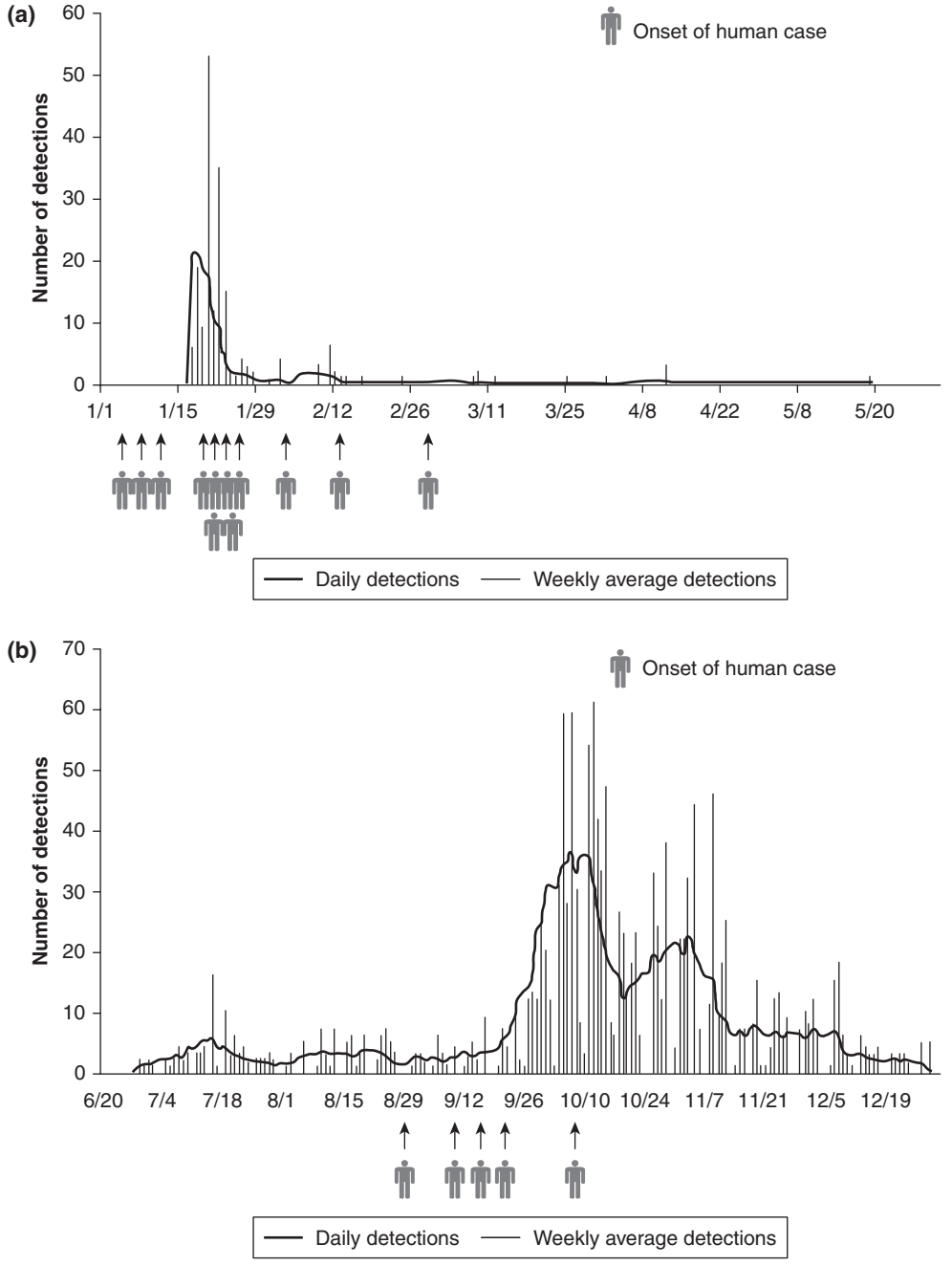


Figure 1 Epidemic Curve of the Confirmed Highly Pathogenic Avian Influenza H5N1 Outbreaks in Poultry in Thailand by Date of Notification

Source: Tiensin et al. (2005).

H5N1 avian (bird) flu in Thailand for two time periods in 2004.

Epidemics are most often caused by infectious agents such as viruses or bacteria, but other causes are possible, including chemical exposure and physical conditions such as extreme heat or cold. Due to improved sanitation, disease surveillance, and medical care, widespread epidemics of infectious diseases are

uncommon in the industrialized world today but still occur regularly in the developing world. Due in part to this shift in focus from infectious to chronic disease, the term *epidemic* is often used today to refer to health behaviors or chronic conditions of widespread concern, such as "the obesity epidemic" or "an epidemic of teenage smoking," or even "an epidemic of child abuse."

—*Sarah Boslaugh*

See also Foodborne Diseases; Infectious Disease Epidemiology; Measles; Outbreak Investigation; Snow, John; Waterborne Diseases

Further Readings

Markel, H. (2004). *When germs travel: Six major epidemics that have invaded America and the fears they have unleashed*. New York: Pantheon Books.

Nelson, K. E., & Williams, C. M. (2007). *Infectious disease epidemiology: Theory and practice* (2nd ed.). Boston: Jones & Bartlett.

Tiensin, T., Chaitaweesub, P., Songserm, T., Chaisingh, A., Hoonsuwan, W., Buranathai, C., et al. (2005). Highly pathogenic avian influenza H5N1, Thailand, 2004. *Emerging Infectious Diseases* [Serial on the Internet]. Retrieved February 19, 2007, from http://www.cdc.gov/ncidod/EID/vol11no11/05-0608.htm.

Epidemiology, History of

Student: What is epidemiology, father?

Teacher: It is all things to all men, my son . . . The plethora of definitions is the very heart of the problem . . . a structure sturdy enough to . . . shelter physicians, dentists, veterinarians, and nurses; very small (micro) biologists and fat chemists; mammalogists, bugmen, birdmen, and spacemen; traffic directors and city planners; engineers mechanical, sanitary, electrical, stationary, and human; sociologists, psychologists, and anthropologists, cultural and otherwise . . . everything!

—Anonymous (1963)

The preceding satirical barb from 1963, titled "Epidemi(olog)^{2}y*," goes on to define epidemiology ("or epidemiologology") in circular fashion as "simply the study of epidemiology," and hints at why a comprehensive history of the subject has neither been written nor, apparently, even attempted. From the vantage point of 2007, looking back over the past few decades, it appears that epidemiology has further expanded in breadth from the diffusion satirized in 1963. Moreover, it has become increasingly absorbed in methods, mathematics, and models; some critics have charged that it has reached its limitations and is spending too much time nibbling at the edges of public health rather than solving the big problems. It is thus hard to escape confusion about epidemiology's identity and origins. Like the different branches of a tree bending down to examine shared roots, the question can be seen from different vantage points leading to different conclusions about what epidemiology is, what its core techniques are, what constitutes epidemiologic practice, who is and who is not a practitioner of epidemiology, and so on. Moreover, borrowing Thomas Huxley's description of science, it seems that fundamentally epidemiology is "nothing but trained and organized common sense" (1854). Tracing the history of commonsensical solutions to health problems in populations is a subject so broad and so all-encompassing that it becomes almost impossible to grasp *in toto*.

This entry is thus focused narrowly on tracing the roots of core and traditional epidemiologic activities such as those canonized by epidemiology textbooks—notably, disease surveillance, outbreak investigation, and discovery of disease etiology and modes of transmission/acquisition. In this effort, we are led disproportionately to the infectious diseases, which have been the principal shapers of epidemiologic endeavor almost up to the present time, placing less emphasis on, for example, chronic and behavioral conditions, and their relatively newer and still-evolving methodologies.

The term *epidemiology* came into being, or at least came to have its modern meaning only fairly recently, at about the same time medical science began to appreciate that communicable diseases might actually be infectious and that such infections not only were *not* distributed randomly in populations but that their specific patterns of "nonrandomness" constituted powerful etiologic clues. These important realizations seem to have been formed around the time of the second cholera pandemic in Europe (1831–1832). Conveniently for historical purposes, the term *epidemiologie* or *épidémiologie* began to be used in the modern sense in Germany, France, the Netherlands, and probably other countries as well, at around the same time, apparently coming to the English-speaking

world a decade later. The eventful year of 1832 also saw the first proposal to establish an academic chair in epidemiology, reflecting a greatly heightened awareness of the importance of systematic and quantitative study of epidemics. In examining the history of epidemiology in both before and after its "birth" in a recognizable form, it is necessary to avoid the sin of "presentism" (imputing to past observers modern concepts and terms of which they were not aware). Therefore, the terms *proto-epidemiology* and *epidemiology* have been used here to refer to epidemiologic efforts and their antecedents occurring, respectively, before and after the period in which recognizably modern epidemiologic concepts and terms came into common use.

Before the 1830s, there seems to have been no term that corresponded to "epidemiology" in the modern sense of the word. "Loimologia," "loimology," and "loimographia," used in the 17th century, referred to the centuries-old chronicling of *loimos* (λοιμοζ), an ancient Greek term that roughly corresponds to "plagues," "pestilences," and major epidemic/pandemic diseases. "Epidemical" and "epidemial," popular in the late 1700s and early 1800s, were adjectives applied to infectious disease epidemics. The Latin term *epidemiologia* can be found in at least one earlier work (*Epidemiologia española* . . .) published in 1802, but the book's subtitle (. . . *ó Historia cronológica de las pestes, contagios, epidemias y epizootias* . . .) makes clear that this "discourse about epidemics" applies not to epidemiology in any modern sense—an approach to understanding the distribution and determinants of diseases in populations—but to the same compiling of chronological lists of important epidemics. Nevertheless, *Epidemiologia española* fits nicely into a then-growing tradition in medical geography. In the hands of its greatest practitioners, for example, Noah Webster (1758–1843), August Hirsch (1817–1894), and Charles Creighton (1847–1927), medical geography eventually joined medical history—practiced by physicians such as Justus Friedrich Karl Hecker (1795–1850), who wrote histories of plague, *tanzwuth*, and other epidemics—to converge with "true" epidemiology, exemplifying some of the difficulties in tracing epidemiology's roots.

The Early Historic Era

The early historic era corresponds to the appearance of planned agricultural communities, the domestication of animals, and eventually urban crowding, all of which were undoubtedly associated with the first infectious disease epidemics. Such diseases (e.g., measles, tuberculosis) probably arose by host switching of enzootic infectious agents in domestic animals to epidemic/endemic forms in man. With no scientific basis for understanding the resulting outbreaks and epidemics, these human diseases were attributed to angry gods and were dealt with by priests and healers (often the same individuals), whose organized rituals to placate the responsible deities aimed to keep the community safe rather than to investigate causality or even provide rational explanations. Apparent epidemic diseases were noted in Old Testament Biblical times (e.g., the Pharaonic "plagues" in the Book of Exodus, written ca. 1552 BCE, some of which are now speculated to have been infectious). Even then, a growing conceptualization of contagion led to public health actions to isolate persons with ostensibly contagious diseases (e.g., *leprosy*, a term that probably subsumed a variety of skin conditions, many of them, such as psoriasis and eczema, being noncontagious). An early "clinical trial" showing the effect of diet on health was even documented in the Book of Daniel (Book I, pp. 3–16, written ca. 167 BCE).

The Greco-Roman Era

The ancient Greeks made two major conceptual contributions to epidemiology: (1) the observation, attributed to Hippocrates (ca. 460–377 BCE) and outlined in *Airs, Waters, Places*, that epidemic diseases are distinct entities that appear in patterns of either pathognomonic or constellationary clinical symptoms and under particular conditions of season, weather, and geological events and (2) the theory propounded by Democritus (ca. 460–370 BCE) that contagious diseases are spread by tiny invisible particles. One of the most infamous epidemics in human history, the "Plague of Athens" (430–425 BCE), which occurred during the Peloponnesian Wars and may have brought about the end of Greece's Golden Age, was well described by Thucydides (ca. 460–400 BCE). Although the disease has not been conclusively linked to any disease known today, Thucydides's account is a landmark in epidemiology because it represents the first recorded attempt to specifically characterize a disease as a distinct entity on the basis of its clinical and epidemiologic pattern. Thucydides's account includes discussion of variations in disease presentation, distinction

between signs/symptoms and complications, discussion of attack rates, mortality, suspected risk factors, suspected modes of transmission, and other epidemic features and details never before chronicled. In almost obligatory fashion, Thucydides's description came to be discussed and analyzed in countless accounts spanning the centuries since the events in question, still having a profound impact on thinking about epidemics and epidemiology well into the 19th century. Discussions and reviews of the Plague of Athens are still fairly common today, perhaps because of the tantalizing mystery of a disease so richly described yet still escaping attempts to identify it.

Another noteworthy development in this era, credited to the Roman Empire, is the organized population census. When national censuses became commonplace nearly two millennia later (the late 1700s to late 1800s), they soon became essential epidemiologic tools, providing population denominators for calculation of attack and mortality rates, allowing disease occurrence to be described in quantitative terms, and leading to the identification of demographic groups at risk.

The Middle Ages

Although there were no breakthroughs in proto-epidemiologic thinking in the first millennium AD, the lessons of the Greco-Roman Era were not entirely forgotten. In imitation of Hippocrates and Thucydides, epidemics were still being chronicled, however simplistically, and a few of them eventually came to be recognized as distinct nosologic, if not etiologic entities. Disease distinction was a critical first step in the development of epidemiology because without it there could be no basis for studying disease determinants or distributions. Persian-born physician Abū Bakr Muhammed ibn Zakarīyā al-Rāzī ("Rhazes"; AD 860–932), for example, clearly distinguished smallpox from measles. Three centuries thereafter, the Scotsman Bernard de Gordon (ca. AD 1260–1318) drew on Rhazes and the Persian Abu-'Ali al-Husayn ibn Abd Allah ibn-Sina ("Avicenna"; ca. AD 980–1037) to list eight distinct diseases he thought were communicable (six of which he identified correctly: anthrax, *ignis sacer*—probably corresponding to erysipelas, although the term was later applied to zoster and ergotism as well—leprosy, plague, trachoma, and tuberculosis).

By the Middle Ages, there was growing suspicion that because epidemic and endemic diseases seemed to be different and to exhibit characteristic clinical features and patterns of spread, they might be distinguishable from each other, and might, furthermore, have different causes and be responsive to different treatment or control measures. However, there was yet no theoretical or scientific basis for identifying those causes, a reality that—as much as anything else—prevented the emergence of epidemiology.

The most significant epidemiologic event of the Middle Ages was the pandemic of "Black Death" (bubonic/pneumonic plague) that swept much of the known world in the 14th century. The pandemic wiped out an estimated 40 million people overall, including a third of Europe's population, making it the most fatal catastrophe in human history up to the time of the 1918 to 1919 influenza pandemic (50 to 100 million deaths). Historians have shown that this plague pandemic had a profound and complex impact on the subsequent course of human history, leading directly to the European Renaissance and its emphasis on scientific thought and critical reasoning. In addition, it opened an important chapter in public health and created a role for proto-epidemiologic activism. For example, the pandemic's arrival in Europe in 1347 demanded a public health role for governments. Before the pandemic, public health activities such as isolation of ostensibly contagious persons usually fell to religious and charitable orders. But spread of the pandemic from overseas locales into European seaports emphasized and broadened the concept of contagion, thus adding an international dimension to it.

This threat demanded civic action, leading to establishments of the first quarantine (from the Italian *quaranta*, referring to the customary 40 days of vessel sequestration in ports). Moreover, Europeans demanded that planning to prevent epidemic importation become a civic responsibility, as was made clear in the preventive medicine–oriented tractate of Jehan Jacme ("Jacme d'Agramont," d. 1384) dated April 24, 1348. Jacme's tractate, republished and widely read throughout Europe for decades thereafter, acknowledges the several contagious diseases of de Gordon and his predecessors, and goes on to exhort the city fathers as follows:

> Everyone must reflect on and prevent the causes that can produce a universal or local pestilence [an epidemic or an outbreak]. And if perchance these can be removed they must be removed. And to this

> end an effort should be made by the Lords [municipal officials] and their officers whose duty it is to look after the usefulness and well-being of the community. (Jacme, 1348, unpaginated)

Without explicitly realizing it, the generation that suffered the Black Death first articulated a rationale for civic actions that would someday make epidemiology a necessary tool for maintaining the public health.

The Enlightenment

Until the era referred to as The Enlightenment, the historical antecedents of epidemiology had been largely confined to recording, cataloguing, and attempting to distinguish between human diseases, supplemented in the more recent centuries by increasing attempts to prevent diseases whose course of spread might be anticipated, and to investigate those that were not prevented. Around the year 1700, corresponding to the beginning of The Enlightenment, a profound rethinking of all things scientific and phenomenological began to propel proto-epidemiology inexorably forward.

Among the many important events of this era and the decades immediately preceding it were the efforts of John Graunt (1620–1674) and Sir William Petty (1623–1687) to quantify and identify patterns in the causes of death recorded in the English Bills of Mortality; the discovery in 1675 by Sir Edmund Halley (1656–1742) that errors in measuring the positions of stars were actually nonrandom, in fact normally distributed, indicating a concealed order within a universe supposed to be incomprehensible to mortals; the discovery in the same year of tiny living organisms under the microscope by Antonie van Leeuwenhoek (1632–1723); and the development by scientists, such as François Boissier de Sauvages de la Croix (1706–1767), Carl von Linné (1707–1778), and William Cullen (1710–1790), of nosologic schemes to classify and differentiate human diseases. Such nosologic schemes were of enormous importance to epidemiology because in the absence of etiologic concepts, they provided a theoretical basis for differentially diagnosing, treating, and initiating epidemiologic responses to diseases. In essence, in the early Enlightenment, disease nosology was the precursor of etiology and served as an organizational framework for the explosion of knowledge in all fields relating to biomedicine.

Given that epidemiology largely developed and grew in response to the challenges of infectious diseases, it might be considered ironic that some of the first "classic" outbreak investigations were conducted in response not to infectious but to occupational diseases. This was so because in the late 1600s and early 1700s, when the first such studies were reported, it was possible to identify cause-and-effect linkages between certain distinctive diseases and prior occupational/industrial exposures to known industrial agents, whereas, lacking any concept of an infectious agent, cause-effect associations for the more common and more serious infectious diseases were not yet possible.

The 1697 outbreak investigation of Eberhard Gockel (1636–1703) linking *colica pictonum* to lead poisoning in mine workers (20) and the corpus of occupational disease studies of Bernardino Ramazzini (1633–1714) published several decades later anticipate modern epidemiologic studies. A 1751 clinical trial conducted by British Navy physician James Lind (1716–1794) showing that citrus juice could prevent scurvy (an occupational disease of sailors, then of unknown cause) is particularly noteworthy for the use, uncommon up to that time, of controlled experimental methods (clinical trial) in determining cause-effect relationships.

If occupational proto-epidemiology matured earlier than its infectious disease counterpart, an explosion of new information and ideas soon caused the former to be overtaken by the latter. Shortly after Leeuwenhoek's discovery of microorganisms (1675), the mites of scabies and mange were seen under the microscope. While we now understand these conditions to be infestations rather than infections, the impact on thinking about epidemic diseases was profound. Acceptance of contagion mediated by microscopic living entities was soon crystallized by a century-long series of European cattle epizootics, caused largely by rinderpest, and to a lesser extent by foot-and-mouth disease and anthrax. Animal diseases provided better opportunities for controlled study and implementation of control measures, human quarantines being difficult to implement in societies that increasingly demanded freedom of citizens from government interference.

The comprehensive public health responses to these epizootics were targeted specifically to preventing the spread of hypothesized microscopic organisms, for example, by burning and deeply burying corpses of dead animals, by disinfecting fomites, and by forming *cordons sanitaires.* Students in veterinary schools were deputized to conduct epizootic outbreak investigations and initiate control measures long before such activities took place in medical schools and long before outbreak

investigators were to be found in any major city. Particularly in France, several successive decades of attempts to investigate and control anthrax led to a fairly sophisticated epidemiologic understanding of the disease by the 1770s, a century before the causative organism was conclusively identified. By the latter 18th century, effective public health measures to contain these devastating epizootics had been refined and widely implemented, and the epizootics had declined in frequency and severity.

An inspirational and transitional figure at the end of The Enlightenment, intimately involved in epizootic investigation/control, was the physician-anatomist Félix Vicq-d'Azyr (1746–1794), arguably the foremost proto-epidemiologist of his era and one of the more obvious "fathers" of epidemiology. Ennobled shortly after the ascension of Louis XVI, Vicq-d'Azyr spent 2 years in the field conducting epizootic investigations throughout France and was then named "Perpetual Secretary" of the newly created Société royale de médecine. Vicq-d'Azyr quickly set up the world's first international (indeed, national) disease surveillance system, based on "correspondents" who reported regularly from the various provinces and overseas locales, not only on epidemic and epizootic diseases but also on climatic, atmospheric, and other environmental conditions suspected of being associated with their prevalences.

Like the Epidemic Intelligence Service of the U.S. Centers for Disease Control and Prevention (CDC), Vicq-d'Azyr ordered and conducted outbreak investigations either from Paris or the provinces themselves, and from his base as a professor at the Alfort veterinary school outside Paris, he conducted some of the first experimental attempts to transmit diseases in animals. The transmission efforts largely failed, but some 20 years after the revolution (and Vicq-d'Azyr's death), similar experiments were taken up more successfully by the next generation of Alfort investigators, under the influence of Vicq-d'Azyr's work. This eventually led to the characterization of the first epidemic infectious disease, anthrax, by Casimir-Joseph Davaine (1812–1882) and Robert Koch (1843–1910), between 1850 and 1876, and to the intimately related births of microbiology and experimental medicine.

Although few were probably aware of it at the time, the concept of contagion was being refined and advanced throughout The Enlightenment; in retrospect, it is clear that progress was moving in the direction of conceptualizing infectious microbial agents. As noted above, the idea that epidemics were caused by tiny transmissible agents arose in ancient Greece, and was further refined in the Renaissance and Enlightenment by theorists such as Girolamo Fracastoro (1484–1553), who in 1546 conceived of disease propagation by living *seminaria*; Athanaseus Kircher (1602–1680), who propounded a theory of contagious microorganisms in 1658; and Marc Anton von Plenciz (1705–1786), who in 1762 hypothesized specific and distinct contagious *animata* for each different human, animal, and plant disease. At the same time, there was growing appreciation that specific infectious diseases were often followed by lifelong refractoriness to the same disease, but not to other diseases, the observational basis of acquired immunity. Such concepts had become so firmly established by 1720 that when smallpox inoculation was introduced that year from Turkey into Europe and the American colonies, it was not only comprehended but widely accepted, a conceptual advance of such importance that it motivated American scientist-physician Benjamin Franklin (1706–1790) to document inoculation efficacy in Boston by conducting one of the first known historical cohort studies.

At the end of The Enlightenment, as the Industrial Revolution began to swell urban populations, creating fertile ground for urban epidemics, seemingly unrelated events were converging to give birth to "modern" epidemiology. These included the establishment of municipal and national censuses, early surveillance and disease-reporting systems, mounting evidence that contagious diseases were caused by replicating microorganisms, the acceptance in Europe and America of smallpox inoculation (1720), and then of Jennerian vaccination (1798)—the latter implying both cross-species infection and specific immunity—experience with outbreak investigations for a variety of human and animal diseases, experience with quarantine and disinfection, development of comparative and experimental medicine, establishment of public health training for physicians (particularly in France), and, especially in the German states, the growth of state-sponsored preventive medicine ("medical police").

After the Napoleonic wars, a generation of physicians returned to their European homes, some with limited training in the radically new subjects of public health and social medicine, and many of them ready to take up a new type of medical practice aimed not just at curing but also at preventing diseases in all segments of the population, prominently including the poor and disadvantaged.

The Birth of Epidemiology: Paris, 1819 to 1832

Among these returning military physicians were two who would lead, in the 1820s and 1830s, a small group of like-minded Parisians in studying the incidence and prevalence of various health conditions in populations, making use of census data and medical arithmetic, and identifying disease-specific demographic risk factors such as sex, age, locale, occupation, crowding, socioeconomic status, and so on. In essence, these physicians systematically studied the distribution and determinants of diseases in open populations, the first true epidemiology.

Louis-François Benoiston de Châteauneuf (1776–1856) and his young protégé Louis-René Villermé (1782–1863), along with colleagues such as Alexandre-Jean-Baptiste Parent-Duchâtelet (1790–1836), have sometimes been referred to as ''hygienists,'' but their role in inventing modern epidemiology from proto-epidemiologic antecedents is clear from perusal of their works, for example, the 536-page report of the 1832 Parisian cholera pandemic. This work, in fact, appears to have been inspirational reading for young medical students such as John Snow (1813–1858) and William Farr (1807–1883), the latter of whom trained in France, and both of whom would make indelible marks on epidemiology in the next (third) cholera pandemic less than 20 years later. Though not part of this epidemiology circle, the hospital-based Paris clinician Pierre-Charles-Alexandre Louis (1787–1872), immortalized decades later by Sir William Osler (1849–1919) as one of the ''fathers'' of American medicine, was working in parallel with Benoiston de Châteauneuf and his colleagues by using structured observations and medical arithmetic to improve clinical therapy. Thus, the forces that led to the first population-based epidemiology and the first clinical epidemiology, though separate and distinct from the beginning, arose at the same time, in the same place, and in the same European scientific and cultural milieu.

Early Epidemiology Responds to Discovery of Infectious Diseases

Political events in France soon undermined the further development of epidemiology there, but not before influencing health workers in other countries, many of whom had studied medicine in Paris or had at least been reading French texts and journals. In the 1830s, the British Sanitary Movement led to establishment of a national Register of Births and Deaths in England and Wales. By this time, many major European and American cities (e.g., Hamburg) had or were developing sophisticated public health data collection systems and were using them to address public health problems, endeavors that created a need for the further development of epidemiology and an environment in which it could grow and thrive.

Between about 1840 and 1866, the influence of epidemiology exploded in Europe and America. Among other notable epidemiologic milestones of the era, in 1840 the young German epidemiologist Friedrich Gustav Jacob Henle (1809–1885) published a comprehensive treatise that was decades ahead of its time in anticipating infectious diseases, a treatise that also suggested to one of Henle's students, Robert Koch, the principles that would later become ''Koch's postulates'' (or the ''Henle-Koch postulates'') for experimentally establishing that a disease is infectious. Many other young European physicians began to be influenced by epidemiology, seen, for example, in the outbreak investigations of Englishman George Budd (1808–1882) and the disease theories of his younger brother William (1811–1880), who later distinguished typhus from typhoid, and most famously in the studies and theories of John Snow, whose separate investigations of cholera incidence rates by contaminated and uncontaminated London water supplies and of cholera contamination at Soho's Broad Street pump were eventually recognized as landmarks. In Vienna, Ignaz Philipp Semmelweis (1818–1865) used epidemiology to demonstrate the cause and prevention of iatrogenic puerperal fever. In Denmark, the medical student Peter Ludwig Panum (1820–1885) investigated and published in 1847 a remarkably insightful treatise on a measles epidemic, further evidence that the generation that came of age during the birth of epidemiology in the 1820s and 1830s had grasped the importance of, and the great potential for, this new approach.

Many of the studies conducted by these early epidemiologists seem recognizably modern today (2007) because of their use of epidemiologic reasoning, calculations of frequencies and, when possible, incidence rates, and their inductive puzzle-fitting approaches that led inexorably from discrete facts and observations to inferences supporting public health control measures. By the mid-19th century, disease outbreaks

were being investigated with arithmetical techniques, epidemic curves were being drawn and studied, and while cohort studies were not yet common, incidence rates were being compared and contrasted to understand causation and risk variables based on the presence or absence of suspected risk factors.

The most significant conceptual breakthrough for epidemiology during the 19th century, however, was undoubtedly establishment of the "germ theory." Following efforts by David Gruby (1810–1898) and others in the 1840s, between 1850 and 1876 the remarkable researches of Davaine and Koch led to the establishment of the first human infectious disease: anthrax. In the 25 years between 1876 and the end of the century, there followed a burst of microbiological triumphs (e.g., establishment of the etiologies of tuberculosis, plague, and cholera) that led to new public health control measures and to clinical therapies, including vaccines (e.g., rabies, 1885), passive immunotherapies (e.g., diphtheria antitoxin, 1890), and to environmental control (e.g., for yellow fever, shown to be a mosquito-borne infection by Walter Reed [1851–1902] and colleagues).

These advances had a profound effect on epidemiology by creating the ability to identify the etiology and verify the presence or absence of the disease, to study the diseases in populations, and to study the efficacies of prevention and treatment modalities. Microbiology and epidemiology developed in a partnership that was so close that they were often thought of as two aspects of the same discipline. The growing sophistication of epidemiology, and its increasingly close relationship with microbiology, can perhaps be most clearly seen at one point in time and place: the remarkable body of American scientific data produced during the 1918 to 1919 influenza pandemic. As late as the 1920s, American epidemiologists working in health departments saw the microscope as one of the key epidemiologic tools and often used it on a daily basis.

Epidemiology After the Microbiological Era

The 1920s also saw epidemiology develop in response to opportunities coming from different directions, including infusions of interest and expertise from the social sciences, such as more rigorous/better standardized methodologies to link risk factors to outcomes by comparing incidence rates in exposed and unexposed persons (cohort studies), or by comparing prior exposure frequencies in ill and well persons (case-control studies, e.g., those of Janet Elizabeth Lane-Claypon [1877–1967]). In the United States, these developments were heavily influenced from about 1920 to about 1980 by the national experiment in establishing numerous academic schools of public health, and the development of academic population-oriented epidemiology under the leadership of rigorous methodologists such as Johns Hopkins University's Wade Hampton Frost (1880–1938), whereas European epidemiology still reflected a largely clinical orientation.

These changes in the development of epidemiology had many consequences, including increasing sophistication of epidemiologic methods and increasing attention paid to noninfectious disease problems, for example, early-20th-century studies of pellagra, a supposedly infectious disease proven to be a nutritional disease in a series of classic studies by Joseph Goldberger (1874–1929); the series of studies conducted by Sir Richard Doll (1912–2005) and Sir Austin Bradford Hill (1897–1991) on the effects of cigarette smoking in British physicians; the linking of dental caries to inadequate levels of halogens in natural water supplies; and the linking of specific cancers to specific environmental or occupational exposures (e.g., vinyl chloride/hepatocarcinoma, prenatal estrogen exposure/vaginal cancer in offspring). During these years, American epidemiology became a dominant force, in part because of its expansion in academic centers, its development in state and local health departments, and also because of the remarkable growth and influence of the U.S. Centers for Disease Control and Prevention (CDC), which supported and worked closely with health departments and contributed greatly to disease discovery and characterization.

From the vantage point of 2007, it seems likely that epidemiology will continue to expand in the direction of methodological orientation in the United States and also that in the United States and elsewhere, the need for basic practical epidemiology in the service of disease control will remain. Advances in genomics, proteomics, and bioinformatics are likely to strongly shape epidemiology. Although future directions are difficult to predict, it seems likely that the development of epidemiology will remain dynamic for the foreseeable future.

—David M. Morens

See also Budd, William; Etiology of Disease; Farr, William; Koch, Robert; Public Health, History of; Snow, John

Further Readings

Anonymous. (1963). Epidemi(olog)^{2}y * . *American Journal of Public Health, 53*, 82–84.

Benoiston de Châteauneuf, L.-F., Chevallier, J.-B., Devaux, L., Millot, L., Parent-Duchâtelet, A.-J.-B., de Maurienne, P., et al. (1834). *Rapport sur la marche et les effets du choléra-morbus dans Paris et les communes rurales du département se la Seine.* Paris: Imprimerie Royale.

Eyler, J. (1979). *Victorian social medicine: The ideas and methods of William Farr.* Baltimore: Johns Hopkins University Press.

Huxley, T. H. (1882). The method of Zadig. In T. H. Huxley (Ed.), *Science and culture, and other essays.* New York: D. Appleton.

Huxley, T. H. (1854). On the educational value of the natural history sciences (1854). In T. H. Huxley (Ed.), *Science and education: Essays by Thomas H. Huxley* (pp. 38–65). New York: D. Appleton.

Jacme, J. [aka Jacobi, J.; "Jacme d'Agramont"]. (1348). *Regiment de preservació a epidemia o pestilència e mortaldats.* Lerida: Council of Lerida.

Lilienfeld, A. M. (Ed.). (1980). *Times, places and persons: Aspects of the history of epidemiology.* Baltimore: Johns Hopkins University Press.

Rosen, G. (1993). *A history of public health* (Expanded ed.). Baltimore: Johns Hopkins University Press.

Vinten-Johansen, P., Brody, H., Paneth, N., Rachman, S., Rip, M., & Zuck, D. (2003). *Cholera, chloroform, and the science of medicine: A life of John Snow.* New York: Oxford University Press.

Winslow, C.-E. A., Smillie, W. G., Doull, J. A., Gordon, J. E., & Top, F. H. (1952). *The history of American epidemiology.* St. Louis, MO: Mosby.

Epidemiology in Developing Countries

The disparity in health across nations is dramatic, as depicted in Table 1 by a 50-year range in life expectancy at birth. Quality of life is also widely varied across nations. What causes these substantial differences in life length and quality? Economic status is a major factor, leading to nutrition- and environment-related risks, injuries, high-risk behaviors, and inadequate health care. These underlying factors cause the infectious diseases, childbirth complications, trauma, and chronic diseases that ravage some developing countries.

Familiarity with the measurements underlying the summary health profiles of each country is essential to understanding the epidemiology of developing countries. Additionally, study methodologies vary worldwide according to the respective needs and cultures, but they are largely limited by the lack of resources in these countries. Also, it is important to consider ethical issues that arise in international study, particularly studies conducted jointly with U.S. investigators.

This entry discusses three aspects of epidemiology in developing countries. The section on burden of disease provides an overview of the epidemiology of developing countries; the section on research methods reviews epidemiology training programs and methods for epidemiologic study in developing countries; and, finally, ethical considerations are discussed in a section on the key components of ethical research conducted jointly by developed and developing nations.

Burden of Disease

While overall mortality rates, years of potential life lost, and life expectancy provide important insight into the health of a population, none provides information about *quality* of life. Rather, they are limited to information about *quantity* of life. Yet disabilities are important considerations when understanding population health. To this end, the World Bank funded the Burden of Disease Study, a collaborative effort of the World Health Organization and university researchers to describe the burden of disease for nations, incorporating both quantity of life and quality of life. The research also pulled together information on the major risk factors for each nation's burden of disease.

Burden of disease is estimated by combining information measured by years of potential life lost (YPLL) and years of healthy life lost due to disability (YLD). This combination of information forms disability-adjusted life years (DALYs), which estimate the years of healthy life lost. For example, if a child dies of pneumonia at age 1, the child has 54 more years of YPLL compared to an individual who dies of pneumonia at age 55. Deaths in childhood have substantial impact on the measure of YPLL; thus, for countries with high childhood mortality, the YPLL and DALY are substantial per capita. Similarly, if an infectious disease, such as onchocerciasis (or "river blindness"), causes blindness of a 5-year-old, the YLD and DALY will be 39 years more than if the same disease causes blindness of a 44-year-old. Combining this information across all diseases in a country creates one summary measure for the country that provides a measure of health. The DALYs can also be computed by gender and for specific diseases and risk factors to understand their relative importance.

Table 1 Life Expectancy by Region and Country, 2006 Estimates

		Mortality Experience	
Region	*Countries (Life Expectancies in Years)*	*Child*	*Adult*
Africa	Angola (38.6)	High	High
	Liberia (39.7)		
	Sierra Leone (40.2)		
	Niger (43.8)		
	Guinea-Bissau (46.9)		
	Nigeria (47.1)		
	Chad (47.5)		
	Burkina Faso (48.9)		
	Mali (49.0)		
	Equatorial Guinea (50.0)		
	Guinea (50.0)		
	Cameroon (51.2)		
	Benin (53.0)		
	Mauritania (53.1)		
	Gambia (54.1)		
	Gabon (54.5)		
	Madagascar (57.3)		
	Togo (57.4)		
	Ghana (58.9)		
	Senegal (59.3)		
	Comoros (62.3)		
	Sao Tome and Principe (67.3)		
	Cape Verde (70.7)		
	Seychelles (72.1)		
	Mauritius (72.6)		
	Algeria (73.3)		
	Swaziland (32.6)	Very high	Very high
	Botswana (33.7)		
	Lesotho (34.4)		
	Zimbabwe (39.3)		
	Mozambique (39.8)		

(Continued)

(Continued)

		Mortality Experience	
Region	*Countries (Life Expectancies in Years)*	*Child*	*Adult*
	Zambia (40.0)		
	Malawi (41.7)		
	South Africa (42.7)		
	Namibia (43.4)		
	Central African Republic (43.5)		
	Tanzania (45.6)		
	Rwanda (47.3)		
	Cote d'Ivoire (48.8)		
	Kenya (48.9)		
	Ethiopia (49.0)		
	Burundi (50.8)		
	Congo Democratic, Republic of the (51.5)		
	Uganda (52.7)		
	Congo, Republic of the (52.8)		
	Eritrea (59.0)		
North and South America	Cuba (77.4)	Very low	Very low
	United States (77.9)		
	Canada (80.2)		
	Grenada (64.9)	Low	Low
	The Bahamas (65.6)		
	Guyana (65.9)		
	Trinidad and Tobago (66.8)		
	Belize (68.3)		
	Suriname (69.0)		
	Honduras (69.3)		
	El Salvador (71.5)		
	Dominican Republic (71.7)		
	Brazil (71.9)		
	Colombia (71.9)		
	Antigua and Barbuda (72.2)		
	Saint Kitts and Nevis (72.4)		
	Barbados (72.8)		

	Jamaica (73.2)		
	Saint Lucia (73.8)		
	Saint Vincent and the Gernadines (73.9)		
	Venezuela (74.5)		
	Dominica (74.9)		
	Paraguay (75.1)		
	Panama (75.2)		
	Mexico (75.4)		
	Argentina (76.1)		
	Uruguay (76.3)		
	Chile (76.8)		
	Costa Rica (77.0)		
	Haiti (53.2)	High	High
	Bolivia (65.8)		
	Guatemala (69.4)		
	Peru (69.8)		
	Nicaragua (70.6)		
	Ecuador (76.4)		
Eastern Mediterranean Region	Iran (70.3)	Low	Low
	Syria (70.3)		
	Lebanon (72.9)		
	Oman (73.4)		
	Qatar (74.0)		
	Bahrain (74.5)		
	Tunisia (75.1)		
	United Arab Emirates (75.4)		
	Saudi Arabia (75.7)		
	Libya (76.7)		
	Kuwait (77.2)		
	Cyprus (77.8)		
	Jordan (78.4)		
	Djibouti (43.2)	High	High
	Afghanistan (43.3)		
	Somalia (48.5)		
	Sudan (58.9)		

(Continued)

(Continued)

		Mortality Experience	
Region	*Countries (Life Expectancies in Years)*	*Child*	*Adult*
	Yemen (62.1)		
	Pakistan (63.4)		
	Iraq (69.0)		
	Morocco (70.9)		
	Egypt (71.3)		
Europe	Croatia (74.7)	Very low	Very low
	Czech Republic (76.2)		
	Slovenia (76.3)		
	Ireland (77.7)		
	Portugal (77.7)		
	Denmark (77.8)		
	Finland (78.5)		
	United Kingdom (78.5)		
	Belgium (78.8)		
	Germany (78.8)		
	Luxembourg (78.9)		
	Netherlands (78.9)		
	Malta (79.0)		
	Austria (79.1)		
	Greece (79.2)		
	Israel (79.5)		
	Norway (79.5)		
	France (79.7)		
	Monaco (79.7)		
	Spain (79.7)		
	Italy (79.8)		
	Iceland (80.3)		
	Switzerland (80.5)		
	San Marino (81.7)		
	Andorra (83.1)		
	Turkmenistan (61.8)	Low	Low
	Azerbaijan (63.9)		
	Uzbekistan (64.6)		

	Tajikistan (64.9)		
	Kyrgyzstan (68.5)		
	Romania (71.6)		
	Armenia (71.8)		
	Bulgaria (72.3)		
	Turkey (72.6)		
	Macedonia (73.9)		
	Slovakia (74.7)		
	Poland (74.9)		
	Georgia (76.1)		
	Albania (77.4)		
	Bosnia and Herzegovina (78.0)		
	Moldova (65.7)	Low	High
	Kazakhstan (66.9)		
	Russia (67.1)		
	Belarus (69.1)		
	Ukraine (69.9)		
	Latvia (71.3)		
	Estonia (72.0)		
	Hungary (72.7)		
	Lithuania (74.2)		
Southeast Asia	Indonesia (69.9)	Low	Low
	Thailand (72.3)		
	Sri Lanka (73.4)		
	Bhutan (54.8)	High	High
	Nepal (60.2)		
	Burma (60.9)		
	Bangladesh (62.5)		
	Maldives (64.4)		
	India (64.7)		
	North Korea (71.7)		
	Brunei (75.0)		
Western Pacific Region	New Zealand (78.8)	Very low	Very low
	Australia (80.5)		

(Continued)

(Continued)

		Mortality Experience	
Region	*Countries (Life Expectancies in Years)*	*Child*	*Adult*
	Japan (81.3)		
	Singapore (81.7)		
	Laos (55.5)	Low	Low
	Cambodia (59.3)		
	Kiribati (62.1)		
	Vanuatu (62.9)		
	Nauru (63.1)		
	Mongolia (64.9)		
	Papua New Guinea (65.3)		
	Tuvalu (68.3)		
	Fiji (69.8)		
	Tonga (69.8)		
	Federated States of Micronesia (70.1)		
	Philippines (70.2)		
	Marshall Islands (70.3)		
	Palau (70.4)		
	Vietnam (70.9)		
	Samoa (71.0)		
	Malaysia (72.5)		
	China (72.6)		
	Solomon Islands (72.9)		
	South Korea (77.0)		

Source: Based on life expectancy data from https://www.cia.gov/cia/publications/factbook/rankorder/2102rank.txt; categories developed by the Global Burden of Disease Study (DOI: 10.1371/journal.pmed.0010027.t002)

Approximately 58 million of the 6.5 billion people in the world died in 2006. Shockingly, about 20% of these deaths are children, disproportionately from developing countries. Developing countries also have substantial deaths in the adolescent age range, while wealthy countries carry a relatively high death rate among those in young and middle adulthood (ages 15 to 59 years). Infectious diseases, particularly respiratory and diarrheal diseases, maternal and perinatal conditions, and malnutrition, continue to play an important role in mortality for the people of developing countries. Malaria ravages sub-Saharan Africa, currently the second leading cause of death after HIV/AIDS in that region. Tuberculosis also is a growing cause of death and disability in developing countries worldwide, becoming increasingly lethal due to drug resistance and coinfection with HIV. However, many developing countries have undergone a key transition in mortality: With the exception of sub-Saharan Africa and South Asia, chronic diseases now outpace

infectious diseases as the cause of death. Heart disease and cerebrovascular disease are now the leading causes of death. Injuries, including violence, occupation-related injuries, falls, and traffic crashes, also take their toll in lives.

Infectious diseases, malnutrition, chronic diseases, and trauma also impact the quality of life for large portions of developing-country populations. Additionally, mental disorders not typically related to substantial loss of life can have a tremendous impact on the quality of life and, thus, contribute to the measures of DALYs. In the 2001 Global Burden of Disease Study, it became clear that neuropsychiatric conditions (e.g., depression, post-traumatic stress disorder, senility), vision problems and hearing loss, and alcohol disorders had substantial impact on the DALYs, but very little impact on mortality rates. It is estimated that psychiatric conditions account for over 35% of YLD among those 15 years of age and older worldwide. While these conditions are prevalent in developed countries, they have substantially more impact on quality of life in developing countries, as measured by YLD and DALYs. The same factors that predispose people in developing countries to infectious and chronic diseases, including poverty and illiteracy, are also risk factors for psychiatric illnesses.

Adding to the burden of disease, countries that were not historically considered "developing" now meet this description. The post-Soviet East has suffered substantial declines in health since independence starting around 1990. The burden of disease in Russia and Central Asia has increased by over 30% between 1990 and 2000, with this increased disease burden disproportionately affecting men, likely due to excessive alcohol use.

Historically, assessment of population health has focused on the immediate causes of morbidity and mortality, but more emphasis is now being placed on the underlying risk factors, or what are sometimes called the actual causes of death and disability. In 2000, over 15% of DALYs were caused by childhood and maternal malnutrition, including deficiencies in weight, iron, vitamin A, and zinc. Poor eating habits and sedentary lifestyle accounted for about 11% of DALYs, followed by unsafe sex (7%), addictive substances (e.g., tobacco, alcohol, illicit drugs; 9%), environmental risks (e.g., unsafe water, pollution, lead; 7%), and occupational risks (e.g., injuries, carcinogens; 1%). These figures represent worldwide DALYs attributable to such risk factors; however, they are disproportionately important in developing countries. These risk factors are also generally reversible.

Research Methods

Epidemiology Training Programs

First and foremost, developing countries need trained professionals to help build the public health infrastructure and carry out research projects. This mission is championed in the United States by the Fogarty International Program, funded by the National Institutes of Health. The development of the public health infrastructure provides the best opportunity for future progress in health. Physicians typically work with academic epidemiologists in the United States for 6 months to 2 years, including completing course work and developing research projects jointly, followed by mentored research, once formal training is complete. Similarly, the Fulbright scholarship program affords physicians and other health workers the opportunity to experience a well-developed health system and learn about research program implementation from academic and research institutes.

The U.S. Centers for Disease Control and Prevention (CDC) also trains physicians in developing countries through the Field Epidemiology Training Programs (FETPs). These programs are designed to teach applied epidemiology from a public health perspective, with a particular focus on disease control. Epidemiologists from the U.S. CDC work with trainees in their home countries throughout the training and collaborate on public health problems as they arise. Specific training in disease surveillance systems and outbreak investigations are central to the programs. Trainees also are encouraged to develop studies for a range of disease areas relevant to their country's health issues. A particular benefit of the FETPs is the opportunity for epidemiologists in training to network with epidemiologists from across the world. The potential for rapid global transmission of communicable disease outbreaks (e.g., avian influenza) makes it more important than ever to have trained epidemiologists in developing countries who are also connected to a worldwide network of epidemiologists focused on disease control. Furthermore, because the developing country's public health system is typically more disadvantaged than clinical practice, a concerted effort to improve public health

first will advance the infrastructure of the entire health system.

Epidemiology Research in Developing Countries

Epidemiologic research should be driven by the priorities of health officials and scientists from the local public health community. Collaborations that will be fruitful for the country are likely to be successful. Other hallmarks of a successful collaboration include respect for team members' scientific knowledge and approach, and the ability to integrate beliefs. One way to ensure a respectful collaboration is to allow local colleagues to direct the research agenda. While discussion of priorities is always appropriate, invited researchers should be prepared to put their preconceived ideas and plans aside and consider research studies recommended by local scientists.

After discussion with local scientists, the next step in research topic selection is a good literature review. While much information from developing countries is not published in journals, local agencies and scientists may be knowledgeable about previous work and can help identify available information. Experienced epidemiologists recognize that decentralized organizations often have a remarkable amount of data and that preliminary studies may already have been conducted. Once a complete literature review has been performed, supplemented by a search for other sources of relevant data, the study methods can be thoughtfully developed.

Organizational Challenges

Many developing countries regularly undergo political transition. For countries in transition, the Ministry of Health structure also may change frequently resulting in changes in funding priorities and administrative structure that can affect the conduct of research. Thus, conducting even a 5-year follow-up study risks interruption. While impossible to fully anticipate, understanding the administrative potential for research, particularly longitudinal research, is important to contemplate the study design process.

Another challenge is the lack of available public health training and disinterest in public health positions. Financial incentives are often greater in clinical practice, pharmaceutical companies, and Western nongovernmental organizations than in public health research positions. In some developing countries, only physicians can serve as epidemiologists, further limiting the pool of qualified professionals. Additionally, young researchers trained in the West compete for positions in organizational structures that historically honored seniority, creating tensions in the transition to new methods of research.

Methodological Challenges

Study Design

In general, classic epidemiology study designs are the foundation of research in developing countries. Randomized trials remain the hallmark, and this design is recommended when appropriate. When research ethics prohibit randomization, cohort studies, case-control studies, cross-sectional studies, and more complex forms of these designs (e.g., case-cohort, nested case-control) are appropriate to consider.

Cross-sectional designs are often an appropriate starting point, due to their ability to provide basic disease and exposure distribution information, often have a low cost, and provide relatively quick implementation. Additionally, many of the logistics needed to successfully conduct more complex studies can be tested and developed in a cross-sectional design.

Data Collection Options and Validity

Most self-report instruments have been validated only in developed countries and may have markedly different validity in developing countries; thus, local validity studies need to be conducted before measures are used in epidemiologic research. For some measures, this validation is simple and inexpensive, but for other important exposures, such as risky behaviors, it is extremely difficult. For example, in studying injection drug use behaviors as a risk factor for HIV, self-report validity appears good if study participants are identified in a high-risk setting such as a needle-exchange program. Individuals have essentially already identified themselves as injection drug users by using the services; providing information for research about their drug use does not put them at much added risk. These same questions included on a survey for the general population, health care workers, or patients are unlikely to provide valid data on injection drug use. To date, there is no validated questionnaire that provides accurate data on this type of high-risk

exposure in the general population for most regions in the world.

Worldwide, issues of social desirability affect survey responses on sensitive topics. The sensitivity of questions can vary tremendously between countries; thus, researchers need to understand the culture to understand how to collect data and interpret data. For example, in some countries where women are still expected to live traditionally conservative roles in relationships, it is difficult to obtain accurate information about risky sexual behavior. Even if interviews are done by qualified interviewers or the questionnaire is self-administered, we rarely see reports of multiple sex partners from women. For men, multiple partners are more acceptable and the responses about sexual behavior may have stronger validity. Interestingly, in some countries where women are expected to have traditional roles, abortion is widely accepted so information about abortion history (e.g., number of abortions) may be valid. Thus, for accurate measurements to be developed, it is important to understand not only local customs but also the implications of social acceptability of behaviors.

Other data collection challenges are less obvious. For example, in at least one developing country, a child's immunization record may include sign-offs by physicians for vaccines that were never administered. While the frequency of this is unknown, anecdotal evidence suggests that it clearly happens because the caregiver perceives that the administration of vaccine poses a greater risk than the disease due to the reuse of needles and use of multidose vials of vaccine. Lack of proper staff training and monitoring and limited resources can also affect the extent to which the researcher can measure the exposure and outcome accurately.

Epidemiology research methods are comparable for both developed and developing countries, but the methods to capture reliable and valid data may differ. Many developing countries do not have the advanced databases and technologies required for some studies. Conversely, some approaches that are cost prohibitive in developed countries are feasible in developing countries The cost of staff salaries is often substantially (e.g., 10-fold) less in developing countries, while costs of laboratory tests, medical treatments, and computer technologies tend to be comparable with those of developed countries. This inversion of the typical cost structure of a developed country provides an opportunity for researchers to select different methods in developing countries. For example, given a relatively small budget, it may be cost prohibitive to interview 2,000 study participants in a study conducted in the United States, but it may be quite feasible with a relatively small grant in a developing country. In general, medical records useful for epidemiologic research are less likely to be available in developing countries, routine laboratory tests are more likely to have poor sensitivity and specificity, and medical care may be provided through non-Western approaches, making it difficult to measure.

Ethical Issues in Joint Research Between Developing and Developed Countries

Research conducted in developing countries jointly by scientists from developed countries has raised substantial ethical issues and controversies during the past two decades. The issues are important and complex; it is rare for a study to be designed without significant focus on clarifying the ethical issues and reviewing approaches to address these issues.

One ethnical question concerns what should be considered a reasonable standard treatment for which an alternative treatment can be compared in a randomized trial. Early in the HIV epidemic, when it became clear that antiretroviral medications could reduce vertical (mother-to-infant) transmission, researchers decided to conduct a randomized study in Africa that compared no treatment (the current standard in the region) to a single antiretroviral medication (below the standard of care in the United States at the time). The rationale for the study was that the standard of care in the region was no treatment, and if a benefit for the single medication was efficacious, then it might be economically feasible to implement a prevention program based on the study's findings. The study started a controversy that continues to this day. Some believe that there should be one universal "best practices" and that this is the standard from which comparisons should be made. Others disagree. While the controversy continues, a substantial case has been made that researchers from developed countries should not conduct research that uses treatments that are not and would not be acceptable in their home countries.

A second ethical question concerns epidemiologic studies that often use diagnostic tests to identify individuals with infections of interest. If the study does

identify infected individuals but the researchers do not have funding to cover treatment that would be available in developed countries, the ethics of the study can come into question. The details of the study and the related ethical issues need to be carefully reviewed to determine if the study should be conducted.

These are just two examples of significant ethical issues that can arise when research is conducted in developing countries. Even though ethical considerations related to the conduct of human subject research are universal, economic and cultural diversities may result in different ethics issues disproportionately arising more often in developing countries than in developed countries.

Informed Consent

One of the issues all researchers struggle with is obtaining truly informed consent. For example, HIV voluntary counseling and testing (VCT) programs, such as prevention of mother-to-child transmission (PMTCT) for pregnant women, raise concerns about participant recruitment. Women might be too aggressively encouraged to accept VCT by obstetricians and not given sufficient information about their right to refuse. However, most practicing physicians in post-Soviet countries worked during the Soviet era when voluntary services and extensive counseling were not characteristic features of the health system, and HIV testing was mandatory for some population groups, including pregnant women. This is why during the training sessions for health care workers, it is very important to emphasize the principles of voluntary counseling and testing, the concept of informed consent, and the rights of patients to refuse HIV testing and study participation.

Regional variation exists in the comfort level individuals may feel about agreeing to participate in a study without speaking to personal advisors. In some Asian countries, for instance, it is difficult to obtain informed consent from the potential study participant directly, without discussion with family members. Before signing informed consent, study participants often need time to review the opportunity with their families (spouses and even in-laws). In some regions of Africa, tribal leaders need to be consulted before individuals are willing to entertain study participation. In these instances, individual autonomous informed consent is still required. However, understanding the culture provides insight into how to provide a process that maximizes the opportunity for individuals to utilize information sources that are important to them for making decisions. Once these discussions occur, potential study participants need to have an environment when an autonomous decision can be made and communicated to researchers. Researchers need to ensure that such an opportunity exists in the informed consent process.

For collaborative research, institutional review board (IRB) approvals from both the developed country and the local institutions are required. In some countries, research ethics review is strong; in other countries, little experience in reviewing research ethics exists, and the committees tend to be token bodies and lack the expertise to thoroughly review and enforce the protection of human rights. Instead, they may err on the side of simply reviewing the paperwork and records before granting approval; they are unable to thoroughly investigate the human subjects' protection for each study. Further complicating the review process, the IRB members of developed counties often have no international research experience, have never visited developing-country institutions, and often cannot understand the barriers and challenges for research in the developing world. The questionnaire, consent forms, and research protocol are submitted to the U.S. IRBs in English, and their subsequent comments are sometimes hard to translate back to the native language adequately for implementation.

In some cases, verbal consent should replace signed consent. When the study is conducted among less educated persons who cannot read and understand the content of informed consent, the interviewer should carefully explain and obtain verbal consent from the study subject.

Financial Compensation

Developing countries are especially vulnerable with respect to the ethics of financial compensation for participation in the study. In countries where people have monthly income below US$50, it is very easy to obtain consent for participation in research with minimal compensation. This makes potential study subjects vulnerable to coercion and requires careful consideration when offering compensation. Money should not be at a coercive level. Another factor that may potentially influence the participation decision is the potential participant's desire to please his or her doctor. Even in developed countries, there

may be a generational difference in the tendency to accept a physician's direction without question. While not an obvious financial issue on the surface, access to a physician and his or her services may be a financial necessity, thus making the physician's interest compelling for a potential participant.

Trust

It is often problematic to gain the trust of the local community and individual study participants in developing countries. Local researchers born and raised in these communities have a good sense of how to build trust, and they should make every effort to do so before starting the research. People in developing countries often do not believe that the goal of epidemiologic research is ultimately to improve health outcomes in their region. Their first concern is the fairness of becoming a participant of experiments for research of interest only to the scientists of the developed country. For instance, in the 1980s, when HIV/AIDS was the overwhelming concern in the United States and Europe, researchers from these developed regions pushed to have funding strongly realigned with their priorities. At the same time, many sub-Saharan Africa researchers requested that funding be expanded for malaria, a current cause of tremendous morbidity and mortality. However, these requests largely went unanswered until recently, negatively affecting the ability to build strong, trusting relationships. The researcher must present the benefits and risks of participating in a culturally acceptable manner and ensure that potential recruits feel free to ask questions about their doubts and concerns. The development of trust typically starts with researchers and leaders of a community and then expands to potential study participants. The process of building trust, on both sides, involves communication that allows for relevant information about the society and research to progress until all involved find the proposed research a value to the community. It is at this point that the research process can move forward to protocol development and study initiation.

Ethics issues related to international research continues to be an evolving area of thought. Epidemiologists need to routinely reflect on foundational documents of ethics philosophy, such as the Belmont Report, and update their review of the ethics literature as it pertains to their specific area of study. Many epidemiologists also require the study methods and procedures to pass two additional tests: (1) Would the study be acceptable for a family member? and (2) "Can I sleep at night?" used by many researchers to assess their peace of mind with the decisions made. Only studies consistent with the Belmont Report or other research ethics foundational documents, laws of all jurisdictions involved, and these last two tests should be considered.

—Maia Butsashvili, Ghazwan G. Toma, and Louise-Anne McNutt

See also Cultural Sensitivity; Ethics in Public Health; Health Communication in Developing Countries; Health Disparities

Further Readings

Lopez, A. D., Begg, S., & Boss, E. (2006). Demographic and epidemiological characteristics of major regions, 1990–2001. In A. D. Lopez, C. D. Mathers, M. Ezzati, D. T. Jamison, & C. J. L. Murray (Eds.), *Global burden of disease and risk factors* (pp. 17–44). Washington, DC: International Bank for Reconstruction and Development/World Bank. Retrieved July 22, 2007, from http://www.dcp2.org/pubs/GBD.

MacQueen, K. M., Karim, Q. A., & Sugarman, J. (2003). Ethics guidance for HIV prevention trials. *British Medical Journal, 327*, 340.

MacQueen, K. M., & Sugarman, J. (2003). Back to the rough ground: Working in international HIV prevention research as ethical debates continue. *IRB: Ethics Hum Res*, *25*(2), 11–13.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. Retrieved April 18, 1979, from http://ohsr.od.nih.gov/guidelines/belmont.html.

Nelson, K. E., & Williams, C. M. (Eds.). (2006). *Infectious disease epidemiology: Theory and practice* (2nd ed.). Boston: Jones & Bartlett.

Rodgers, A., Ezzati, M., Vander Hoorn, S., Lopez, A. D., Lin, R.-B., & Murray, C. J. L., for Comparative Risk Assessment Collaborating Group. (2004). Distribution of Major Health Risks: Findings from the Global Burden of Disease Study. *PLoS Medicine, 1*(1), e27.

Epigenetics

When one considers inheritance, typically it is at the level of the DNA, gene, or chromosome: typical Mendelian genetics. However, not all hereditary

information is passed on in our genes. The field that studies these heritable differences in gene expression that do not involve a change in the DNA sequence is called *epigenetics.* This field can include studies of how gene expression is modified as cells divide and differentiate into specific types; how environmental factors (such as chemicals or radiation) change the way genes are expressed; and how patterns of gene expression are passed on from parent to daughter cells.

All multicellular organisms start off from a single cell, which contains all the genetic information present in an organism. However, from this single cell, vastly different daughter cells arise, eventually going on to form all the different tissues in our bodies. One way these differences are possible is via epigenetic regulation of individual cells. The first phenomenon of this type to be described was genetic imprinting, whereby one set of genes in a diploid individual is silenced. This silencing is based on its parental origin: either maternal or paternal. Epigenetic modifications can also include methylation (the addition of methyl groups) to the DNA backbone. The addition of these methyl groups changes the structure and appearance of the DNA itself, therefore, altering the interaction of methylated genes with messengers (such as regulatory proteins) that typically control gene expression.

It has also been suggested that epigenetic regulation plays a role in the development of some cancers and other diseases. Disorders caused by epigenetic mechanisms include Prader-Willi syndrome, which results in short stature and a complete lack of sexual development in adulthood, and Angelman syndrome, symptoms of which include developmental delays, seizures, and frequent laughter and smiling. Both diseases are caused by the loss of expression of a part of chromosome 15, which can be due to loss of a portion of the chromosome or epigenetic imprinting abnormalities eliminating expression of these genes. Beckwith-Wiedemann syndrome, which causes large size, an enlarged tongue, abdominal wall defects, and hypoglycemia, has also has been associated with an epigenetic imprinting abnormality. Epigenetic processes have also been suggested to be involved in various psychiatric disorders, including schizophrenia, bipolar illness, and depression, though these have not been fully characterized.

Research also suggests that certain cancers, including Wilm's tumor associated with Beckwith-Weidemann syndrome, may have an epigenetic origin. By examining DNA samples collected from tumors and normal tissues, researchers have found that the development and proliferation of cancerous tumors may be associated with a loss of imprinting at a number of different loci. As epigenetics is a quickly growing field of research, it is likely that additional epigenetic-associated diseases will be identified in the future.

—Tara C. Smith

See also Cancer; Gene; Genetic Disorders; Genetic Epidemiology

Further Readings

Allis, C. D., Jenuwein, T., & Reinberg, D. (Eds.). (2007). *Epigenetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Jiang, Y.-H., Bressler, J., & Beaudet, A. L. (2004). Epigenetics and human disease. *Annual Review of Genomics and Human Genetics, 5*, 479–510.

ESCHERICHIA COLI

Escherichia coli, more often known as *E. coli*, is a bacterium commonly found in the digestive system of humans and animals. There are hundreds of different strains of *E. coli.* Although they are generally harmless, there are some that cause severe illness. The most widely known harmful strain is *E. coli* O157:H7.

E. coli O157:H7 is a shiga-toxin-producing gram-negative bacterium that causes foodborne illness. It is usually found in the intestine of cattle and produces large quantities of the shiga-like toxin, a verotoxin, which causes severe damage to the lining of the intestine in infected individuals. Ground beef is the most common source of contamination, typically caused by the undercooking of beef that became infected during slaughter. Other sources of contamination include unpasteurized milk and juice, lettuce, sprouts, and water.

E. coli O157:H7 was first identified as a source of illness in 1982 in an outbreak of severe diarrhea among 47 people from Michigan and Oregon caused by hamburgers from a fast-food restaurant contaminated with the bacteria. It became a reportable disease in the United States in 1987. Most cases have been reported in the United States, Canada, and the

United Kingdom. However, the largest recorded outbreak of *E. coli* O157:H7 occurred in Japan in 1996. It affected more than 6,300 children and caused two deaths. This outbreak was associated with consumption of radish sprouts served in elementary school lunches in Japan.

Infection is diagnosed by detecting *E. coli* O157:H7 in a stool sample. Infected individuals typically experience symptoms that include severe abdominal pain and bloody diarrhea. These symptoms can last for approximately 5 to 10 days. In severe cases, kidney failure may result, called hemolytic uremic syndrome. While rare, this most commonly occurs in the elderly and young children, and treatment typically includes blood transfusions and kidney dialysis. Although it can affect individuals of all ages, young children are the most at risk of developing serious complications.

Although transmission is primarily foodborne, infection can be secondarily passed from person to person and has been documented in some day care and nursing home outbreaks. Secondary transmissions are typically the result of improper hand washing and poor hygiene, resulting in the transmission of bacteria from fecal contamination. *E. coli* O157:H7 causes approximately 73,000 illnesses in the United States each year and 60 deaths. However, these are likely underestimates of the true prevalence given the numbers of infected individuals who do not seek medical treatment and, therefore, are not captured in the public health surveillance system. Preventive measures such as proper hand washing and thoroughly cooking meat are suggested as ways to minimize transmission of *E. coli* O157:H7.

—Kate Bassil

See also Foodborne Diseases; Notifiable Disease

Further Readings

Hideshi, M., Kazuhiro, A., Shunsaku, M., Satoshi, T., Nobumichi, S., Motonobu, M., et al. (1999). Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. *American Journal of Epidemiology, 150*(8), 787–796.

Rangel, J., Sparling, P. H., Crowe, C., Griffin, P. M., & Swerdlow, D. L. (2005). Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. *Emerging Infectious Diseases, 11*(4), 603–609.

Ethics in Health Care

Although health care is primarily delivered to individual patients, health care ethics examines issues that relate to many stakeholders, including patients (first and foremost), families, providers, institutions, third-party payers, and society. Ethics in health care can richly overlap with other areas of ethical enquiry such as organizational and business ethics, as well as research and public health ethics. Health care ethics is a young and loosely defined interdisciplinary field with a diverse community of scholars; however, it has been recognized as a field of study within the university, and it has a shared vocabulary, specialized journals, research funding, and an identifiable body of topics of concern. Scholars in the field of health care ethics seek consensus where possible, and when consensus is not possible, the goal is dialogue with mutual respect to advance understanding.

Tasks and Methods

Health care ethics draws on a wide variety of tools from a number of disciplines. The diversity of tasks and methods used in the field can be illustrated by considering the following vignette.

> Mr. Decline is a 48-year-old man whose life is in jeopardy from end-stage kidney disease. In the past, he received dialysis regularly, but he recently discontinued all treatments. His physician has explained that without dialysis he will die very soon. However, Mr. Decline refuses dialysis because he says it is uncomfortable and inconvenient and he generally feels well. He says he will undergo dialysis when he needs it, but right now he feels fine.

Many hospitals in the United States have multidisciplinary ethics committees that provide clinical ethics consultations. The consultative services provided by ethics committees can be used to educate practicing health care professionals, but they are more commonly called on to facilitate the resolution of disputes. Methods for dispute resolution include neutral framing of ethical questions, soliciting additional consultations (e.g., psychiatric or legal consultations), mediation, determining who should make decisions, facilitating discussions, and providing nonbinding advice. In the

case of Mr. Decline, a clinical ethicist would likely review Mr. Decline's chart, interview him and the staff who regularly provide care to him, and request a psychiatric consultation to evaluate his capacity to make medical decisions. If the psychiatric evaluation revealed that Mr. Decline did not have the capacity to make informed decisions, the ethics consultant would likely search for advance directives and/or next of kin for input. If no relatives could be found, the ethics consultant might work with a local court for the appointment of a guardian *ad litem.* On the other hand, if Mr. Decline were found to have decision-making capacity, his wish to decline treatment would be honored. This is grounded in the legal and ethical doctrine of patient self-determination, which recognizes a competent adult patient's right to refuse medical treatments.

Clinical ethicists who work in health care environments frequently play additional roles in cases involving issues such as informed consent and decision-making capacity. For example, they may provide education to staff, may advise on health care policies, or serve as expert witnesses in legal cases that involve related disputes.

Methods of Ethical Inquiry

Health care ethics draws on a wide range of tools from various disciplines.

Philosophical Methods

The discipline of ethics originated in philosophical reflection on the nature of the good and the right. Philosophy typically divides ethics into three categories: normative ethics, descriptive ethics, and metaethics. Normative ethics addresses questions of "ought" and tries to provide answers to the question, "What ought one do?" Descriptive ethics addresses empirical questions such as "How do people in fact behave in a certain situation?" Metaethics is the most esoteric of the three branches. It deals with questions about the meanings of ethical terms and the character of ethical knowledge. Metaethics is the foundation that normative ethics builds on, though it has been observed that greater consensus exists regarding specific moral principles than the metaethical foundations of ethics.

Normative ethics is the heart of ethical analysis, being informed by both metaethics and descriptive ethics. Thus, ethical considerations in philosophy seek to build a normative framework to guide moral decisions. There are myriad philosophical frameworks for guiding the moral life, usually centered on a theory of right action or of good character (virtue). Examples of theories of right action include utilitarianism (briefly stated as the greatest good for the greatest number) and Kantian deontology (acting only according to inviolable universal maxims of right action, such as always treating persons as ends rather than as means). Such simple theories are not commonly advocated today, as they are perceived to be inadequate to many of the complex tasks in health care ethics. Some currently used models are outlined below in the section on resolving ethical disputes.

Additionally, philosophical ethics may engage in clarification of basic concepts or the practical implications of moral rules. In the case of Mr. Decline, the tools of philosophy could be used to address a series of normative and conceptual questions such as "Is a patient morally obliged to accept all treatments that will prolong his or her life, or is it acceptable to forego treatments that are futile or overly burdensome?" When a patient foregoes life-sustaining treatments, is this tantamount to suicide? What is the scope of patient autonomy or of a physician's duty to heal?

Theological Methods

Many theological traditions enrich ethical discourse with a great array of perspectives. There is no one "right" theological method in secular ethics, but secular ethics simultaneously accepts input from many faith traditions. In Western secular ethics, the most common theological voices are those of Christianity and Judaism. Under these models, divine revelation and respect for traditions play central roles. Ethical problems are seen in light of received wisdom, and they are generally solved in the larger context of leading a moral life. The ethical analytic process has much deeper content for members of strongly defined moral communities, and they are generally able to resolve a great number of questions that remain debated in secular ethics. For example, most religious traditions have well-defined norms regarding the provision of health care at the extremes of life (i.e., norms dealing with abortion and euthanasia)—areas in which secular ethical consensus is not likely in the foreseeable future.

For a case such as Mr. Decline's, theology would be likely to ask questions concerning the appropriateness of life-sustaining treatments and the ways in which patients are expected to address illness and suffering.

Legal Methods

Legal scholars in ethics study how law and ethics intersect and inform each other. For instance, if studying informed consent, they would investigate its various manifestations (e.g., as explicit, presumed, or implied), as well as situations in which there might be exemptions such as emergency treatment. Legal ethics scholars are in a position to contribute substantially to policy debates. Some topics that legal ethics frequently discuss include universal access to health care, physician-assisted suicide, pharmaceutical regulation, privacy issues, and pain management. Legal scholarship in ethics helps define boundaries of human behavior through an understanding of case law, constitutional law, legislation, and administrative law. The interaction between law and ethics has been particularly fruitful in the area of medical research ethics. The ethical recommendations in the document known as the Belmont Report heavily influenced the Code of Federal Regulations that currently governs all biomedical research in the United States.

Legal considerations in Mr. Decline's situation would focus on issues of legal competence, advance directives, surrogate decision making, patient self-determination, and the scope of physician duties as defined by constitutional, statutory, and case law as well as standards articulated by the profession.

Qualitative and Quantitative Methods

Qualitative and quantitative methods are used both to describe actual moral judgments and behaviors (descriptive ethics) and to contribute data relevant to normative analysis. Both quantitative and qualitative methods draw from branches of the social sciences such as psychology, sociology, and anthropology. In addition, many quantitative methods also derive from processes found in the life sciences.

Broadly considered, qualitative methods are those methods that do not yield results that can be analyzed statistically. They generally involve asking open-ended questions to small groups of participants. Qualitative research methods seek to discover values and perspectives in populations. Some types of qualitative methods are focus groups, interviews, observation, phenomenology, and ethnography.

Quantitative methods are those that yield data that are best understood by statistical analysis. They ask questions such as "What percentage of patients with Alzheimer's disease retain the capacity to make medical decisions?" Quantitative researchers in ethics frequently make use of surveys and assessment instruments. A great deal of effort must be expended to craft reliable survey or assessment items so that the data gleaned from them will withstand rigorous statistical scrutiny.

Both qualitative and quantitative studies have addressed issues similar to Mr. Decline's. Qualitative studies might explore decision-making capacity, for instance, by interviewing patients to assess how they understand their treatment, prognosis, and options. A quantitative study might apply these findings further by trying to correlate them to objective measurements such as tests of cognitive function.

Topics

Health care ethics addresses a very wide array of issues from many perspectives. Commonly engaged topics include issues surrounding reproduction and the beginning of human life, such as fertility control, prenatal genetic screening, in vitro fertilization, or abortion; end-of-life issues such as futility judgments, code status, organ donation, or palliative and hospice care; the social dimension of health care, such as allocation of resources, managed care, eugenics, or conflicts between patient interests and society's interests; matters pertaining to informed consent, such as decision-making capacity, presumed consent, or children's right to assent to treatment; human research ethics, including the use of placebos in clinical trials, data and safety monitoring, and justice in recruiting participants; and professional responsibilities, such as truth-telling, the protection of confidentiality, and competency. The list of topics addressed within health care ethics will continue to evolve as new technologies develop. While new technologies often solve some ethical problems (e.g., an effective treatment for kidney disease might eliminate a host of ethical issues surrounding kidney donation, allocation, and transplantation), frequently new technologies generate a host of new ethical questions (e.g., our growing ability to diagnose the genetic correlates to diseases for which no cures currently exist poses numerous questions pertaining to cost-effectiveness, beneficence, justice, and eugenics).

Resolving Ethical Disputes

Any problem in health care ethics may include legal, medical, philosophical, religious, and sociological

dimensions. The problems most often identified with ethics are those that involve disputes over competing values, principles, or core commitments. Many patients rely on their faith tradition for answers to bioethical questions, and most traditions provide guidance on a variety of issues. Nevertheless, within the U.S. health care system, health care ethics is most often a secular endeavor embedded in the context of a pluralistic society, and as such it lacks a shared, canonical ethical framework for dispute resolution. However, some proposed frameworks, limited in scope, have been influential in shaping a language of secular ethics discourse.

The following are summaries of some influential theories in secular health care ethics.

Permission-Based Ethics (Engelhardt)

Engelhardt argues that reason has failed to generate a set of universally acceptable principles to guide moral decisions. Therefore, consensus is generally not possible across competing visions of the good because reason cannot mediate outside a particular context. Engelhardt distinguishes two kinds of moral relationships: those between "moral friends" (those who share fundamental convictions) and "moral strangers" (those who embrace nonshared convictions). Because of the failure of reason to establish a consensus on important ethical matters, society must embrace the principle of permission. Many, if not most, morally serious people would have to permit much that they view as unethical for a consensus based on permission to provide an outline of acceptable behavior that interdicts only the grossest of ethical infractions. Because of this provision, Engelhardt sees permission as effectively the lowest common denominator among competing beliefs. This model does not propose itself as a standard to reach for, but rather as the most robust public moral arrangement possible in a postmodern, pluralistic society, if we eschew the use of violence or coercion.

Principlism (Beauchamp and Childress)

Principlism seeks to identify common ground among competing views of morality. It does this in a unique way through the use of four "mid-level principles": autonomy, nonmaleficence, beneficence, and justice. Principlism assumes that competing moral theories do not have a common groundwork and also that they do not necessarily agree on what the proper ends of morality ought to be. Nevertheless, drawing from "common morality," principlism uses principles that emerge across competing theories. These principles may provide a shared language and a conceptual lens for analyzing ethical problems. For example, the case of Mr. Decline might be described as a conflict between patient autonomy (if Mr. Decline is competent to make decisions) and medical beneficence.

Virtue Ethics (Pellegrino and Thomasma)

According to classical virtue theory, virtues are behavioral characteristics that lead human beings to achieve their natural goal of full human flourishing. Pellegrino and Thomasma extrapolated this concept from individuals to the professions, arguing that professional virtues are the characteristics that lead professionals to the attainment of the goals of the profession. For example, if the primary goal of medicine is to heal, then virtue ethics seeks to find what character traits promote that goal. These traits, or professional virtues, should be fostered in professionals because they are essential to achieving the goals of their professions. Virtue ethics does not rely on abstract principles to guide moral actions but instead appeals to society's common frame of reference in the form of virtuous behavior.

Casuistry

Casuistry is a discipline of ethical problem solving that emphasizes looking to past precedent to resolve current disputes. Casuistry has deep roots in Western thought, appearing prominently in the ideas of the sophists against which Plato and other ancient philosophers fought. It has fallen in and out of favor over the centuries and was most recently revived in the 1960s, in part because of bioethics and its need for case analysis. Casuists study myriad facets of cases that have occurred in the past to see their similarities and differences, learn from them, and apply those lessons to the resolution of present cases. There is a strong analogy between casuistry and the legal concept of case law.

Bioethics and the Law

The law can narrow the range of ethical debate on an issue, though some debates continue to rage after the law has addressed an issue (e.g., abortion). Though at

first blush both law and ethics seem to deal with setting boundaries for human behavior, the two have important differences beyond enforcement. The chief separation between law and ethics is how each one sees the purposes of the limits it sets. On at least one account, law sets groundwork and provides punishment for violations, whereas ethics is largely aspirational and seeks to inform good behavior. On the other hand, ethics can help to shape law. Professional standards of practice are one example. Violations of standards of practice are the most common tort causes of action, and ethics contributes significantly to definitions of what constitutes best practice.

—James M. DuBois and Nathaniel J. Brown

See also Ethics in Human Subjects Research; Ethics in Public Health; Qualitative Methods in Epidemiology; Tuskegee Study

Further Readings

American Society for Bioethics and Humanities. (1998). *Core competencies for health care ethics consultation*. Glenview, IL: Author.

Annas, G. J. (1993). *Standard of care: The law of American bioethics*. New York: Oxford University Press.

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics* (5th ed.). New York: Oxford University Press.

Beauchamp, T. L., & Walters, L. (2003). *Contemporary issues in bioethics* (6th ed.). Belmont, CA: Wadsworth-Thompson Learning.

Dubler, N. N., & Liebman, C. B. (2004). *Bioethics mediation: A guide to shaping shared solutions*. New York: United Hospital Fund of New York.

Engelhardt, H. T. (1996). *The foundations of bioethics* (2nd ed.). New York: Oxford University Press.

Jonsen, A. R. (1998). *The birth of bioethics*. New York: Oxford University Press.

Jonsen, A. R., Siegler, M., & Winslade, W. J. (2006). *Clinical ethics: A practical approach to ethical decisions in clinical medicine* (6th ed.). New York: McGraw-Hill.

Moreno, J. D. (1995). *Deciding together: Bioethics and moral consensus*. New York: Oxford University Press.

Pellegrino, E. D., & Faden, A. I. (1999). *Jewish and Catholic bioethics: An ecumenical dialogue*. Washington, DC: Georgetown University Press.

Pellegrino, E. D., & Thomasma, D. C. (1993). *The virtues in medical practice*. New York: Oxford University Press.

Reich, W. T. (1995). *Encyclopedia of bioethics* (Rev. ed.). New York: Macmillan.

Sugarman, J., & Sulmasy, D. P. (2001). *Methods in medical ethics*. Washington, DC: Georgetown University Press.

Ethics in Human Subjects Research

During the latter part of the 20th century, significant attention was paid to abuses of human subjects and to the development of a system for overseeing human research and a framework for the ethical conduct of such research. The Code of Federal Regulations (C.F.R.) governs much of the research conducted in the United States or by U.S. investigators abroad. Nevertheless, compliance with the regulations governing human research does not guarantee the ethical conduct of research. The regulations require ethical insight to be interpreted and applied appropriately, and regulations may not capture ethical issues that emerge in light of social changes or scientific and technological developments. The Office of Human Research Protections (OHRP), part of the U.S. Department of Health and Human Services, periodically offers additional guidance; professional organizations and institutions may produce codes of ethics or guidelines that address the ethical conduct of research on humans; and several international codes and statements influence discussions of research ethics. In the developed world, it is common for nations to have substantive laws, regulations, guidelines, or policies concerning research.

This entry describes the historical development of ethical principles governing research with human subjects, reviews public policy in this area, and considers specific issues of importance in epidemiology. The ethical conduct of epidemiological research involves balancing the interest in conducting such research so as to understand health and disease and ultimately improve health and prevent disease against individuals' interests in privacy and confidentiality and in avoiding nonvoluntary participation in research.

History

The contemporary approach to the ethical conduct of research grew out of a history of abuses and scandals. The abuses of human subjects that surfaced during the trial at Nuremberg of Nazi physicians and scientists who had experimented on concentration camp victims led to the first significant 20th-century changes in human subjects' protections. The necessity of these protections was demonstrated not only because the often fatal experiments conducted by the Nazis were

made known, but also because as part of their defense, the Nazi defendants spoke of studies conducted by the victorious powers that shared some parallels to their experiments, including the failure to obtain consent.

The Nuremberg Code was developed as part of the judgment against the Nazi physicians and scientists. The Code consists of 10 principles that the court said should be followed when using humans in research, including the obligations to (1) obtain subjects' voluntary consent; (2) allow subjects to withdraw and end their participation at any time; (3) ensure that it is necessary to conduct the proposed research to obtain the information sought and that the research is expected to be beneficial to society; (4) minimize the risks to and suffering of subjects; and (5) ensure that the expected benefits outweigh the anticipated risks.

The publication of Henry Beecher's 1966 article in the *New England Journal of Medicine* was a pivotal event in the history of research ethics. Beecher described 22 biomedical studies that he believed were unethical. In the Jewish Chronic Disease Hospital study conducted in 1963, liver cancer cells were injected under the skin of elderly patients to study immune responses. Most of the subjects suffered from dementia, were not told the truth about the study, and were not asked for permission to participate. At the Willowbrook State School for the Retarded in New York City, where children typically developed hepatitis while living in the institution, children were infected with hepatitis on admission to study the progress of the disease and possible methods of treatment. The study raised a range of issues, including the use of undue influence to obtain parental permission by putting children who would participate in the study on a fast track for admissions and the failure to take basic precautions to prevent the spread of hepatitis in the institution.

Around the time of Beecher's publication, the U.S. Public Health Service (USPHS) began requiring peer review, including consideration of subject's rights and welfare, of all USPHS studies. Despite this requirement, failure to consider the subjects' welfare was overwhelming in the infamous USPHS syphilis study conducted in Tuskegee, Alabama, for four decades starting in 1932. This study prompted radical changes in the regulation of human research in the United States and led to the development of the framework used for assessing the ethical conduct of research. The study enrolled African American sharecroppers in Alabama with promises of meals, free medical care, and burial insurance. Subjects were not told they were participating in syphilis research, were not asked for permission to enroll them in the observational study, and were not given information about the fact that they had syphilis. The study remained strictly observational even after penicillin became widely available for treatment of syphilis. Throughout the course of the study, and especially after the introduction of penicillin, subjects were discouraged and prevented from seeking treatment outside the study sites. A 1972 article in the *Washington Post* informed the public of the study, and significant changes in human research began to occur. In 1972, the Office for the Protection from Research Risks (now OHRP and housed under the Department of Health and Human Services) was created under the National Institutes of Health (NIH) to protect the welfare of human research subjects. In 1974, the U.S. Congress passed the National Research Act, requiring that an institutional review board (IRB) review research protocols and appointing the National Commission for the Protection of Research Subjects of Biomedical and Behavioral Research. The Commission was to identify principles and guidelines that ought to govern research on humans. The Commission published a number of reports, the most well known of which is the 1979 Belmont Report, which identified three basic ethical principles for the ethical conduct of research:

- *Respect for Persons.* The judgments of autonomous persons, those "capable of deliberation about personal goals and of acting under the direction of such deliberation," must be respected. Persons not capable of self-determination must be given special protections (§B.1). This principle generates the obligation to obtain valid, free, and voluntary informed consent for research participation.
- *Beneficence.* One must protect persons from harm and make efforts to secure their well-being. This is the basis of the obligation to ensure that the anticipated benefits (to subjects and/or society) outweigh the risks, that risks be minimized, and the subjects be told of a study's risks (§B.2).
- *Justice.* The principle of justice requires that one avoid denying someone a benefit to which he is entitled and that one avoid imposing an undue burden on a person (§B.3). This principle requires that subject recruitment and selection be conducted in such a way that no group disproportionately bears the burdens of research participation and that no population be denied access to research participation without justification.

Public Policy

The C.F.R. and the *Belmont Report*, on which the regulations are largely based, are of primary importance in the U.S. research environment. A number of codes and international guidelines also influence discussions surrounding the ethical conduct of research, especially the Nuremberg Code, the Declaration of Helsinki, and the Council for International Organization of Medical Sciences (CIOMS) (an organization established by the World Health Organization and the United Nations Educational, Scientific, and Cultural Organization [UNESCO] in 1949) Guidelines.

C.F.R. Regulations

The Department of Health and Human Services regulations at 45 C.F.R. 46.102(f) define research as a "systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge." Generally, any investigation that is intended to result in the dissemination of information through an article, a presentation, a poster session, or some other means is considered to be aimed at contributing to generalizable knowledge. A human subject is "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) [d]ata through intervention or interaction with the individual, or (2) [i]dentifiable private information." When both conditions are met, investigations are generally subject to the C.F.R. Three main sections of the C.F.R. concern human subjects research protections: 45 C.F.R. 46 (the Common Rule); and Food and Drug Administration (FDA) regulations at 21 C.F.R. 50 and 21 C.F.R. 56. The Common Rule reflects a 1991 decision by 17 federal agencies and departments to adopt common regulations governing human research; the FDA maintains its own set of very similar regulations.

Although the regulations in the C.F.R. govern a significant portion of research conducted on persons in the United States or by U.S. investigators engaged in research abroad, not all research falls under the C.F.R.'s purview. However, all research sponsored by the federal government, conducted in an institution that has a federalwide assurance (FWA), conducted by any investigator employed by an institution that has an FWA, or any research the results of which are to be submitted to the FDA must be conducted in accordance with the C.F.R. A number of scholars, scientists, and government panels have argued that all human research conducted in the United States or by U.S. investigators ought to be governed by the C.F.R.

Institutional Review Boards

A key requirement in the C.F.R. is that research protocols be reviewed and approved by a duly constituted IRB and that IRBs monitor studies on an ongoing basis and review approved protocols at least once per year. Traditionally, IRBs have been housed within the institutions where proposed research will be conducted, such as academic medical centers or hospitals. Today, some institutions contract the review of at least some studies to independent or commercial IRBs. An IRG must have at least five members, at least one of whom must be a scientist and one of whom must not have a background in science. At least one member must not be affiliated with the institution conducting the research. If a proposed study will involve prisoners, at least one prisoner or one person who can represent the interests of prisoners must be on the board, and the majority of the IRB members must have no affiliation with the prison(s) involved in the study. The C.F.R. exempts some research from the requirement of IRB review and allows some research to be reviewed by an IRB in an expedited fashion rather than by the full board. For example, studies using only existing data that will be recorded without identifiers or codes may be exempt from IRB review. Many institutions require that investigators conducting exempt research inform the IRB of their intent to do so and that a representative of the IRB determine that the proposed study meets the criteria for exempt research. Such review can ensure that investigators have appropriate plans to conduct the research such that it in fact qualifies as exempt, for example, that they have plans to anonymize data and protect subjects' confidentiality. Studies that pose only minimal risk to subjects and meet one of nine criteria enumerated in the *Federal Register* of November 9, 1998, may be reviewed on an expedited basis. Examples of research that generally qualify for expedited review include certain studies that involve only venipuncture where limited amounts of blood will be drawn from adults and studies that involve only a survey instrument or other noninvasive means of collecting data.

IRBs review protocols to ensure that they comply with the requirements set forth in the C.F.R. and that any other ethical or safety concerns are given proper

consideration. IRBs are responsible for ensuring that risks to subjects are minimized; the expected benefits of a study outweigh the risks; research participants will be recruited and selected in a fair manner such that no group disproportionately bears the burdens of research participation or is denied access to research participation without justification; investigators plan to obtain and document informed consent appropriately; data safety monitory plans appropriate to a protocol have been developed; and appropriate measures are taken to protect subjects' confidentiality and privacy. The requirements for informed consent vary based on the type of study proposed and the population to be included. If a study will recruit adults who may be unable to make their own decisions regarding research participation because they are unable to understand and appreciate information, appropriate plans to involve a legal surrogate must be made. If a study will involve children, the permission of one or both parents is required, depending on the nature of the study and the reasonable availability of both parents. As appropriate for a given study, the information to be provided as part of the informed consent includes information about the fact that the proposed intervention is for research; the purpose of the research; the number of subjects to be enrolled; the time research participation will require; the research procedures; the foreseeable risks and discomforts and, if appropriate, acknowledgment that that there may be unforeseen risks; the expected benefits to subjects and/or others; subjects' alternatives to participation; confidentiality; the availability of compensation or treatment available to subjects who may be injured; contact information for persons who may help subjects with questions about their rights and who should be contacted in the event of injury; the voluntary nature of participation; the right to withdraw from a study and the consequences of withdrawing early; the circumstances under which a subject may be withdrawn from a study without the subject's permission; additional costs subjects may incur as a result of participation; and the fact that new information obtained during the course of a study will be communicated if it is deemed relevant to subjects' desire to remain enrolled in a study. Particularly, when patients enroll in research, it is incumbent on the researcher to counteract subjects' tendency to misunderstand the purpose of research and assume that research is an extension of their clinical care or that participation in research will give them access to the newest and best treatments (therapeutic misconception), compromising informed consent.

Other Ethical Issues

In recent years, a number of concerns have emerged that are not explicitly addressed in the C.F.R. but that have become important in discussions of the ethical conduct of research. One is that investigators and institutions conducting research may have conflicts of interest or commitments that could interfere with their obligations to patients and/or subjects. Financial conflicts of interest have been scrutinized in the wake of cases in which subjects were injured or died, and it was later revealed that the investigators and/or research institutions had significant financial interest in the research.

Even in the absence of significant financial conflicts of interest, investigators who also serve as clinicians, and organizations that also serve as patient care institutions, may have conflicts between their commitment to and duties of care and their research aims. Following the Kennedy Krieger–led abatement study ruling in Maryland, in which the court asserted that investigators owed a duty of care to the subjects enrolled in the study, there has been much discussion of what duties researchers may have to patients and nonpatients enrolled in research. The conduct of research in developing nations, either with the aim of ameliorating disorders or conditions endemic to those nations or for the purpose of testing interventions that primarily will be used in the developed world, has raised important new questions. For example, after the trials on preventing vertical transmission of HIV in the 1990s, during which a shorter course of AZT was tested against placebo in pregnant women (rather than against Protocol 076, which had become the standard of care in the developed world and according to which pregnant women were given AZT orally five times a day during the last several months of pregnancy and intravenously during labor; newborns were given AZT in syrup form for the first 6 weeks and were not breastfed), a debate emerged about what it means to provide standard of care (e.g., worldwide global best of care or local standard) to subjects and what obligation researchers have to provide standard of care. Finally, there is ongoing discussion of what is owed to research subjects who are injured as a result of their participation and whether sponsors (or others) must make provisions for compensating individuals' research-related injuries.

Epidemiological Research

The regulations governing human research emerged in response to studies that exposed subjects to high levels of physical risk, yet this framework has been applied to all areas of research involving human subjects, including epidemiological research and behavioral and social sciences research. Research in these areas raises some unique ethical concerns, and it may be appropriate to develop a more nuanced regulatory framework for oversight of human research in response to the concerns that emerge in different areas of research. Epidemiological research can raise questions about privacy, confidentiality, informed consent, study design, cultural sensitivity and imperialism, reporting of results, the blurring of the line between advocacy and scientific research, and the possibility of conflicts of interest among funders who expect certain results.

The ethical issues that arise in epidemiological research can be examined by distinguishing three categories of research: studies that (1) use only existing data, (2) involve collecting new data, or (3) are interventional or experimental in nature.

Use of Existing Data

Privacy, confidentiality, and lack of consent are the primary ethical concern raised by studies that involve only the use of existing data, records, specimens, or other information. In many cases, requiring consent of individual subjects would render it impossible to conduct epidemiological research and important information would remain unavailable. If identifying information is not collected with the private information that is gathered from existing sources, many of the confidentiality concerns are resolved and such research is considered exempt under the C.F.R., and informed consent is waived. In many institutions, IRBs make the final determination that such studies are exempt from full board review, and IRBs may examine protocols to ensure that the confidentiality and privacy protections proposed are sufficient. A person's privacy, nevertheless, may be invaded by such research because investigators may initially have access to identifying information even if they record data without identifiers. An emerging area of concern is the use of research involving tissue samples previously collected for research or as part of clinically indicated procedures. The OHRP offers guidance on when research may be conducted using stored specimens, when such research may be considered exempt, when the persons from whom the specimens were obtained must be contacted for consent to conduct the research, and what procedures should be followed when collecting specimens to be stored for possible future research. Because the primary concerns associated with such research involve privacy and confidentiality, research involving specimens that have been de-identified or that cannot be linked to individuals raises fewer concerns than research involving identifiable samples.

Collecting New Data

When epidemiological research involves the collection of new data, consent generally must be obtained, but it often is not documented through a formal consent form. If the new data will be collected from surveys or comparable means, typically investigators must disclose the fact that a person is being asked to participate in research and offer them information about the study. Participants indicate their consent by responding to interview or survey questions. If the data are collected only through surveys or comparable means without identifiers and the research does not involve minors, the research typically is exempt. Individual institutions may require some oversight of such research. Studies involving minors require IRB review and approval, and there is disagreement about how parental permission should be obtained. Some favor approaches that require parents to opt out rather than to opt in as a means of increasing participation; others are concerned that at least some research should require a clear indication that parents have chosen to allow their children to participate.

Interventional or Experimental Studies

When communities are randomized to receive different interventions or to serve as controls receiving no interventions to alter behaviors or practices that affect health as part of interventional or experimental epidemiological studies, additional ethical concerns emerge. Examples of these studies include research aimed at reducing smoking in a community, increasing childhood vaccination rates, or improving cancer-screening rates. Questions about how communities are chosen and randomized and how to show proper respect for cultural practices that may be of interest to epidemiologists emerge. A number of questions revolve around informed consent, including whether there is an obligation to obtain permission from community members to have their community be subject to interventions and

have outcomes measured and compared; if an obligation to obtain community permission exists, how such permission should be obtained and how much information must be disclosed (especially in light of the possibility that disclosing too much information could invalidate the results), and whether all communities or only those subject to experimental interventions must give their permission to participate are questions that need to be answered.

—Ana S. Iltis

See also Ethics in Health Care; Ethics in Public Health; Informed Consent; Institutional Review Board; Tuskegee Experiment

Further Readings

Advisory Committee on Human Radiation Experiments, Department of Energy. (1995). *Advisory Committee on Human Radiation Experiments Final Report.* Washington, DC: Government Printing Office.

Brody, B. (1998). *The ethics of biomedical research: An international perspective*. New York: Oxford University Press.

Council for International Organization of Medical Sciences. (1991). *International Guidelines for Ethical Review for Epidemiological Studies.* Retrieved March 8, 2006, from http://www.cioms.ch/frame_1991_texts_of_guidelines.htm.

Couglin, S. S., & Beauchamp, T. L. (1996). *Ethics and epidemiology*. New York: Oxford University Press.

Emanuel, E. J.,, Crouch, R. A., Arras, J. D., Moreno, J. D., & Grady, C. (2003). *Ethical and regulatory aspects of clinical research*. Baltimore: Johns Hopkins University Press.

Jones, J. H. (1981). *Bad blood: The Tuskegee syphilis experiment*. New York: Free Press.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report*. Washington, DC: Government Printing Office.

Office for Human Research Protections. (2004). *Guidance on research involving coded private information or biological specimens*. Retrieved March 6, 2006, from http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.pdf.

Ethics in Public Health

Public health practice and research raise unique ethical dilemmas. In contrast to medicine, which is patient-centered, public health activities aim to protect and promote population health and must balance risks and benefits to individuals against those to communities or society as a whole. Students and practitioners of public health should recognize the ethical basis of public health activities and make decisions that are consistent with the underlying ethical values and norms of the field. Because key public health threats change over time, ethical values evolve, and multiple disciplines inform the efforts of public health activities, defining a distinctive ethical orientation or normative framework for public health is challenging. This entry aims to reflect areas of consensus that have emerged in recent literature.

Ethical Analysis in Public Health

Public Policy, Law, and Ethical Theory

There is a societal expectation that the government should assume some degree of responsibility for protecting public health. In the United States, public health institutions have legal (police) powers. Legislative statutes, court cases, and administrative policies frequently address public health matters. Ethics, public policy, and law are overlapping sources of guidance for determining acceptable behavior for citizens, professionals, and government officials, allocating limited resources, and defining individual rights and responsibilities. Policies may be inconsistent or outmoded, they may not always reflect society's moral consensus, and they do not provide guidance for decision making in every situation. Some laws may even be considered by certain citizens to be unethical. For these reasons, careful ethical analysis is necessary to provide additional legitimacy for state-sponsored public health activities.

Ethical theories and frameworks are conceptual tools to aid decision making and the evaluation of ethical arguments. They do not provide automatic solutions to complex problems. Theories identify and justify basic moral principles, but in a given situation, facts must be considered, and principles or values can sometimes conflict.

The leading ethical framework in medicine is principlism. The four prima facie principles of respect for autonomy, beneficence, nonmaleficence, and justice are widely recognized as shared professional values from which ethical rights and duties can be derived. Public health shares a commitment to these ethical principles but may weigh them differently in a particular circumstance due to an emphasis on collective

risks and benefits. Additional principles such as utility and consideration of the common good are also characteristic of public health.

Several ethical theories are also relevant to public health. Kantian deontology connects morality and reason; certain actions are intrinsically right or wrong and, therefore, ethically mandated or prohibited. Kant emphasizes individual rights and is perhaps best known for arguing that no person should ever be treated as a means to another's end, but instead, all persons should be treated as "ends in themselves." This emphasis on respect for persons provides balance against considerations of utility and efficiency. In contrast to deontology, for which consequences are relevant only if the action is morally permissible, utilitarianism presupposes that actions are right insofar as they promote the greatest good for the greatest number. For the utilitarian, expected consequences are the primary consideration in decision making. Public health is often described as utilitarian because policies and interventions aim to maximize aggregate health outcomes. However, both Kantian deontology and utilitarianism are somewhat limited in their application to public health because they do not consider collective social goals. Communitarian theory, which asserts that ethics cannot be separated from the shared history, traditions, customs, and institutions of particular communities, may be more consistent with the ethical perspective of public health. Human rights theory may also provide a normative ethical framework for public health, emphasizing universal civil, political, and social rights and recognizing socioeconomic, cultural, and environmental influences on health.

Professional Codes of Ethics

Professional codes of ethics can assist public health practitioners and researchers to recognize and address ethical issues. While professional codes are general and do not provide definitive answers to specific problems, they do outline the basic rights and obligations of citizens, public health practitioners, governments, and other public and private institutions and are therefore useful in setting standards and expectations.

Morality and Health

Public health issues invoke images of virtue and vice. Moral judgments and stereotypes influence public health interventions. Negative perceptions of individuals who engage in risky behaviors affect public attitudes toward policy, funding, and education, including the rhetoric used in interventions and political decisions to criminalize certain behaviors. Prescriptive messages regarding health-related behavior may be judgmental or stigmatizing.

Ethical Principles of Public Health

The fundamental ethical principles underlying public health practice and research include the following: beneficence, respect for the common good, respect for persons, utility, social justice, prevention, commitment to science, honest communication, and community participation. These principles are not weighted in any particular order but rather are all considered prima facie obligations.

- *Beneficence.* One of the basic ethical mandates of public health is to provide benefits. In addition to protecting the population from immediate and future threats to health and safety, public health interventions and research should also avoid causing direct harm and make every reasonable attempt to minimize risks.

- *Respect for the Common Good.* An individual's actions affect not only his or her own health but also the health of other humans, nonhuman organisms, and ecosystems. Most definitions of public health emphasize interdependence and collective action, obligation, and benefit. Focus on the common good (specifically population-level health outcomes) is the primary difference between the ethics of public health and medical ethics. Public health efforts often require pooled resources and cooperation. For example, vaccination can effectively prevent the spread of infectious disease (provide herd immunity) only if a majority of citizens participate.

- *Respect for Persons.* Public health activities should respect individual rights, but autonomy must be balanced with promotion of the common good. In some cases, it may be necessary to override individual rights in order to protect community health. For example, in the case of infectious disease, the population-level benefits of quarantining exposed persons may justify a temporary infringement on personal freedom. However, public health officials must be certain that proportionality is respected and that community benefits outweigh individual sacrifices.

- *Utility.* While not always strictly utilitarian in the classic sense, public health strives to achieve the greatest good for the greatest number and emphasizes producing the maximal balance of benefits over harms and other costs. Efficiency is related to utility; stewardship of limited resources is required. However, the costs and savings of prevention are not always clear, and measurement techniques and metrics vary. Furthermore, efficiency must be balanced with equity.

- *Social Justice.* Public health's social justice perspective supports equitable access to—and equitable distribution of—health care and other resources necessary for health, as well as the benefits and burdens of public health interventions and research. Social justice also provides the foundation for public health's focus on eliminating health disparities, respect for cultural diversity, and protecting the health of vulnerable populations (e.g., children).

- *Prevention.* The public health perspective prioritizes prevention over cure. This focus influences resource allocation decisions. Generally, efforts to prevent disease are prioritized over treatment of those who are already ill. Prevention also mandates thoughtful anticipation and sound scientific research.

- *Commitment to Science.* Public health interventions should be based on scientific knowledge. Public health practitioners have an ethical responsibility to seek out necessary information to the extent possible. However, knowledge is not morally neutral. Facts can be interpreted differently by various stakeholders, and the relationship between facts and values is complex. While knowledge may often demand action, knowledge is not always necessary for action. In some cases, such as an infectious disease outbreak, action may be necessary before all the facts are available.

- *Honest Communication.* Public health activities necessarily involve a great deal of communication. Truthfulness, promise keeping, and transparency about motives and actions promote public trust. Timeliness and accountability are also imperative, particularly in emergencies.

- *Community Participation.* Public health values the participation and collaboration of all stakeholders in planning, implementation, evaluation, and research. Opportunities for community input are particularly important to ensure appropriateness of resource allocation, quality of intervention design, and adequate informed consent for research.

Ethical Challenges for Public Health

Ethical dilemmas familiar to public health researchers and practitioners will be discussed, highlighting those ethical norms that should guide decision making. Examples presented are not meant to be exhaustive but rather illustrative of the types of ethical conflicts arising in public health.

Health Promotion and Disease Prevention

The appropriate parameters for health education and health promotion efforts—and the degree to which governments should regulate health-related behavior—are contentious. Some would argue that the state's only obligation is to protect citizens from imminent threats that necessitate action at the population level (e.g., communicable diseases). Others argue that government paternalism is justified on the grounds of efficiency, utility, and economics as well as the desire of many citizens for assistance in being healthy. However, educational campaigns run the risk of being tyrannical or moralistic and raise ethical questions regarding the appropriateness of moral exhortation, value-laden language and imagery, and excessive focus on individual responsibility. The use of financial incentives or coercive measures (such as the threat of imprisonment) to influence health-related behavior may also infringe on individual rights, stretch governmental limits of power, or violate social justice.

Determining and Valuing Risk

Defining, determining, and communicating risk raises ethical challenges. Determining safe or socially acceptable levels of risk (whether related to behavior, exposure, or interventions) involves value judgment; consideration of various social, political, and ethical factors; and compromise.

The equitable distribution of risk across a population is a matter of social justice. Recent studies suggest that the potential (known and unknown) risks of exposure to industrial pollution and other environmental hazards disproportionately burden low-income populations throughout the world. This inequality exacerbates the health problems of already disadvantaged groups.

Another ethical tension in public health exists between interventions that focus on risk reduction versus those that focus on total risk elimination (or "abstinence" in the case of behavioral interventions).

For example, while some would argue that the only way to eliminate HIV infection among injection drug users is to completely stop (or never start) use, others promote clean-needle-exchange programs as a more realistic and immediate way of reducing transmission. Ethical disagreement also arises over the messages risk-reduction efforts send to youth regarding the safety and social acceptability of unhealthy behaviors.

Public Health Research

The Belmont principles for ethical research with human subjects—respect for persons, beneficence, and justice—were developed with clinical research in mind. They are somewhat inconsistent with public health's emphases on prevention and population health and do not consider community risks and benefits. For example, if research suggests that a certain ethnic group or geographic community is at increased risk of a particular illness or risk behavior, publicizing findings may result in harm (e.g., employment or insurance discrimination, social stigma) not only to study participants but also to any member of the group.

Equipoise, or genuine uncertainty about the comparative merits of two or more interventions, is a condition of ethical research and ensures that subjects are not disadvantaged by research participation. Public health's commitment to eliminating health disparities necessitates consideration of the costs of different types of interventions, and these economic considerations present a challenge to equipoise. A particularly controversial issue in public health (especially international) research is the ethicality of studying or implementing less expensive—and predictably less effective—interventions when a known effective intervention is too expensive for a particular community or country.

Epidemiological Surveillance

Epidemiological surveillance involves the gathering and application of information to determine appropriate interventions and distribute resources for disease prevention. Surveillance is ethically imperative as detailed, accurate information is needed to plan interventions or determine if further investigation is warranted (e.g., in the case of cancer clusters). Surveillance may involve identifying specific individuals or communities to determine the source(s) of the health problem. Like research, surveillance presents risks related to the breach of privacy and confidentiality. However, research guidelines focus on protecting individuals and producing generalizable knowledge and, therefore, cannot be directly translated to the public health setting, where the protection of communal welfare is prioritized.

Infectious Disease Control

Until the emergence of HIV/AIDS in the 1980s, communicable diseases—including sexually transmitted diseases—were handled in a fairly consistent manner. Infected individuals were reported by physicians to local public health authorities, and contact tracing was conducted to identify other exposed and possibly infected individuals. Stigmatization, the labeling of HIV/AIDS as a disease of homosexuals, and the rise of individualism and privacy expectations led to what some have called "HIV exceptionalism." Recently, rising infection rates worldwide have led to the consideration of widespread screening of low-risk populations rather than a turn to rigorous reporting and partner notification practices. However, routine screening does not eliminate ethical concerns related to breach of confidentiality and other risks.

Emergency Response

Emergency response during a disease outbreak or natural or man-made disaster (such as a bioterrorist attack) may involve quarantine, isolation, forced treatment, or other threats to civil liberties. Emergencies may also necessitate breaches of privacy and/or confidentiality such as calling names, providing treatment in open spaces, or publicly disclosing disease or exposure status. Limited health care resources may need to be quickly rationed, and not everyone will receive what they perceive to be adequate and timely treatment. Public discontent may be exacerbated because there is limited information, fragmented communication, and inadequate opportunities for community input.

Emerging Threats

As both an area of moral inquiry and a set of shared professional values, public health ethics continues to develop. Emerging threats to public health, including pandemic infectious diseases, growing health disparities, and the rising costs of health care, will shape

future discourse on the ethical norms and standards of the field.

—*Emily E. Anderson*

See also Ethics in Health Care; Ethics in Human Subjects Research; Governmental Role in Public Health; Informed Consent; Institutional Review Board

Further Readings

American College of Epidemiology. (2000). *Ethics guidelines*. Retrieved September 22, 2006, from http://www.acepidemiology2.org/policystmts/EthicsGuide.asp.

American Public Health Association. (2002). *Principles of the ethical practice of public health.* Retrieved July 3, 2007, from http://www.apha.org/NR/rdonlyres/1CED3CEA-287E-4185-9CBD-BD405FC60856/0/ethicsbrochure.pdf

Anand, S., Peter, F., & Sen, A. (2004). *Public health, ethics, and equity*. Oxford, UK: Oxford University Press.

Association of Schools of Public Health. (2003). *Public health ethics: Model curriculum*. Retrieved September 22, 2006, from http://www.asph.org/document.cfm?page=782.

Bayer, R., & Fairchild, A. L. (2004). The genesis of public health ethics. *Bioethics*, *18*(6), 473–492.

Beauchamp, D. E., & Steinbock, B. (Eds.). (1999). *New ethics for the public's health*. New York: Oxford University Press.

Boylan, M. (2004). *Public health policy and ethics*. Dordrecht, The Netherlands: Kluwer Academic.

Callahan, D., & Jennings, B. (2002). Ethics and public health: Forging a strong relationship. *American Journal of Public Health, 92*(2), 169–176.

Childress, J. F., Faden, R. R., Gaare, R. C., Gostin, L. O., Kahn, J., Bonnie, R. J., et al. (2002). Public health ethics: Mapping the terrain. *Journal of Law, Medicine, and Ethics*, *30*(2), 170–178.

Coughlin, S. S., & Beauchamp, T. L. (Eds.). (1996). *Ethics and epidemiology*. New York: Oxford University Press.

Coughlin, S. S., Soskolne, C. L., & Goodman, K. W. (1997). *Case studies in public health ethics*. Washington, DC: American Public Health Association.

Kass, N. E. (2004). Public health ethics: From foundations and frameworks to justice and global public health. *Journal of Law, Medicine, and Ethics*, *32*(2), 232–242.

Etiology of Disease

The etiology of disease refers to the causes or to the study of the causes or origins of disease. The term *etiology* is derived from the Greek *aitiologia* meaning "statement of cause." The roots come from *aitia* "cause" + *logia* "speaking." The primary focus of investigations of disease causation within the discipline of epidemiology has changed over time. However, the elucidation of the relationships between potential causal factors and diseases or health outcomes has remained a central concern in epidemiological research.

Historical Background

Investigations of disease etiology in the modern period of epidemiology have continued to evolve with the dominant paradigms of disease causation of each era. The sanitary era of the early 19th century was dominated by the miasma theory, a view of disease as arising from foul emanations from impure water, soil, and air. English sanitary reformers who held to this theory, such as Edwin Chadwick and Florence Nightingale, were thus mistaken in their understanding of the specific agents of disease etiology; however, their overall emphasis on improving social and physical conditions in the urban industrial environment led to major improvements in public health.

The work of Louis Pasteur and the development of Koch's postulates next contributed to the shift in clinical medicine and public health toward the germ theory, which attributes disease causation to microorganisms. During the ensuing era of infectious disease of the late 19th and early 20th centuries, etiologic research accordingly consisted of searching for single causative agents of diseases. The classic model that reflects the influence of this era of infectious diseases is the epidemiologic triad of agent, host, and environment.

By the mid-20th century, a shift had occurred in more developed countries, with chronic rather than infectious diseases becoming predominant. Etiologic research in the era of chronic disease epidemiology focused on multiple proximate risk factors for chronic diseases in individuals and benefited from advances in epidemiological and biostatistical methods.

The latter part of the 20th century witnessed increasing developments within epidemiology and related fields that have contributed depth and breadth to the field on both a micro- and macrolevel. Advances in the subdisciplines of molecular epidemiology and human genome epidemiology enabled epidemiologists to use findings from the Human Genome Project and to study the specific pathways, molecules, and genes, and interactions between genes and the environment that

influence the risk of developing disease. Concomitantly, a renewed focus on the effects of the social and physical environments on health and an increasing interest in intergenerational influences were reflected in the development of a life course approach to chronic disease epidemiology and the contributions of social epidemiology to a multilevel perspective of disease etiology.

—Helen L. Kwon

See also Causation and Causal Inference; Epidemiology, History of; Human Genome Project; Koch's Postulates; Life Course Approach; Molecular Epidemiology; Nightingale, Florence; Pasteur, Louis; Social Epidemiology

Further Readings

Schwartz, S., Susser, E., & Sussser, M. (1999). A future for epidemiology? *Annual Review of Public Health*, *20*, 15–33.

Susser, M., & Susser, E. (1996). Choosing a future for epidemiology: I. Eras and paradigms. *American Journal of Public Health, 86*(5), 668–673.

Eugenics

Eugenics, defined as the study of improving the human race through selective breeding, coincided with the Bacteriological Era of epidemiology during the late 19th century into the early decades of the 20th century. As a science, eugenics developed gradually from basic ideas to lofty, and often unfounded, scientific conclusions. Its rapid growth in popularity with the general public and a handful of socially inclined scientists led to a series of events that has permanently influenced the design of modern population studies and the practice and study of medicine.

Nineteenth-Century Origins of Eugenics

Eugenics began when Sir Francis Galton, cousin of Charles Darwin, invented and defined the term *eugenics* in 1883. Many scientists, philosophers, and socialites, including Galton, Herbert Spencer, Charles Davenport, and John Harvey Kellogg, played pivotal roles in legitimizing eugenics and promoting the concept of protecting the human race from genetic and moral decline. England, the United States, and Germany were among the forerunners in the establishment of eugenics, which eventually expanded to more than 30 countries worldwide. People feared diseases and social problems that were thought to be hereditary, including morally inept behavior, congenital abnormalities, and alcoholism, which propelled the eugenics movement forward. Techniques such as sterilization by vasectomy or tubal ligation, forced imprisonment in asylums, and euthanasia were seen by some as ways to cleanse human society.

Scientific advances by three leading figures in the late 1800s led to the initial concepts that would become the cornerstone of eugenics at the turn of the century. Darwin's controversial book *On the Origin of Species*, published in 1859, proposed that inheritance was a natural, controlled process. Darwin's theory of natural selection was based on scientific evidence showing that variation of targeted characteristics in animals influenced selection and survival within a population. While Darwin remained dedicated to studying populations of animals, Galton was increasingly interested in applying selection processes to human populations. He believed that cognitive ability could be increased through improving human biological variation and viewed such improvement as an extension of Darwin's theory. At the time, Darwin thought only on the level of observable traits in a population and did not appear to have considered the complexity of genetics as the driving force behind selection. The only scientist to have recognized the intricacies of inheritance, by studying the transmission of characteristics in peas, was Gregor Mendel. His article on experiments with plant hybridization was published just 6 years after Darwin's book; however, it was poorly understood and disregarded. It wasn't until Mendel's work was rediscovered in the early 1900s that its impact was realized.

During the time between 1860 and 1900, Galton promoted the concept of natural selection in humans. At the same time, the two most significant discoveries in epidemiology during the Bacteriological Era were made. First was the transmission of disease by arthropods, generally from an animal source, and second, the transmission of disease from healthy carriers. However, it was the idea that an unsanitary physical environment contributed to disease, a hypothesis presented centuries earlier, that found a new appreciation in Victorian-Era Britain, as well as in other countries around the world. Since

the British ruling classes were viewed as superior and free from unsanitary environments, many people accepted the perspective that biological inheritance determined leadership qualities and social status. Herbert Spencer, an Englishman, championed Galton's ideas and introduced Social Darwinism, which suggested that a proper and discriminating society depended on the quality of the individuals within the society. Spencer's book *First Principles*, published in 1862, attracted a large following. The book drew comparisons between society and biological organisms and implied that the process of selection was entirely natural, even in humans. Spencer relied heavily on the concept of evolution and considered competition as the fundamental element to evolutionary progress.

The rediscovery of Mendel's laws at the turn of the century was perhaps the single most important scientific detail that finally united early eugenicists with scientists in a wide variety of fields. Mendel's laws provided scientists with the ability to predict the transmission pattern of traits and allowed them to understand the basic mechanism driving inheritance. This understanding provided a way to relate eugenics, concerned with purifying human society, to epidemiology, concerned with the occurrence of disease in large populations. At the time, it was believed that diseases such as typhoid, yellow fever, and syphilis could be inherited, and the decision whether to treat and save people afflicted with these diseases was a constant conflict for eugenicists and public health officers.

Eugenics in the 20th Century

In 1904, one of Galton's former students, Charles Benedict Davenport, established the Station for Experimental Evolution at Cold Spring Harbor in Long Island, New York. Research at Cold Spring Harbor was to be concerned with nonhuman genetic and evolutionary studies. However, Davenport was clearly studying the role of eugenics in human evolution, and he used human pedigree analysis and Mendelian principles to predict tendencies toward traits such as feeble-mindedness or disease. In 1906, John Harvey Kellogg created the Race Betterment Foundation in Battle Creek, Michigan, and just 4 years later, in 1910, Davenport established the Eugenics Record Office (ERO) at Cold Spring Harbor, appointing himself as director and hiring Harry H. Laughlin as superintendent. Laughlin was responsible for collecting data on the inheritance of human traits, analyzing census data and population trends, encouraging courses in eugenic principles, and distributing eugenics materials in the United States and Europe.

Between 1910 and 1916, discussion of eugenics spread rapidly in the mass media. Articles published in periodicals written for the average citizen, such as *Good Housekeeping* and *The Saturday Evening Post*, captured the imaginations of housewives and working husbands. In addition, books targeted for public consumption included such works as *The Passing of the Great Race* by Madison Grant (1916) and *The Rising Tide of Color Against White World Supremacy* by Lothrop Stoddard (1920). A prosterilization film, *The Black Stork* (1917), and a Public Health Service film, *The Science of Life* (1922), were released to demonstrate the dangers of passing on defective traits. In 1929, philosopher and mathematician Bertrand Russell, who received the Nobel Prize for Literature in 1950, published *Marriage and Morals*, in which he wrote, ''In extreme cases there can be little doubt of the superiority of one race to another'' (p. 266). Later, Russell turned against eugenics, becoming a leader in the fight against racism. ''Better Baby'' and ''Fitter Family'' contests were held at state fairs as ways to entertain and educate American families, providing a visible distinction between those who were genetically fit and encouraged to reproduce and those who were not. In the 1930s, eugenics was commonly taught in high schools and colleges, and sermons augmenting the virtues of eugenics resulted in its widespread acceptance.

By 1907, the role of politics in eugenics began to appear when Indiana became the first state to pass an involuntary sterilization law, targeting individuals in mental institutions, sex offenders, and the feeble-minded. By 1935, 26 states had similar laws, and others were preparing to pass legislation. When the repeal of eugenics laws finally became a reality in 1979, more than 64,000 individuals had been involuntarily sterilized. A second major area of eugenics political action involved immigration. The Johnson Immigration Act of 1924 (the Immigration Restriction Act) was the most widespread and restrictive immigration legislation to be passed in the United States. Evidence presented to Congress by Laughlin depicted immigrants from nations in Eastern Europe and other countries as undesirable

because they possessed lower intelligence and high rates of negative attributes, including alcoholism and intractableness, which were then thought to be heritable conditions. However, members of the scientific community were not as easily convinced as Congress, and opposition to the eugenics movement increased. In the early 1920s, it was already understood by geneticists that some traits emphasized by eugenicists could not be attributed to a single gene and were the result of complex interactions among multiple genes. Environmental factors were shown to play a large role in gene expression, and, more important, it was found to be mathematically impossible to eliminate a recessive gene in a population; even if involuntary sterilization were to be used, random mutation will occur.

Loss of support for eugenics began in earnest with the occurrence of the Great Depression. When people in the higher echelons of society suddenly found themselves with no money and no job, they realized the false impression that eugenics had made. Their genetic backgrounds did not determine their superiority over those in the lower classes and heredity had little, if anything, to do with economics. Additionally, in 1933 when Hitler and the Nazi party came to power, the similarities between eugenics in America and the eugenic ideals of Nazi Germany became painfully obvious.

Perhaps the most significant impact of eugenics on epidemiology occurred in 1932 when the U.S. Public Health Service (PHS) initiated the Study of Untreated Syphilis in the Male Negro, more commonly known as the Tuskegee Syphilis Experiment. While the study is often criticized for its length (it lasted until 1972, continuing even after the discovery of the use of penicillin for the treatment of syphilis) and poor scientific reasoning, it is an example of a research program in which it is now apparent that eugenics beliefs infiltrated the science. Three PHS officers, Hugh Smith Cumming, Taliaferro Clark, and Raymond Vonderlehr, founded the study, which was conducted in Macon County, Alabama, where more than 400 black males were screened for syphilis and then not treated so that investigators could learn more about the effects of syphilis. Misleading procedures, lack of treatment, and deceptive tendencies flawed the study. However, the largest flaw was that the study was plagued with the presence of race-conscious scientific assumptions and cultural intolerance that influenced its motives and faulted the study from its very inception.

While Davenport and other eugenicists believed that their scientific methods and hypotheses were valid, the eugenics movement serves as an example of the consequences of confusing scientific findings with societal values. Students and doctors who were educated in the principles of eugenics in the 1920s and 1930s continued to believe in and practice eugenics into the second half of the 20th century despite the failure of eugenics as a science and philosophy. Racism was enforced by eugenics, and even with the substantial progress made to dissipate racial tension over the second half of the 20th century, racial issues in science will remain sensitive. The notion that one race can be made better than another through genetic manipulation is incorrect, and it will take constant educating and surveillance to avoid a similar conflict in the future due to the significant progress being made in the field of genetics today.

—Kara E. Rogers and F. John Meaney

See also Ethics in Health Care; Ethics in Human Subjects Research; Heritability; Social Hierarchy and Health; Tuskegee Study

Further Readings

Allen, G. E. (1997). The social and economic origins of genetic determinism: A case history of the American Eugenics Movement, 1900–1940, and its lessons for today. *Genetica, 99*, 77–88.

Kevles, D. J. (1995). *In the name of eugenics.* Cambridge, MA: Harvard University Press.

Lombardo, P. A., & Dorr, G. M. (2006). Eugenics, medical education, and the public health service: Another perspective on the Tuskegee Syphilis Experiment. *Bulletin of Historical Medicine, 80*, 291–316.

Pernick, M. S. (1997). Eugenics and public health in American history. *American Journal of Public Health, 87*, 1767–1772.

Russell, B. (1929). *Marriage & morals*. New York: Horace Liveright. (Liveright paperbound edition, 1970)

European Public Health Alliance

The European Public Health Alliance (EPHA) is composed of more than 100 European and international nongovernmental and not-for-profit organizations that work together toward the common goal of protecting and promoting the public health of all Europeans. A second mission of the EPHA is to support collaboration

between European Union (EU) institutions, nongovernmental organizations (NGOs), and EU citizens in support of public health policies in Europe. The EPHA was founded in 1993 and has its headquarters in Brussels. It receives funding from the European Commission, EPHA members, and from its publication subscriptions.

The early work of the EPHA was to increase NGO activity and not-for-profit involvement in public health policy at the European level, in order to enlarge the role these organizations played in the process of making health policy process and to stimulate involvement in important health programs. Members were solicited from national organizations and NGOs with an interest in health or health-related areas. EPHA members come from a variety of backgrounds, including health care professionals, health advocacy groups, caregivers, patients, and consumers. Before the EPHA was created, NGO activity in Europe in health-related programs was restricted to organizations working only on one issue, such as cancer. Through the EPHA, the focus was shifted to viewing determinants of health as key aspects of health policy development.

Since its founding, the EPHA has had a significant impact on the development and strategy of European health policy by monitoring the health-policy-making process within EU institutions and by maximizing communication regarding health promotion and public health policy among institutions and the public. The four key themes of EPHA activities include (1) health, human rights, and social justice; (2) health and sustainable development; (3) health and consumer issues; and (4) health and enlargement process. The EPHA actively represents the public health interests of EU citizens and disseminates information on the public health policies and EU policies that affect the health of EU citizens. Additionally, the EPHA helps raise awareness of the public health element in many other EU policy areas such as consumer protection, environment, and agriculture.

The EPHA has several member groups that focus on a specific public health concern. For example, the EPHA Environment Network (EEN) advocates the protection of the environment as a means of improving public health in Europe. The core of the EEN's work revolves around ensuring clean air and water and reducing noise, pesticides, and chemical pollutants. Topics of other working groups include public health policies, health inequalities, healthy lifestyles, and disease-related issues.

Communication among members is enhanced through working groups, conferences and events, e-mail and the Internet, and sharing of information. The EPHA publishes a magazine, the *European Public Health Update*, bimonthly, in addition to an electronic newsletter for its members, and other briefings and press releases throughout the year.

—Britta Neugaard

See also European Union Public Health Programs; Governmental Role in Public Health

Further Readings

European Public Health Alliance. (2006). *European public health alliance: Putting citizens' health at the heart of Europe*. Retrieved April 15, 2006, from http://www.epha.org.

European Union Public Health Programs

The European Union's (EU) Health and Consumer Protection Directorate General (DG) works with the 25 EU member states to protect and promote the health of the EU's citizens. The responsibility for public health is split between the member states and the EU. The EU directs the overall health policy; however, the delivery of health care is left up to the member states. The EU is advised on health policy by the EU Health Forum, a platform for organizations working in the health field to participate in health policy making.

The main goals of the EU's public health policy are to help EU countries develop expertise on health issues, to share best practices, and to improve coordination within the EU and thus make possible a rapid response to public health threats such as infectious disease outbreaks. Recently, the EU has sought to encourage cooperation between health care systems in EU member states as these systems have become more intertwined than ever before. As the connection between the health care systems grows within the EU, new health policy issues arise, such as quality and access in cross-border care. Also, there are cross-border health threats, such as avian (bird) flu, that require collaboration among member states and a coordinated EU health policy. The EU's Health and Consumer Protection DG is a distinct European

Commission department that is divided into three areas: public health, food safety, and consumer affairs. The Health and Consumer Protection DG (also referred to as DG SANCO) has approximately 800 staff members who are based in Brussels, Belgium; Luxembourg; and Grange, Ireland. The EU Commissioner for Health and Consumer Protection oversees the activities of the Directorate. Within the public health area, laws are developed on matters such as safety and quality of blood, blood derivates, and human tissues and human cells used in medical therapies. There are also laws in place regarding the advertising and manufacturing of tobacco. The Health and Consumer Protection DG is charged with making sure that these laws remain up-to-date. However, enforcement of these laws is left up to the national, regional, or, sometimes, local governments of the EU member states.

The Commission's Health and Consumer Protection DG is divided into the following departments: General Affairs, Consumer Affairs, Public Health and Risk Assessment, Animal Health and Welfare, Safety of Food Chain, and Food & Veterinary Office. The Public Health and Risk Assessment department works to ensure human health protection in EU policies, to take actions to improve public health in the EU, and to prevent human illness and diseases. There are a number of topics that the Health and Consumer Protection DG works on, including communicable and rare diseases, HIV/AIDS, injury prevention, tobacco, nutrition and obesity, bioterrorism, and the environment. The Health and Consumer Protection DG publishes a monthly newsletter, *Health and Consumer Voice*, which covers news on the public health developments in the EU.

On January 1, 2005, the European Commission created an Executive Agency for the Public Health Program to improve the EU's community public health programs. The agency is based in Luxembourg and is charged with implementing the public health programs, managing the budget, awarding contracts and grants, and organizing scientific conferences and expert panels. The Executive Agency collaborates with scientific experts in the field of public health and supports scientific committees that are composed of representatives from the EU member states.

The Health and Consumer Protection DG is the Commission liaison for three EU agencies: the European Food Safety Authority (EFSA) in Parma, Italy; the European Centre for Disease Prevention and Control (ECDC) in Stockholm, Sweden; and the Community Plant Variety Office (CPVO) in Angers, France. The ECDC was created in 2005 to help maintain and improve the health of the EU citizens. The ECDC is the European counterpart to the U.S. Centers for Disease Control and Prevention (CDC). The ECDC provides surveillance of communicable diseases within the EU and partners with national public health agencies across Europe. Its role is to increase existing disease surveillance that was managed by the Commission's Communicable Disease Networks. The ECDC is responsible for (1) epidemiological surveillance and laboratory networking; (2) operating the early warning and response system; (3) providing EU members with technical assistance; (4) responding to disease outbreaks; (5) increasing EU countries in preparedness for health emergencies; and (6) communicating health threats.

On September 23, 2002, the European Council and the European Parliament established the Community Public Health Program for the years 2003 to 2008. The public health program was established based on Article 152(4) of the treaty establishing the European Community. The Director of the Executive Agency manages the public health program.

The three objectives of the EU's public health program are (1) to improve information and knowledge to promote public health and health systems, (2) to improve rapid response capability to public health threats, and (3) to promote health and disease prevention through targeting health determinants. Public health program initiatives include creating an epidemiological surveillance system, addressing problems of antimicrobial resistance and bioterrorism, and developing strategies to prevent communicable diseases. The public health program fosters collaboration with international organizations such as the World Health Organization (WHO). A new public health program for the years 2007 to 2013 was adopted on April 6, 2005, by the EU Commission. The new program extends the current EU's public health program but joins together EU health and consumer protection policies under one framework.

—Britta Neugaard

See also European Public Health Alliance

Further Readings

European Commission Health and Consumer Protection Directorate-General. (2006). Retrieved April 20, 2006, from http://europa.eu.int/comm/health/index_en.html.

EuroQoL EQ-5D Questionnaire

The EuroQoL EQ-5D questionnaire is a brief utility index designed to measure health-related quality of life and health preferences. The Panel on Cost-Effectiveness in Health and Medicine recommends the use of societal or community preferences to calculate quality-adjusted life years (QALYs) for the reference case analysis in cost-effectiveness analyses. Preference-based health state classification systems, or "utility indices," represent a class of survey instruments recommended as suitable for providing these community preferences for reference-case analyses. While the mechanisms for creating scoring functions vary for different utility indices, these indices all consist of a descriptive health survey on which the respondent describes his or her state of health on multiple domains. The survey is then scored to produce a summary measure, which represents an average preference score of the population in which the scoring algorithm was derived. The preference scores calculated by this class of instruments are suitable for use in QALY calculations to be used in any analysis combining morbidity and mortality into one summary measure of population health, for example, decision analyses and cost-utility analyses.

When completing the EQ-5D, a respondent rates his or her health on five domains; these responses are then scored to provide a community utility. The survey also has a visual analog scale (VAS) on which the respondent rates his or her health, providing the respondent's own preference for health in addition to a community-based utility score. The EQ-5D's brevity and the availability of translations in numerous different languages have contributed to this instrument's popularity; a recent systematic review by Rasanen and colleagues found that almost half (47%) of all studies reporting QALY calculations based on valid assessment techniques used the EQ-5D for the utility values in QALY calculations.

The Survey

The EuroQoL (a contraction of "*Euro*pean *Q*uality *o*f *L*ife") Group, a multidisciplinary, international group of researchers, developed the EQ-5D; this survey has been in the public domain since 1990. The survey was designed to be a self-administered, simple, generic measure of health. Although originally designed to be used alongside other measures to allow comparability across settings, the EQ-5D has also been used as a stand-alone measure of health.

The EQ-5D is currently available in 60 different official language versions, and additional language translations are in development. The U.S. English language version is shown in Figure 1. The survey consists of one question for each of five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each domain has three levels: no limitations in function, some limitation, and extreme limitations or unable to function, giving a total of 243 different combinations of responses.

The EQ-5D has been shown to be valid in a number of populations and health conditions. In general populations, significant proportions of respondents had the highest score possible on the EQ-5D index (a ceiling effect), although in comparison to another utility index, the SF-6D, the EQ-5D index may have fewer people achieving the lowest possible score (a floor effect).

In addition to using the EQ-5D index to derive health utilities for QALY calculations, the survey has been used as a descriptive measure of health status, for example, to describe the health of populations or the impact of health care interventions such as surgical procedures or medical therapies. The index score can be used to provide a summary score across all five domains of health measured, and the function of individuals on the five separate domains of health has also been reported. While the index score gives a summary measure of preference or value for the person's state of health, a preference score alone does not state *why* one state of health is valued more or less than another. The five-question index describes the function of the respondent on the domains of health represented in the survey, allowing exploration of the specific domain differences that may explain differences in the overall score.

The EQ-5D is a generic instrument; it is designed to measure health across a wide range of settings and health conditions. In addition to using the EQ-5D as a stand-alone instrument, use of a generic instrument can be complementary to disease-specific instruments designed to be particularly responsive to changes in status of particular conditions. The EQ-5D and other generic health-related quality-of-life instruments were intended to cover broad domains of health; they may capture important changes in a domain of health

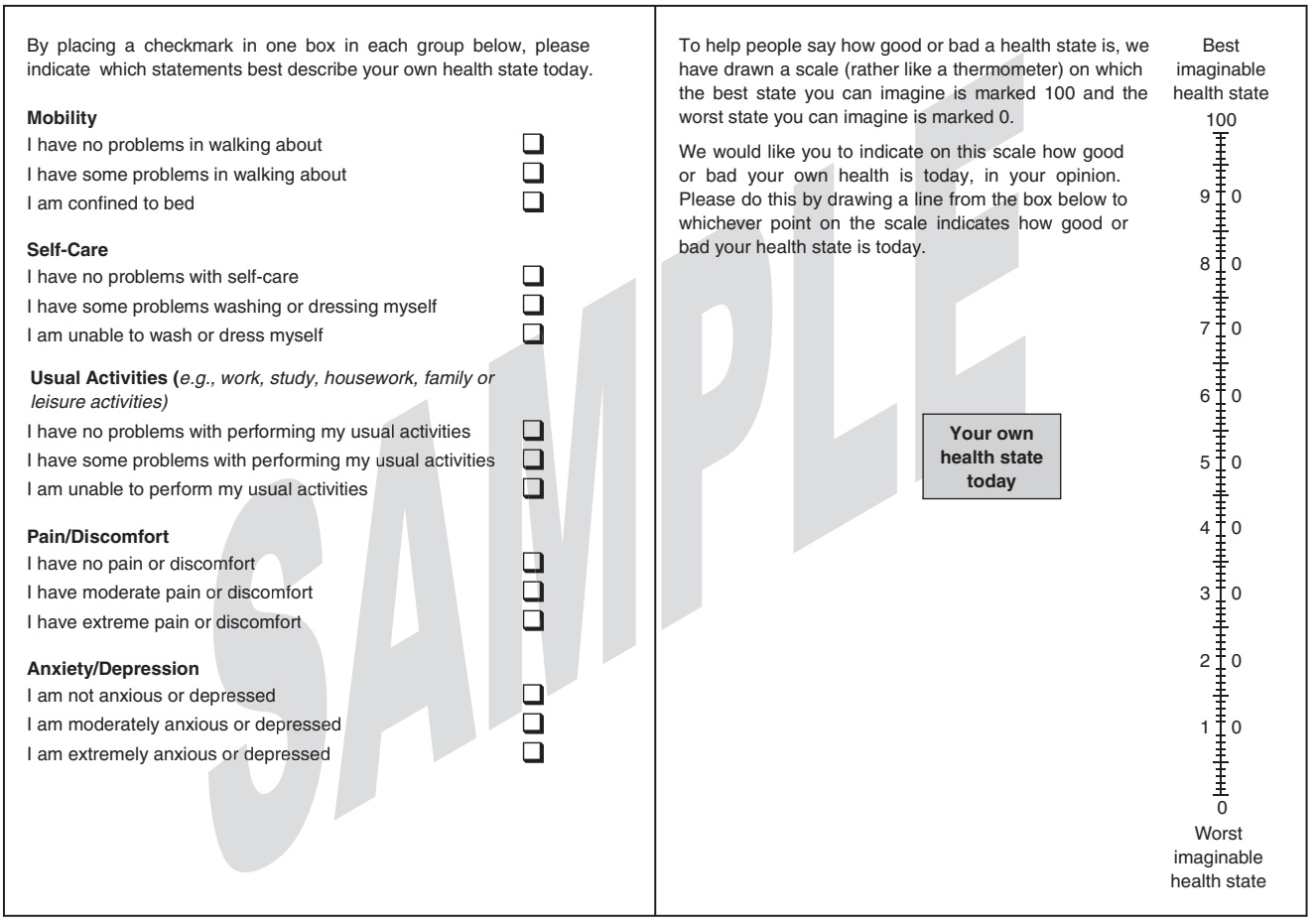


Figure 1 U.S. English Version of the EQ-5D Five-Question Index and Visual Analog Scale

Source: Reprinted with permission of the EuroQoL Group. This figure is for display only; any use for research or clinical purposes requires approval from the developers.

which a disease-specific measure does not cover. Perhaps even more important, use of the EQ-5D adds comparability of the sample in question to other populations. Thus, for example, a sample of patients with depression could be compared on overall or domain-specific health with age- and sex-stratified members of a general population or to patient samples with different conditions.

Scoring the Survey

Scoring algorithms for the five-question index have been developed by surveying a sample of respondents to determine the values that they place on health states as described by subsets of the 243 different possible EQ-5D states. Statistical methods were used to determine the independent contributions of each level of each of the five EQ-5D domains to the overall index score. A number of scoring algorithms are available for the EQ-5D; investigators choosing a specific method to calculate summary scores for the EQ-5D should consider the country or sample the survey will be used in and the health state valuation method that underlies the scoring algorithm. Currently, the EuroQoL Group reports that country-specific scoring protocols are available for eight countries using the VAS to value health states and for six countries using time trade-off (TTO) health utility assessments. Researchers have also developed a VAS-based, single scoring algorithm for six European countries (Finland, Germany, The Netherlands, Spain, Sweden, and the United Kingdom). Investigators comparing TTO-based scoring algorithms have found country-specific differences in values for specific health states.

The underlying valuation techniques used to develop scoring algorithms should also guide the choice among the algorithms. Since TTO assessments are generally considered more theoretically correct assessments of health utility than VAS assessments, a TTO-based algorithm is preferred for use in QALY calculations, for example, for use in cost-utility analyses. Either TTO- or VAS-based scores could be used if estimating formal utilities for quality-adjusted life expectancy is not the goal. Of the English Language surveys, TTO-based scoring algorithms based on nationally representative samples are available for the United Kingdom and the United States. In both these scoring algorithms, a person reporting the best function on all five domains receives a score of 1.0. Those reporting the worst function on all five domains receive scores of −0.594 for the U.K. algorithm and −0.109 for the U.S. algorithm. On a utility scale, 1.0 represents excellent health and 0 represents death, so for both the U.K. and the U.S. scoring algorithms, some states are valued as worse than death. The U.S. scoring algorithm is available programmed for common statistical software through the Agency for Healthcare Research and Quality (AHRQ) Web site.

The VAS is treated as a separate score and is not included in the index score calculation. As opposed to the index score, which represents a community or societal preference for a state of health, the VAS represents the respondent's own preference for his or her state of health. This VAS is anchored by "best imaginable health state" at the top, with 0 representing "worst imaginable health state." The anchor for 0 makes the VAS problematic theoretically for use in QALY calculations, since in the standard health utility metric, death is anchored at 0. However, the VAS can be used as a summary rating of the respondent's health.

Available Data and Norms

Population-based norms data are available for the EQ-5D for 16 countries, including the United States. Recently, Luo, Johnson, Shaw, Feeny, and Coons (2005) have published U.S. population norms using the U.S. TTO-based scoring algorithm in a national sample, and Hanmer, Lawrence, Anderson, Kaplan, and Fryback (2006) used a national sample from the Medical Expenditure Panel Survey (MEPS) to develop U.S. national norms for both the EQ-5D U.S. TTO-based and U.K. TTO-based scoring algorithms.

The MEPS survey Household Component, a nationally representative household survey of the U.S. noninstitutionalized civilian population, included a self-administered written version of the EQ-5D for the years 2000 to 2003. In addition to the EQ-5D, the survey includes other health status information, self-reported health conditions, demographic information, and health care utilization and expenditures. In 2003, EQ-5D data were available for more than 20,000 adults aged 18 years and above. De-identified data are available for public use at the MEPS Web site. In addition to norms data, Sullivan and Ghushchyan (2006) used these data to evaluate the effects of a number of health conditions on EQ-5D scores.

Terms of Use

While the survey and the scoring algorithms are in the public domain, use of the instrument for research or clinical purposes requires approval of the developers. Except for commercial uses, use of the survey is free. Potential users can contact the EuroQoL Executive Office at userinformationservice@euroqol.org.

—William F. Lawrence

Disclaimer: The views expressed in this entry are those of the author, and no official endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services is intended or should be inferred.

See also Disability Epidemiology; Economic Evaluation; Quality of Life, Quantification of; Quality of Well-Being Scale (QWB); SF-36® Health Survey

Further Readings

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care, 35*(11), 1095–1108.

Hanmer, J., Lawrence, W. F., Anderson, J. P., Kaplan, R. M., & Fryback, D. G. (2006). Report of nationally representative values for the non-institutionalized U.S. adult population for seven health related quality of life scores. *Medical Decision Making, 26*, 391–400.

Kind, P., Dolan, P., Gudex, C., & Williams, A. (1998). Variations in population health status: Results from a United Kingdom national questionnaire survey. *British Medical Journal, 316*(7133), 736–741.

Luo, N., Johnson, J. A., Shaw, J. W., Feeny, D., & Coons, S. J. (2005). Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. *Medical Care, 43*(11), 1078–1086.

Rasanen, P., Roine, E., Sintonen, H., Semberg-Konttinen, V., Ryynanen, O. P., & Roine, R. (2006). Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. *International Journal of Technology Assessment in Health Care, 22*(2), 235–241.

Shaw, J. W., Johnson, J. A., & Coons, S. J. (2005). US valuation of the EQ-5D health states: Development and testing of the D1 valuation model. *Medical Care, 43*(3), 203–220.

Sullivan, P. W., & Ghushchyan, V. (2006). Preference-based EQ-5D index scores for chronic conditions in the U.S. *Medical Decision Making, 26*, 401–409.

Web Sites

Agency for Healthcare Research and Quality: http://www.ahrq.gov/rice/EQ5Dscore.htm.

EuroQoL: http://www.euroqol.org.

Medical Expenditure Panel Survey (MEPS): http://meps.ahrq.gov.

Evans's Postulates

See Koch's Postulates

Event History Analysis

See Survival Analysis

Evidence, Legal Admissibility of Scientific

Product liability cases and toxic tort cases (e.g., asbestos, lead poisoning) are frequently in the news, and scientific and epidemiological evidence is often presented to support or refute the claim that the product or toxin has caused harm or injury. For a claim to succeed, the law requires evidence of general causation and specific causation. Specific causation usually requires medical evidence directly on point regarding the claimant and his circumstances. General causation takes on the larger issue of whether the alleged product or toxin can actually cause the injury. General causation requires scientific evidence of general causation and usually requires an expert such as an epidemiologist to present the scientific evidence of general causation. Accordingly, epidemiologists serve an important role in establishing or refuting claims concerning whether substances such as tobacco, alcohol, fried foods, silicone, Agent Orange, asbestos, hair coloring, electromagnetic fields, caffeine, and so on are capable of causing the injury complained of. The courts, however, have certain rules determining who can call themselves experts.

The law classifies witness testimony into two types: the lay witness and the expert witness. The legal distinction between the two is that under the Federal Rules of Evidence (FRE) and many state rules of evidence, experts proffering expert testimony enjoy greater latitude than lay witnesses, on two counts: Expert evidence may rely on hearsay, and expert evidence involves an opinion. While the distinction may appear trivial to the layperson, it is nevertheless one of the most critical distinctions at high-stakes trials and complex litigation. Accordingly, who is entitled to provide expert testimony can be a major point in any complex trial. The rules defining expert testimony have proven crucial in the progress of other medical device and pharmaceutical liability cases.

Until 1923, the FRE provision for expert testimony was determined generally by qualifications of the expert. Then, as a result of the ruling in *Frye v. United States*, a 1923 case concerning a novel polygraph test challenged in an intermediate Federal Court of Appeals, the standard became general acceptance of the expert's method or procedure. This general acceptance standard, known as the *Frye* test, prevailed for 70 years, until the decision in *Daubert v. Merrell Dow Pharmaceuticals* (1993).

The current rules governing the admissibility of scientific evidence in law in the U.S. federal courts were developed in response to *Daubert*, based on a lawsuit charging that Benedictin, an antinausea drug taken by many pregnant women to combat morning sickness, was a teratogen (a substance that causes birth defects). Epidemiologists and other scientists involved with the Benedictin litigation created enough confusion on matters of evidence to elevate the legal issues to the U.S. Supreme Court. The Court set about defining what exactly scientific knowledge is within the FRE, and it then gave the trial courts responsibility for ensuring that evidence proffered as scientific knowledge has evidentiary reliability, defined as scientific validity. To help the trial judge determine whether the scientific testimony is reliable and thus admissible under this standard, the Court provided

trial judges with four nonexclusive factors with which the judges may assess scientific validity: (1) testability of the claims made by the expert, (2) error rate of the method of investigation, (3) peer review of the conclusions reached, and (4) their general acceptance. In addition to these factors, the expert must be qualified and his testimony must "fit" the facts of the case, essentially a higher standard of relevance.

The factor of testability (or falsifiability) concerns the question of whether the scientific knowledge is falsifiable. Scientists and philosophers since Aristotle have identified the character of testability as a necessity to enjoy the status of "knowledge" or "science."

For example, if an epidemiologist hypothesizes that aflatoxins are carcinogenic, he can easily test his hypothesis (albeit with some systematic error) through a case-control study. Using the odds ratio as a measure of the carcinogenic effect, the p value as a measure of statistical significance, and a qualitative assessment of the bias inherent in case-control studies, he has tested his hypothesis to determine whether it is true or false.

Known or potential rate of error is a second factor to consider. Error rate suggests quantifiable errors such as the p value or measures of Type 1 and Type 2 errors. In reality, however, these measures of "error" provide information concerning only random error, while an epidemiologist must also account for measurement error and systematic error. Estimation of measurement error may be a quantitative or qualitative process; if quantitative, it requires establishing the reliability and validity of particular methods of measurement. Assessing systematic error or bias requires a qualitative assessment of the shortcomings found in observational studies, including the lack of randomization, the potential absence of matching, and other methodological trade-offs.

There is no simple threshold error rate to determine whether the testimony qualifies as scientific knowledge and is therefore admissible. Furthermore, the rate in "error rate" is probably an oversight, as the Court was not actually contemplating a "rate" in the technical epidemiological sense of frequency per unit of time.

A third factor is peer review. While there is a presumption that published work vetted through peer review evidences scientific knowledge, there are also exceptions to this. The Court recognized that some work is scientifically sound even though it has not been published or peer reviewed, while some peer-reviewed journals publish work that may not be scientifically sound. Furthermore, publication bias may complicate assessment of this factor. Nevertheless, work that is both published and peer reviewed stands a better chance of being admitted as evidence in a court of law. On the other hand, research done for the sake of litigation is suspicious to most judges. With admissibility in law, as in science, there is a hierarchy of credibility associated with journals and other vetting procedures.

The fourth nonexclusive factor is general acceptance. This refers to how credible certain types of studies are regarded within a profession. In epidemiology, hospital-based case-control studies are inferior to population-based case-control studies and cohort studies but are nevertheless generally accepted methods for epidemiological research. Accordingly, evidence derived from hospital-based case-control studies satisfy the general acceptance factor, although they may be more heavily scrutinized under the "rate of error" factor.

Here, general acceptance is the same as the *Frye* test above but operates as only one of several factors to be taken together as a whole in the determination of admissibility, whereas before *Daubert*, it was the only factor.

The Court's opinion in *Daubert* was designed to be used as guidance and not a checklist. On one hand, placing the "general acceptance" test as a mere one-of-four-nonexclusive factors provided a relaxation of general acceptance as the sole criterion for admissibility and therefore the possibility that novel scientific methods may find their way in court (e.g., the novel polygraph test in the *Frye* case). On the other hand, the other factors raised the bar for admissibility since there are more factors to consider. In practice, *Daubert* has been applied to be more restrictive of expert evidence. Two other cases at the U.S. Supreme Court (*General Electric Co. v. Joiner*, 1997; *Kumho Tire Co. v. Carmichael*, 1999) provided further guidance on the application of *Daubert.* In December of 2000, the lessons of these three cases were woven into an amended FRE 702.

Other courts have adopted other rules of admissibility (and rules for sufficiency) based on other jurisprudential notions. One federal judge developed a rule in the Agent Orange litigation cases captioned *In Re Agent Orange.* Judge Jack Weinstein's rule restricted epidemiological risk factor (antagonistic) evidence if it failed to demonstrate a relative risk

(*RR*; or odds ratio) of 2 or more. His basis for such a rule lay in the following analysis: If the incidence of disease *X* is 20 per 1,000 persons in the general population, then the incidence must be at least 40 per 1,000 in the population exposed to a potential risk factor, such as Agent Orange, to justify a claim that the risk factor could *legally* cause the disease *X*. Such an elevated incidence ensures without any further data that for any given member of the exposed population who contracts the disease, there is as much or a higher likelihood that the disease was associated with the risk factor rather than the result of chance. This "*RR* of 2" rule is not recognized in most courts, but it nevertheless is the rule for some jurisdictions.

Even when expert testimony qualifies for admissibility under *Daubert* and any other rules of the jurisdiction (e.g., *RR* of 2), the evidence may be struck through application of other rules such as FRE 403, which permits the exclusion of evidence when its value is substantially outweighed by "the danger of unfair prejudice, confusion of the issues, or misleading the jury." Most judges are reluctant to exclude evidence under FRE 403, because they believe the best place for testimony to be tried is through the strenuous advocacy of each party before a jury. Nevertheless, FRE 403 holds scientists accountable for a proper presentation of their evidence.

An example of scientific testimony that probably could have been excluded through FRE 403 was presented during the Benedictin litigation. As part of his expert testimony, a toxicologist presented a chart consisting of more than 100 confidence intervals (CIs) calculated from small samples. He arranged the CIs in vertical format with each bar representing one of more than 100 rows and arranged the CIs by their upper boundary. The result was that a few upper boundaries fell short of the *RR* of 1, the method and its measure suggesting no relationship between Benedictin and alleged teratogenic effect. A large majority of the upper boundaries, as expected, conveniently drew the eye toward the bottom and to the right where upper boundaries went beyond an *RR* of 100 (due to small sample sizes). The overall visual effect of the chart was extraordinary: a quasi meta-analysis pointing to an extraordinary antagonistic effect. In reality, however, the average *RR* was much less than the cognitive bias displayed by the chart. Presenting numerous CIs in this manner has no basis in scientific practice.

—Mark Gerard Haug

See also Agent Orange; Asbestos; Love Canal; Thalidomide; Tobacco

Further Readings

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 979 (1993).

Faigman, D., Kaye, D., Saks, M., & Sanders, J. (Eds.). (2002). *Science in the law: Standards, statistics, and research issues.* St. Paul, MN: West.

Foster, K. R., & Huber, P. W. (1997). *Judging science: Scientific knowledge and the federal courts.* Cambridge: MIT Press.

Frye v. United States, 293 F. 1013 (1923).

General Electric Co. v. Joiner, 522 U.S. 136 (1997).

Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).

Reference manual on scientific evidence, Federal Judicial Center, 2d (2000).

Evidence-Based Medicine

Evidence-based medicine (EBM)—which might better be called evidence-based health care (EBHC) because it applies to all parts of the health care system and not just the practice of medicine by physicians—encompasses a set of tools for the enhancement of the practice of medicine. EBM uses those tools, many of which are drawn from epidemiology and biostatistics, to create a bridge between information gained from the study of populations and communities, on the one hand, and medical care provided to a particular individual, on the other. EBM requires that physicians and other medical professionals be able to critically appraise the medical literature and selectively apply information based on these critical appraisal principles to the individual patient.

Definition and History of EBM

To practice the highest quality of scientific medicine, physicians must bring the best information from medical research (and medical technology) to the patient's bedside. Secondary goals are to improve the health of the public through control of epidemic diseases (whether caused by microorganisms or environmental contaminants) and comforting the patient and their immediate social group in times of illness. In the Glossary on its Web site, the Centre for Evidence-Based Medicine at Oxford University in the United Kingdom defines EBM as "the conscientious, explicit and judicious use of current best evidence in making

decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research." The Centre also defines EBHC as an extension of EBM "to all professions associated with health care, including purchasing and management" (Centre for Evidence-Based Medicine).

The practice of medicine requires physicians to make correct diagnoses and choose the best treatment to improve the health or reduce the burden of illness for their patients under conditions of significant uncertainty. The health care worker needs to effectively access the best and most current information from the medical literature, critically evaluate this information, and determine how and when the results will be applied to the patient sitting in front of them.

History of EBM

Elements of EBM can be found in the biblical Book of Daniel where a description of a trial of diet is given. The participants were Daniel's friends, and they were "randomized" to eat only vegetables or the king's food. Hippocrates told the physicians of his day to observe their patients and to perform only those actions that could be helpful and would "first, do no harm," implying the ability to distinguish helpful from potentially harmful therapies. An 18th-century British physician, George Fordyce, demanded that the medical profession provide better evidence for the therapies of the day. Captain James Lind, a British naval surgeon, performed a nonblinded but randomized clinical trial on a dozen Navy seamen with scurvy. The results clearly showed that citrus was vastly superior to the other treatments being tested. Pierre Charles Alexandre Louis was a French physician who applied the new science of statistics to show that bloodletting was unlikely to benefit patients with typhoid fever.

More recently, a 1947 editorial by Austin Bradford Hill in the *British Medical Journal* demanded that physicians use mathematics and statistical methods to evaluate the practice of medicine. American epidemiologist John Paul in the United States coined the term *clinical epidemiology* in the late 1930s, but this concept was neither accepted nor used by mainstream physicians and languished in obscurity until the first modern randomized clinical trial in 1948. This trial of streptomycin for patients with tuberculosis performed under the direction of the Medical Research Council of the National Health Service in the United Kingdom showed that the therapy was beneficial. Between 1950 and the mid-1970s, further elucidation of the nature of EBM in modern medicine occurred when Dr. Alvan Feinstein differentiated the science of clinical epidemiology as distinct from the traditional epidemiology of public health. This served as the foundation of the statistical revolution in medicine beginning in the 1960s. The research of Dr. John Wennberg in the 1970s demonstrated the large variation in the quantity of health care provided to populations living in relatively small geographical areas. During this time period, there was an explosion in the number of medical research articles published. Researchers in clinical epidemiology at McMaster University Health performed outcomes and process research, and this paved the way for wider dissemination of clinical epidemiology and developed a curriculum that incorporated clinical epidemiology into the medical curriculum.

The first modern systematic review of perinatal interventions was done in 1986 by a group at Oxford led by Iain Chalmers. Systematic reviews have become increasingly popular as the amount of medical literature continues to increase. A major source of clinical trial information and systematic reviews is the worldwide online Cochrane Collaboration, founded in 1993 and currently composed of more than 6,000 people in 60 countries. Currently, there are more than 30,000 trials entered into the Cochrane Controlled Trials Registry and more than 1,000 systematic reviews.

EBM soon became synonymous with the explicit application of the results of research published in the medical literature to improve patient care for the individual patient. However, Feinstein warned of the danger that EBM could become a substitute for critical thinking skills, and he defined the role of clinical epidemiology as making physicians' thought processes more transparent and explicit and improving the critical thinking required for modern scientific medical practice, rather than replacing it.

Through the 1990s, there has been an explosion in the number of courses teaching physicians how to become more intelligent consumers of the medical literature. These are often traditional epidemiology and biostatistics courses modified through the use of explicit tools of EBM and the use of statistical methods in medical decision making. EBM as a basic principle of medicine teaches all health care workers the application of critical thinking to improve the care of the individual patient. It encompasses clinical

epidemiology, research methodology, narrative-based medicine, ethics, public health, health policy, social and community medicine, and population medicine. EBM bridges the care for populations and communities with that for the individual.

EBM Sources and the Levels of Evidence

Various groups have tried to develop ways to package critically appraised and filtered evidence to make it more useful to individual practitioners. Access to these predigested "EBM reviews" is done through various online databases around the world. A major center for the dissemination of these sources of best evidence in the United Kingdom has been the Centre for Evidence-Based Medicine at Oxford University. The home of several EBM sources, it includes *Bandolier*, a biweekly summary of recent and interesting evidence evaluated by the Centre, which is free to all and available online. The Centre has other easily accessible and free features related to the practice of EBM on its main Web site. The *British Medical Journal* publishes an updated *Clinical Evidence*, a summary of critically evaluated topics in therapeutics that are regularly updated and available online.

Many commonly used preprepared critical appraisals of various clinical questions are found in the Journal Club Bank (JCB) of the Evidence Based Interest Group of the American College of Physicians (ACP) and the Evidence Based Emergency Medicine Web site. This consists of critically appraised topics (CATs) or summaries of research studies and can also be found on the Evidence-Based Medicine Resource Center of the New York Academy of Medicine. The CAT format developed by the Centre for Evidence-Based Medicine is being made available on CD-ROM for use outside the Centre. The University of Sheffield (United Kingdom) has one of the most complete resources listing EBM-related Web sites, "Netting the Evidence." Other medical organizations are beginning to embrace the CAT format to disseminate critical reviews on the Web.

EBM focuses on evidence that will make a difference for the patient. Disease-oriented evidence (DOE) is not necessarily "patient-oriented evidence that matters" (POEMs) because DOE may reflect a change in disease status that has no direct relationship to outcomes that matter to an individual patient; for instance, drugs such as statins lower cholesterol but may not reduce mortality or improve quality of life. POEMs are a format developed by family physicians for the American Academy of Family Practice; further information can be found on the InfoPOEMs Web site.

Evaluation of the strength of evidence for a particular clinical query has led to several methods of ranking different types of studies from most to least important in having the ability to determine causation for the question at hand. The Centre for Evidence-Based Medicine developed the most commonly used scheme of categorization, which depends on the nature of the clinical query. These have been challenged as being too doctrinaire, and users should employ them in a flexible manner and look critically at each study evaluated, regardless of the study design. There is also concern among EBM scholars that the research agenda has been hijacked by proprietary interests such as pharmaceutical and technology companies as the studies sponsored by these groups are frequently randomized clinical trials with biases designed to achieve results favorable to the sponsoring organization.

The Five-Stage Process of EBM

The phenomenal growth in the amount of medical research information available has made it both more difficult and important for physicians to have the tools to assess this information. Breakthroughs in information systems technology, including Internet access to MEDLINE via PubMed and other medical databases, allow physicians to obtain the most current information to answer educational needs more quickly and easily than in the past.

Once an educational need has been recognized, the next step is to develop a clinical question that maximizes the likelihood of finding good-quality evidence through a search of the literature. This is best done using a four-part PICO question, which includes the following elements: patient or population (P), intervention or exposure to a risk factor (I), comparator (C), and outcome (O).

The next step is critical appraisal of the medical literature, which is the heart of EBM and attempts to identify potential shortcomings of the research study being evaluated. Is the study valid or are there sources of bias? The essence of the critical appraisal part of the EBM process is asking if there are other reasonable explanations for the results of the study. Finally, the reader must draw inferences and apply the results of the study to the care of their individual patient and

integrate the evidence into actual practice. The complete understanding of sophisticated statistical testing is less important than the application of common sense and skeptical evaluation of what is read: Understanding research designs and basic statistical methods will allow physicians to critically evaluate most published clinical research and avoid many errors of interpretation. Evaluation of the methodology of a research study is the most important part of the critical appraisal of the literature and does not require any sophisticated mathematical abilities.

Understanding the research study design will help identify most of the problems that can potentially influence the results of a poorly done research study. To determine causation for diseases with multifactorial causes requires showing that (a) the cause and effect are associated with each other more likely than by chance alone, (b) the cause precedes the effect, and (c) changing the cause changes the effect. These three conditions are known as contributory causes and all three are required to prove causation for a multifactorial disease.

The Hierarchy of Evidence: Study Design and Minimizing Bias

Research answers questions about populations by studying samples of individuals from a given population. Individuals have variable characteristics that might affect outcomes of research, and the design of a study will alert the critical reader to potential problems in the design of a study. The best research design is one that minimizes the chance of bias, and it is the responsibility of the researcher to minimize bias in a study. Since this cannot always be done, it is the responsibility of the reader to determine if biases that exist in a study, whether real or potential, are enough to affect the outcome. The conclusion may not be compatible with the research hypothesis when there is a large degree of bias in the conditions of the research.

As many of the population characteristics as possible should be represented in the study sample. In most types of studies, the sample is divided into two groups to test the hypothesis that they are different in some important characteristic. If the two groups are not equivalent with regard to all other characteristics, confounding of the results can occur, leading to an incorrect conclusion. It might be erroneously concluded that a difference in the outcome between the groups occurred because of a presumed causative factor when in reality it was produced by a difference in the baseline characteristics, and bias is present.

The strongest type of study in the hierarchy of study designs is the randomized clinical trial (RCT) because it is most likely to be able to prove causation and least likely to contain biases leading to incorrect or misleading results. In RCTs, the study subjects are assigned to the treatment (exposed) or comparison (placebo or not exposed) on the basis of chance alone using some technique to assure random placement of each participant. This maximizes the probability that the two groups are equal with respect to all the other characteristics that could affect the outcome under consideration. RCTs are the best design to minimize bias but are usually costly in money and time, and are not efficient if the outcome being studied is rare.

In a cohort study, the subjects are identified based on their exposure or lack of exposure to the risk factor being studied. A researcher determining whether cigarette smoking causes brain tumors could take a sample of smokers and nonsmokers and follow them for a period of time to determine the numbers of brain tumors developed by the subjects in each group. The subjects are not assigned to the two groups and may differ on important characteristics: For instance, smokers may have a higher degree of exposure to other toxins than nonsmokers, which could be the contributory cause of brain cancer. Cohort studies usually cost less than RCTs and may allow for the study of issues for which randomization would be unethical or difficult.

Case-control studies begin with cases who have and controls who don't have the outcome of interest; matching is usually performed to make the two groups as similar as possible on key covariates. For instance, a researcher might study 20 patients with brain tumors and 20 of similar age and gender without brain tumors to look at the proportion of cigarette smokers in each group. Bias is more difficult to avoid in a case-control study, but this type of study can be completed more quickly than an RCT or cohort study and is generally less costly especially if the outcome is rare.

A case series is simply the description of the characteristics or clinical course of a set of individuals with a given exposure or outcome. For instance, a surgeon might describe his or her experience using a new operative technique for brain tumors, including the frequency with which the patients were cured or reached some defined outcome. The reader cannot tell if the new procedure is better than existing operative technique without a comparison group of patients

receiving the standard or no therapy. Case series are valuable for generating hypotheses or suggesting necessary studies for future research.

Cross-sectional studies measure the relationship between variables at one point in time and can calculate the frequency of a variable in a sample at a given point in time, its prevalence. They cannot prove the temporal relationship between cause and effect and are used to generate new research hypotheses.

Assessing the Significance or Impact of Study Results

The impact of a study tells the reader the likelihood of an association between the outcomes of the two groups (treatment and comparator or exposed and nonexposed). All the types of studies described above, except for case series, generally include an evaluation of whether the results were statistically significant. However, even with the best attention to study design, a study may find a difference between groups that is not truly present in the larger population or no difference when one really exists but was simply not found in the study sample. These are known as Type I and Type II errors, or alpha and beta, respectively. Common causes of a Type I error in a study are multiple comparisons and composite outcomes. Multiple comparisons done between two groups of patients are sometimes known as "dredging the data" or "fishing for results" because as more tests are done, it becomes increasingly likely that a statistically significant difference between two study samples will be found due to chance, when in fact no such difference exists in the larger population. Composite outcomes, that is, when several outcomes are combined to create a single outcome, may also make a Type I error more likely when the different outcomes do not have the same values (e.g., death, myocardial infarction, and repeat admission for chest pain). Subgroup and post hoc analysis of the data are other ways in which a Type I error can occur.

The most common cause of Type II errors is insufficient sample size, also known as an insufficiently powered study. Because the power to detect a significant difference between groups rests partly on the sample size, a result that is not significant in one study might well be significant in a larger study: Beyond the fact of whether significance was achieved or not, therefore, the reader must also evaluate the possibility of a Type I or Type II error.

Confidence intervals are commonly reported in the medical literature as well as statistical significance: They report how much the estimate of any outcome may vary if the study is repeated with different samples from the same population. This tells the precision of the result and can help the physician and patient make an informed decision on the certainty of the evidence.

Meta-Analysis and Systematic Reviews

Review articles summarize the literature on a topic; however, this is often done in a subjective manner that may include significant author biases (e.g., in the choice of studies included or the relative emphasis placed on particular studies). Systematic reviews critically combine multiple studies that look at the same research question. The results of multiple studies can be examined statistically in a meta-analysis that "transcends" simple analysis to reconcile studies with different results. Meta-analyses of multiple negative studies may uncover Type II errors due to an inadequate sample size of one or more of the studies included in the analysis. Meta-analysis can also help identify a study with a Type I error or that has outlier results as part of a collection of many other studies.

Performance of a meta-analysis starts with an exhaustive search for studies, including both Medline and unpublished studies and dissertations. The studies found are then critically reviewed and graded using a standardized scheme. The statistical results are then compared and the presence of heterogeneity among the studies can be determined. If studies are heterogeneous, they cannot be directly compared, and this process may uncover one outlier among the studies that caused the heterogeneity. The reasons for this will usually be found in the methodology of the outlier study. Finally, summary statistics can be calculated and conclusions drawn. A technique known as cumulative meta-analysis can be done whenever a new study is reported on a given topic. This type of analysis will determine when in time the intervention first shows statistically significant results. Some believe this should always be done for all the previously done studies whenever any research study is published.

Assessing the Usefulness of Diagnostic Tests and Information

Accurate diagnosis is one of the primary duties of the physician. Diagnostic decision making consists of

gathering useful information to create and refine the differential diagnosis and then sharing this information with the patient in a way that facilitates informed decision making. Physicians must critically assess the value of the information gathered to help their patient, and while some of this information appears to be of value, it may be misleading and not discriminate among diagnostic possibilities. The patient's preferences must be added to the diagnostic process to assess the value of diagnostic information.

Useful clinical procedures and tests should be reliable and valid. Reliable tests are reproducible, and a reliable test that is run more than once on the same specimen will lead to the same result. The validity of a diagnostic test determines the ability to truly discriminate between patients with and without a given disease. Diagnostic tests should be judged against a "gold standard" test that conclusively defines the presence or absence of the disease. For example, the gold standard for bacteremia is a positive blood culture and for a malignant tumor a tissue specimen containing malignant tumor cells. Figure 1 shows the distribution of test results in patients with and without a disease. Extreme test results can definitely define who has the disease and who does not. Values near the cutoff point between normal and abnormal will demonstrate significant overlap for diseased and nondiseased individuals. Sensitivity and specificity are the mathematical descriptions of this degree of overlap. Studies of diagnostic tests that report the correlation between a diagnostic test result and the presence or absence of disease are not helpful to the clinician who needs to know the likelihood of illness after application of a diagnostic test.

Sensitivity and *specificity* are two technical terms used to evaluate the usefulness of diagnostic tests. Sensitivity is the percentage of patients with the disease who will test positive and is also called the true positive rate (*TPR* = ratio of subjects with the disease and a positive test to all subjects with the disease) as shown in Figure 2. Specificity or the true negative rate (*TNR*) measures the percentage of people without the disease who test negative and is the ratio of subjects without disease who test negative to all those who don't have the disease. Sensitivity and specificity are considered fixed characteristics of a test for purposes of decision making and usually do not change with the prevalence of disease in the population from which the patient came. In reality, most diseases have varying levels of severity related to different stages of disease leading the diagnostic test to demonstrate spectrum bias and be more sensitive in patients with classical or severe disease and less sensitive in patients with mild or early disease.

A test with high sensitivity, which minimizes the number of missed cases (false negatives), is preferred if the disease is readily treated or has serious morbidity. A negative result of the test will rule out disease. A test with high specificity, which minimizes the number of falsely identified cases (false positives), is preferred for diseases that have minimal morbidity or in those for which there is dangerous or risky treatment. A positive result on the test will rule in disease. Published sensitivity and specificity values are point estimates and should always be accompanied by 95% confidence intervals.

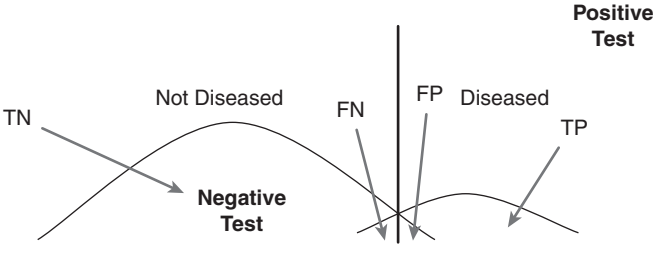


Figure 1 Results of a Diagnostic Test: Two Normal Curves, One for Diseased and One for Healthy (Nondiseased) People

Source: Mayer (2004).

	D+	*D*−
T+	*TP* (*T*+ | *D*+)	*FP* (*T*+ | *D*−)
T−	*FN* (*T*− | *D*+)	*TN* (*T*− | *D*−)
	Disease	**No Disease**

Figure 2 Results of a Diagnostic Test: The 2 × 2 Table Used to Calculate Sensitivity, Specificity, and Likelihood Ratios

Source: Mayer (2004). Reprinted with permission of Cambridge University Press.

Notes: Sensitivity = *TP*/*D*+
Specificity = *TN*/*D*−
LR+ = Sensitivity/(1 − Specificity) = *TPR*/*FPR*
LR− = (1 − Sensitivity)/Specificity = *FNR*/*TNR*
Bayes's theorem: $O_{pre} \times LR = O_{post}$

The clinician treating an individual patient wants to know the probability that their patient has the disease if the test is positive or negative. This is called the positive or negative predictive value or the posttest or posterior probability of disease given a positive or negative test. Positive predictive value is the ratio of true positive test results to all positive tests results or the fraction of patients with positive tests who really have the disease. Negative predictive value is the ratio of true negative test results to all negative tests results or the fraction of patients with negative tests who really don't have the disease. One minus the negative predictive value, the false reassurance rate is the probability that we are falsely reassuring patients who have a negative test that they are disease free, when in fact they actually have the disease. One minus the positive predictive value, the false alarm rate is the probability that we are falsely alarming patients who have a positive test that they are diseased when in fact they are actually disease free.

Predictive values of a test depend on the sensitivity, specificity, and prevalence of disease in the population from which the patient comes. The pretest prevalence, which is also called the prior probability of disease, is determined from the clinical presentation of the patient or the baseline prevalence of the disease in the population of patients with similar signs and symptoms. This is determined by the clinical experience of clinicians. The most experienced ones are better able to recognize a pattern of disease in patients with atypical presentations.

A direct way to calculate the posttest probability of disease use Bayes's theorem and likelihood ratios (*LR*). The *LR* combines sensitivity and specificity into one number and becomes a measure of the strength of a diagnostic test. The *LR* of a positive test ($LR+$) is the sensitivity divided by one minus specificity, and the *LR* of a negative test ($LR-$) is one minus sensitivity divided by the specificity (Figure 2). Strong tests have an $LR+$ greater than 10 or $LR-$ less than 0.1. Fair tests are those with $LR+$ between 2 and 10 and $LR-$ between 0.1 and 0.5. A test is almost worthless if the $LR+$ is less than 2 or an $LR-$ is greater than 0.5. Bayes's theorem revises disease probabilities using the formula Pretest odds $\times$ *LR* = Posttest odds. Although Bayes's theorem is daunting to most physicians, a nomogram is available to go from pretest to posttest probability using the *LR* without doing any computations. Continuous test results, such as the peripheral white blood cell count, can have their results broken into intervals to preserve test information that would be lost in reducing the test to a single normal or abnormal cutoff. The use of intervals creates multiple *LR*s (interval or *iLR*s) for each interval of test results.

Receiver operating characteristic (ROC) curves compare two or more tests or select the best single cutoff for a diagnostic test. The ROC curve plots sensitivity on the *y*-axis against one minus specificity (false-positive rate) on the *x*-axis for all possible test cutoff points. A perfect test represents 100% sensitivity and specificity, so there are no false positives or negatives. The point at the lower left of the ROC curve corresponds to 0% sensitivity and 100% specificity, and there are no false positives and no true positives. When looking for the best cutoff point or comparing two tests represented by curves that do not overlap, the best single cutoff point or test result is the one closest to the (0, 1) point.

The area under the ROC curve (AUC) represents a mathematical description of the likelihood that one can identify a patient with the disease using that test alone. A diagonal line drawn from the lower left to the upper right corner of the ROC curve has an AUC of 0.5 meaning that the probability of identifying a diseased patient from one without the disease is 50% or the same as a coin toss. The AUC is useful for evaluating two tests whose ROC curves cross or a single test to determine its usefulness in general. Ideally the AUC should be as near to one as possible. Before deciding which test to use, the clinician must assess the trade-off of sensitivity for specificity for each test and cutoff point, balancing the harm of missing a patient with the disease and the risk of treating a patient without the disease.

Expected Values Decision Making

Medical decision making combines the probability of an event with its value or risk. Done for generations of physicians "by the seat of their pants," it can now be done using a mathematical method of determining the optimal decision in medicine. Expected values decision making uses the concept of instrumental rationality to determine the optimal course of action based on the combination of probability of an event and the utility or value of the outcome. Instrumental rationality begins by using a decision tree showing all the possible actions that would be taken for a particular therapeutic or diagnostic decision. The starting point is

a place where the physician must make a decision and then each outcome of the decision is determined. If one choice is surgery, the outcomes could be death during the operation, complete cure, or some intermediate outcome such as relief from symptoms but shortened life. A probability and a patient value or utility is associated with each of these outcomes. After the tree has been constructed, the probability is multiplied by the utility for each branch until one gets back to the starting point, and a final expected value is obtained for each decision. The one with the highest value would be the preferred decision. A sensitivity analysis for the tree, incorporating plausible ranges of values for probability and utility for each of the decisions, should be done. If the final outcomes are pretty much the same for these different values, the tree is said to be "robust" and the results considered reasonable.

The Threshold Approach to Diagnostic Testing

Pauker and Kassirer (1980) introduced the concept of the threshold approach to diagnostic testing to help clinicians decide on whether to perform a test. Using this method maximizes the effectiveness of diagnostic testing and limits unnecessary testing. If the clinician judges the prior probability of disease to be below the testing threshold, then the patient is so unlikely to have the disorder that a diagnostic test would not raise the probability sufficiently to change the decision not to treat for the disease. For prior probabilities below this level, the test should not be done. If the clinician judges the prior probability of disease to be above the treatment threshold, then the patient is so likely to have the disorder and a diagnostic test would not lower the probability enough to change the decision to treat for the disease. For prior probabilities above this level, the patient should be treated and the test should not be done. If the prior probability is between the two thresholds, the patient should be tested and treatment should be based on the test result. Thresholds are determined by balancing the benefits and risks of appropriate therapy, the risks of inappropriate therapy and of doing the test, and the test sensitivity and specificity using formal decision analysis.

—Dan Mayer

See also Bayes's Theorem; Causation and Causal Inference; Clinical Epidemiology; Hill, Austin Bradford; Lind, James; Meta-Analysis; Quantitative Methods in Epidemiology; Sensitivity and Specificity; Study Design

Further Readings

Concato, J., Freinstein, A. R., & Halford, T. S. (1993). The risk of determining risk with multivariate analysis. *Annals of Internal Medicine, 118*, 200–210.

Daly, J. (2005). *Evidence based medicine and the search for a science of clinical care*. Berkeley: University of California Press/Milbank Memorial Fund.

Evidence-Based Medicine Working Group. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association, 268*, 2420–2425.

Fagan, T. J. (1975). Nomogram for Bayes' theorem [Letter]. *New England Journal of Medicine, 293*, 257.

Feinstein, A. R. (1967). *Clinical judgment*. Malabar, FL: Robert E. Krieger.

Hill, A. B. (1947). Statistics in the medical curriculum. *British Medical Journal, 2*, 366–368.

Lau, J., Schmid, C. H., & Chalmers, T. C. (1995). Cumulative meta-analyses of clinical trials builds evidence for exemplary medical care. *Journal of Clinical Epidemiology, 48*, 45–57.

Mayer, D. (2004). *Essential evidence-based medicine.* Cambridge, UK: Cambridge University Press.

Medical Research Council. (1948). Streptomycin treatment of pulmonary tuberculosis. *British Medical Journal, 2*, 769–782.

Olson, C. M. (1994). Understanding and evaluating a meta-analysis. *Academic Emergency Medicine, 1*, 392–398.

Pauker, S. G., & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *New England Journal of Medicine, 302*, 1109–1117.

Paul, J. R. (1966). *Clinical epidemiology* (Rev. ed.). Chicago: University of Chicago Press.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence based medicine: How to practice and teach EBM*. Edinburgh, UK: Churchill Livingstone.

Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V., & Chalmers, T. C. (1987). Meta-analyses of randomized clinical trials. *New England Journal of Medicine, 316*, 450–455.

Sox, H. C., Blatt, M. A., Higgins, M. C., & Marton, K. I. (1988). *Medical decision making*. Boston: Butterworth Heinemann.

Trohler, U. (2000). *To improve the evidence of medicine: The 18th century British origins of a critical approach*. Edinburgh, UK: Royal College of Physicians of Edinburgh.

Upshur, R. E. G. (2002). If not evidence, then what? Or does medicine really need a base. *Journal of Evaluation in Clinical Practice, 8*, 113.

Wennberg, J. E., & Gittelsohn, A. M. (1973). Small area variations in health care delivery. *Science, 182*, 1102–1108.
Woolf, S. H. (1999). The need for perspective in evidence based medicine. *Journal of the American Medical Association, 282*, 2358–2365.
Wulff, H. R., & Gotzsche, P. C. (2000). *Rational diagnosis and treatment: Evidence-based clinical decision making* (3rd ed.). Oxford, UK: Blackwell Science.

Web Sites

Bandolier Journal: http://www.jr2.ox.ac.uk/bandolier.
Cochrane Collaboration: http://www.cochrane.org/index.htm.
Evidence Based Emergency Medicine: http://ebem.org/index.php.
Evidence-Based Medicine Resource Center of the New York Academy of Medicine: http://www.ebmny.org.
InfoPOEMs: http://www.infopoems.com.
James Lind Library: http://www.jameslindlibrary.org/index.html.
Oxford Centre for Evidence-Based Medicine: http://www.cebm.net.
University of Sheffield "Netting the Evidence": http://www.shef.ac.uk/scharr/ir/netting.

Exposure Assessment

Exposure assessment is a critical component of epidemiologic research, and unfortunately for many studies, it is this component that introduces many limitations. Exposure assessment has been defined by Last (2001) as the "process of estimating concentration or intensity, duration, and frequency of exposure to an agent that can affect health" (p. 66). It involves preferably quantitative, but often qualitative, procedures to estimate and assign an individual's past or current exposures. Various types of exposures are studied in epidemiology. Some common examples are summarized in Table 1.

A person's behavior, dietary patterns, smoking history, family history, personal characteristics, and exercise are frequently studied in epidemiologic research as potential exposures, risk factors for exposure, and potential confounders. However, generally the area of exposure assessment in epidemiologic research refers to its application in occupational and environmental epidemiology.

Exposure assessment is also used for purposes beyond epidemiologic research such as risk assessment and regulatory compliance. As an element in risk assessment, exposure assessment has been defined by Jayjock, Lynch, and Nelson (2000) as "the activity that describes the nature and size of the various populations exposed to a chemical agent and the magnitude and duration of their exposures" (p. 26). In regulatory compliance, exposure assessment serves as a quantitative method by which measures of exposure can be compared with established exposure limits. This entry discusses exposure assessment in the

Table 1 Types of Environmental Agents or General Factors to Which One Can Be Exposed

Agent Type	*Examples*
Chemical	Solvents, pesticides, acrylamides, drugs
Metal	Arsenic, nickel, chromium, lead, cadmium
Physical	Radiation, temperature, noise, ergonomic, physical force
Particulate	Dusts, fibers, molds, silica, fumes
Biological	Viruses, fungi, parasites, bacteria, prions
Psychosocial	Violence, stress, social networks
Nutritional	Fiber in diet, meat consumption, vegetable consumption
Behaviors	Smoking, alcohol consumption, exercise
Family history	Family history of specific disease
Personal characteristics	Height, weight, race, sex, body mass index (BMI)

Source: Adapted in part from Rom (1992) and Plog (2002).

context of occupational and environmental epidemiologic studies, although these principles can be applied to all types of epidemiologic research.

Epidemiologic Approaches to Exposure Assessment

Exposure assessment in epidemiology can be broken into various phases that include (1) identification and characterization of the exposed population; (2) evaluation of potential exposure pathways; (3) assessment of the frequency, intensity, and duration of exposure; and (4) exposure classification.

The first issue to consider in exposure assessment is choosing the group that represents the most appropriate and feasible population to study, given the specific research questions. Often, worker populations are selected because of the availability of exposure information and the fact that higher exposures typically occur in workplace settings; however, it may be more appropriate to study specific community members, consumers, or the general population depending on the research question. Most environmental or occupational epidemiologic studies estimate exposure for groups of individuals who are expected to share similar exposure characteristics (e.g., levels, variation).

The second issue to consider is the potential route of exposure, which can include dermal uptake, oral intake via ingestion or mouthing (for children), inhalation, or a combination of pathways. In the case of physical agents (e.g., electric or magnetic fields), simply being present in a field or experiencing a physical force constitutes the route of exposure. For exposure to occur, a "complete" exposure pathway must exist. Examples of incomplete pathways, which prevent exposure to contaminants, include a concrete cap in a foundation that could prevent intrusion of a soil contaminant into a home, or vegetative cover that could reduce release of metal-laden dusts. The amount of agent taken into the body depends on various characteristics such as particle size (for inhalation), bioavailability, human behavior, whether natural barriers or personal protective equipment prevent actual exposure, and whether the substance is transformed within the body.

If complete exposure pathways exist, information on the duration and frequency of exposure can significantly improve a study's ability to assess health risks to a particular agent. Peak exposures, or those occurring over a short time frame, may be important for short-term or intermittent tasks or exposures to some types of chemical or physical agents. For example, a worker injury may result from short-term exposure to a high force. For tasks that are longer in duration or conducted on a more frequent basis, time-weighted averages (TWA) might be more appropriate. For example, for a worker conducting similar tasks on a daily basis over a typical workday, an 8-hr TWA is likely the most appropriate metric. Often the cumulative exposure over a worker's career is estimated as average exposure multiplied by the duration of exposure. When data are available, period-specific variation in average exposure is accounted for in the calculation of cumulative exposure. Similarly, exposures to the general population from agents in the environment might be best characterized by lifetime exposure.

Exposure data can be classified into three categories depending on the detail of information: quantitative, semiquantitative, and qualitative. Quantitative methods for exposure classification include the use of personal or biological monitoring, which collects concentration or dose information for individuals.

Qualitative exposure classification methods often use contemporaneous or historical environmental/industrial hygiene data, surrogates of exposure or exposure proxies, models, questionnaires, personal interviews, or diaries to obtain qualitative data regarding an individual or group exposure. Surrogates of exposure can include historical occupational history records or residence information, chemical inventory records, environmental discharge data, job exposure matrices (which provide estimated job-specific exposure levels based on job titles), use of a certain product, or duration in a job title or geographic location. It is highly recommended that studies using exposure proxy variables include a validation substudy to compare the correlation between the actual exposure and the exposure surrogate values and to assess potential exposure misclassification that may result when relying on proxy measures of exposure.

Finally, direct but qualitative methods can also include collection of personal interview, diary or questionnaire data that specifically query about exposure history/experience. These data may be classifiable only into a "yes/no" category or may be useful in assigning individuals into high, medium, and low categories.

Quantitative methods can be used to estimate specific concentrations or doses, which can be expressed

as average or cumulative exposure levels such as ppm-years or milligram per kilogram (continuous). Groups can be qualitatively classified as ever/never exposed (dichotomous) or as being high/medium/low-exposed (semiquantitative). Either quantitative or qualitative data can be used to place individuals within exposure categories that are then assigned a score (ordinal ranking) representing a range of exposures (e.g., 0 to 5, with 5 being the most highly exposed group).

Exposure Measurements in Occupational or Environmental Settings

Exposure measurement highly depends on the type of exposure, route of exposure, available measurement technology, and practical issues such as access to study subjects and method of data collection (e.g., "expert" observer, self-report). In the case of instrumentation, data can be limited by the "limit of detection" inherent in the method, instrument, or analytical laboratory for quantifying the concentration of a contaminant. In the use of self-report, accuracy can be affected by recall bias, or the ability to remember events accurately.

Measurement of exposure in air can be conducted for fumes, smoke, mists, gases, and vapors that would be expected to enter the breathing zone of workers and inhaled. Table 2 provides a list of the types of air sampling that can be conducted. Air sample results are usually reported in ppm (parts of vapor or gases per million parts of contaminated air by volume at room temperature and pressure) or milligrams or micrograms per cubic meter (milligrams or micrograms of substance per cubic meter of air).

Wipe sampling can be conducted if either dermal or oral exposure from the hand-to-mouth pathway is suspected. Wipe sampling can be conducted by wiping a specified area with an appropriate wipe. Results are usually reported in mass/wipe, which can then be converted to mass/area if the area is known. In general, wipe sampling to assess chemical exposure is not common, but it is most often used to assess lead or other metal contamination on surfaces.

Biological monitoring involves measurement of changes in composition of body fluid, tissue, or

Table 2 Types of Air Sampling

Type	*Description*	*Examples*
Active	Air actively drawn through sampling media; contaminant of interest is deposited	Battery-powered pump draws air through appropriate device
Passive	Natural air movement sufficient for deposition of contaminant on sampling media, or air movement is unnecessary (e.g., radiation badges)	Dosimeter badge adsorbs contaminant
Grab	Small volume of air rapidly drawn through sampling media or into a bag/canister	Stain (colorimetric) detector tube measures contaminant concentration in air (i.e., changes color)
Short-term/ long-term	Air contaminant collected over a variable time period (e.g., during specific task, entire shift)	Radiation badges worn daily and analyzed periodically to assess cumulative exposure
Area	Sampler placed in a single area	Stainless steel canister under vacuum opened to collect air sample
Personal	Sampler placed as close to contaminant's portal of entry as possible (e.g., breathing zone)	A filter cassette attached to pump placed on worker's lapel to assess particulate exposure
Continuous	Direct-reading instrument logs contaminant concentration at specified intervals	Continuous noise data logger logs sound power levels to determine peaks in noise level
Average	One concentration value obtained per measurement period	Device collects contaminant over a work shift; mass/volume calculated to provide time-weighted average concentration

Source: Adapted in part from Plog (2002) and Nieuwenhuijsen (2003).

expired air to establish absorption of a contaminant. For example, lead concentrations in urine and blood are often used to determine lead exposure. Specific biomarkers of chemical or metal exposure will provide more accurate and reliable estimates of historical exposure than nonspecific biomarkers, which can reflect exposures to various agents. In addition, the half-life of the biomarker and the representativeness of samples with respect to a person's entire exposure history should be considered when evaluating biomarker information.

Exposure simulation involves recreating or simulating exposures to certain contaminants during specific activities and is usually performed if it is not feasible to conduct present-day sampling (e.g., products or conditions no longer exist) or if historical data are not adequate or available. Simulation tests may involve conducting an activity suspected of generating a certain contaminant in an enclosed chamber with a known ventilation rate.

Models are often used when sampling or simulation is impractical or when exposure has already occurred but no sampling data are available. Plume modeling is an example of a type of modeling that predicts the spread of a contaminant through a community when there has been a release.

Exposure Classification, Summarization, and Analysis

The final phases of the exposure assessment process involve (1) assessment of the accuracy of exposure assignment and (2) the summarization and analysis of the exposure data. Exposure misclassification can occur in two forms: differential and nondifferential. With differential misclassification, one group in the study (e.g., cases or those exposed) has a higher probability of misclassification than another group (e.g., controls or those not exposed). Nondifferential misclassification, on the other hand, occurs when there is equal probability of exposure misclassification in the different study groups. In general, exposure misclassification of either type will reduce study power.

The impacts of exposure misclassification depend on the type of misclassification and whether exposure is measured as a dichotomous, ordinal, or continuous type variable. Nondifferential exposure misclassification with a dichotomous exposure variable will lead to bias in the relative risk estimate. The direction of this bias will be toward the null value. For other exposure conditions (i.e., ordinal or continuous exposure variables), this "rule of thumb" does not necessarily apply, although the assumption of nondifferential exposure misclassification and bias toward the null is frequently invoked, but rarely are data provided to verify this assumption. The impact of differential misclassification is much less predictable and dependent on the specific situation in each study.

The final aspect of exposure assessment is the manner in which exposure data are summarized. Various strategies are commonly used when some form of continuous data are available. Continuous exposure data can be modeled in its original form in epidemiologic analyses. Such data can be used to classify the exposure group as those from the top percentile category of exposure (e.g., 90th or 75th percentile) or can used to define the exposed group based on a biological/clinical/physical basis (i.e., if exposure exceeds a certain threshold or the safety level). Another aspect in the classification of exposure is the consideration of the latency period or the time between exposure to the agent and onset of disease. This is especially relevant in cancer epidemiology. If it is assumed, based on biological and clinical considerations, that a disease has a minimal latency period (e.g., 10 years), then exposure occurring within this period prior to disease onset will not be relevant. Exposure windows, another time-dependent variable in exposure assessment, refer to specific points in time for which a person is classified for exposure status. Some examples of exposure windows include exposure during fetal development, adolescence, and other periods of life.

Exposure assessment methods are critical in determining overall study quality in assessing health risks. The need to accurately reconstruct historical exposures makes this one of the most challenging aspects of epidemiologic research. Continually advancing technologies in exposure biomarkers to evaluate current and past exposures, geographical information systems for visualization of proximity to contaminant sources, and statistical modeling to predict likely exposure concentrations and outcomes should help to improve future exposure assessment.

—Fionna S. Mowat, Mona Shum, and Michael A. Kelsh

See also Biomarkers; Environmental and Occupational Epidemiology; Lead; Pollution

Further Readings

Armstrong, B. K., White, E., & Saracci, R. (1992). *Principles of exposure measurement in epidemiology.* New York: Oxford University Press.

Byrd, D. M., & Cothern, C. R. (2000). *Introduction to risk analysis: A systematic approach to science-based decision making.* Rockville, MD: Government Institutes.

Hawkins, N. C., Norwood, S. K., & Rock, J. C. (Eds.). (1991). *A Strategy for occupational exposure assessment.* Fairfax, VA: American Industrial Hygiene Association Press.

Hulka, B. S., Wilcosky, T. C., & Griffith, J. D. (1990). *Biological markers in epidemiology.* New York: Oxford University Press.

Jayjock, M. A., Lynch, J. R., & Nelson, D. I. (2000). *Risk assessment principles for the industrial hygienist.* Fairfax, VA: American Industrial Hygiene Association Press.

Last, J. M. (Ed.). (2001). *A dictionary of epidemiology* (4th ed.). New York: Oxford University Press.

National Research Council. (1983). *Risk assessment in the federal government: Managing the process.* Washington, DC: National Academy Press.

Nieuwehnhuijsen, M. J. (Ed.). (2003). *Exposure assessment in occupational and environmental epidemiology.* New York: Oxford University Press.

Plog, B. A. (Ed.). (2002). *Fundamentals of industrial hygiene* (4th ed.). Chicago: National Safety Council.

Rom, W. N. (Ed.). (1992). *Environmental and occupational medicine* (2nd ed.). Boston: Little, Brown.

Smith, T. J. (2002). Issues in exposure and dose assessment for epidemiology and risk assessment. *Human and Ecological Risk Assessment*, *8*, 1267–1293.

F

Factor Analysis

Not all research is focused on hypothesis testing. Sometimes researchers are interested in identifying the structure of a particular phenomenon. For example, suppose a team of researchers were interested in developing a tool that would adequately measure and reflect the concerns of women considering undergoing genetic testing for familial breast cancer. A literature review indicated to these researchers that the structure of this construct called *concern* had been identified and described for other populations (e.g., adult caregivers of cancer patients) but needed to be redefined for their population of interest: women at risk for familial breast cancer. The methods of factor analysis can be used in developing such an instrument.

Characteristics of Factor Analysis

Factor analysis is not a single statistical method. Rather, it involves a complex array of statistical procedures that provide a way to identify interrelationships among a large set of observed variables. Much subjectivity and artistry are involved in this technique. The goal of factor analysis is to arrive at a parsimonious set of factors that have common characteristics and that summarize and describe the structural interrelationships among a set of identified items in a concise and understandable way. A *factor* is a cluster of related, observed variables that represent the underlying dimension of a construct that is as distinct as possible from other factors in the solution.

Exploratory Versus Confirmatory Factor Analysis

Factor analysis can be used for theory and instrument development and for assessing construct validity of an established instrument when administered to a specific population. There are two basic forms of factor analysis: exploratory and confirmatory. In exploratory factor analysis (EFA), the researcher does not know initially how many factors are necessary to explain the interrelationships among a set of characteristics. EFA is used, therefore, to explore the underlying dimensions of a construct. It is available in a number of statistical computer packages, including SPSS and SAS.

In contrast, confirmatory factor analysis (CFA) is used to assess the extent to which a hypothesized organization of identified factors fits the data. It is used when the researcher has prior knowledge about the underlying structure of the construct under investigation. CFA could also be used to test the utility of the dimensions of a construct identified through EFA, to compare factor structures across studies, and to test hypotheses concerning the structural relationships among a set of factors associated with a specific theory or model. To undertake CFA analyses, the researcher needs to use a structural equation modeling program, such as LISREL, AMOS, or EQS. Because CFA is addressed elsewhere (see the entry "Structural

Equation Modeling''), the focus of this discussion will be on EFA and, specifically, its use in instrument development.

Assumptions of Exploratory Factor Analysis

A basic assumption of EFA is that within a collection of observed variables, there exists a set of underlying factors, smaller in number than the observed variables, that can explain the interrelationships among those variables. Because the initial steps of EFA are performed using Pearson product moment correlations, many of the assumptions associated with this correlation coefficient are applicable to factor analysis (e.g., large sample sizes, continuous distributions and sufficient variation within the items, and linear relationships among the correlated variables). Since the response categories for each individual item are often constructed using dichotomous *yes, no* responses or ordinal-level Likert scales, normality of distributions is not always a strict requirement.

Sample Size Requirements

EFA is not a technique to be used with small sample sizes. Typically, it is expected that there will be 10 to 15 subjects for each item that is initially being considered. Comrey and Lee indicate that a sample of 50 is very poor, 200 is fair, 500 is very good, and 1,000 or more is excellent. Realistically, the number of available subjects may be restricted especially when the researcher is examining disease entities in which there are small numbers in the population. For example, it may be very difficult to identify and recruit 1,000 subjects who are considering undergoing genetic testing for cancer.

The Process of Exploratory Factor Analysis

There are eight basic steps to EFA. The first two, specifying the problem and generating the items, are undertaken prior to data analysis and are major determinants of a successful factor analysis. The following is a brief overview of these eight steps. For greater detail on the process for undertaking a factor analysis, the reader is referred to the suggested readings.

1. Specifying the Problem

Often, researchers think that they know the dimensions of the construct they want to measure—until they begin to specify the problem and to generate items related to that construct. During this initial phase, the observed indicators of the construct of interest need to be conceptualized and operationally defined. *What is it, exactly, that the instrument will measure? Are there other constructs that are related to this construct? Will the instrument measure the construct broadly or a specific aspect of the construct?* These questions must be addressed before undertaking a factor analysis. Without careful conceptualization, the resulting instrument will likely have poor construct validity.

2. Generating the Items

The next step is to generate items or empirical indicators that accurately reflect the construct of interest. These indicators need to be organized with a meaningful format that will allow data to be collected effectively and efficiently. Pilot testing of the instrument, its format, design, and layout with respondents who are similar to the target group will help set the stage for a successful factor analysis.

3. Evaluating the Adequacy of the Correlation Matrix

Once the data have been collected, the variation in item responses should be carefully examined. Without adequate variation, the interitem correlations will be low and the utility of conducting a factor analysis questionable. The correlation matrix also needs to be evaluated to determine whether there are adequate correlations among the items to justify a factor analysis. Several approaches are available in the statistical packages to assess the initial factorability of a correlation matrix. These include an evaluation of the determinant, Bartlett's test of sphericity, and the Kaiser-Meyer-Olkin test.

4. Extracting the Initial Factors

Assuming that the correlation matrix is factorable, the task of the beginning extraction process is to determine the number of initial factors that appear to represent the dimensions of the construct being measured. This extraction process begins with an

initial estimate of the total amount of variance, or *communality*, in each individual item that can be explained by the extracted factors. These communality estimates can range from 0 to 1.0 with higher values indicating that the extracted factors explain more of the variance in an individual item.

Initially, these communalities are unknown; they cannot be identified until after a factor analysis has been run. Yet to begin a factor analysis solution, estimated communalities need to be placed on the diagonal of the correlation matrix. There are two basic approaches to estimating these communalities: principal components analysis (PCA) and common factor analysis.

Principal Components Analysis (PCA)

PCA assumes that potentially all the item variance can be explained and that there is as much variance to be analyzed as there are observed variables. Since each item, if standardized, has a mean of 0 and variance of 1.00, the initial estimate of communality for each item is 1.00. This is what appears on the diagonal of the correlation matrix when using PCA.

Common Factor Analysis

Common factor analysis assumes that only the common variance that an item shares with other items in the correlation matrix can be explained by a small number of underlying factors. The variation that is unique to the item including error variance cannot be explained. Because common factor analysis focuses on the common variance shared among the items, the amount of variance that can be extracted from the correlation matrix by these estimated factors is less than 100%. Therefore, initial values less than 1.00 will appear on the diagonal of the correlation matrix. One common solution in common factor analysis, principal axis factoring (PAF), places the squared multiple correlation (R^2) of each item with all other included items on the diagonal. These squared multiple correlation coefficients (range: 0 to 1.0) provide an initial indication of the strength of the interitem relationships.

Eigenvectors and Eigenvalues

Whether using PCA or common factor analysis, each extracted factor is represented by an *eigenvector*. An eigenvector is a column of weights, each weight of which is associated with an item in the correlation matrix. If there are eight items in the matrix, there will be eight weights in that eigenvector. These weights are called *factor loadings* and represent the correlation of each item with the given extracted factor. Table 1 presents an initial factor-loading matrix generated in SPSS for Windows from an 8×8 correlation matrix using PAF. Item c6, for example, has an initial factor loading of .712 with Factor 1.

The *eigenvalue* associated with an extracted factor is equal to the sum of its squared factor loadings and represents the amount of variance in the items that can be explained by that particular factor (Table 1). Dividing the eigenvalue by the number of items indicates the proportion of total item variance accounted for by a given factor. For both PCA and common factor analysis, the eigenvalues for each extracted factor are typically largest for the first factor extracted and lowest for the last extracted factor. Table 2 presents a computer-generated output using SPSS for Windows for the total variance explained for the same 8×8 matrix using PAF. The initial eigenvalues presented in Table 2 are similar to those that would be obtained in a PCA. However, because only common variance is being extracted in PAF, the extracted eigenvalues are smaller.

Extracting the Factors

Extracting factors from a correlation matrix is an iterative procedure that consists of repeatedly refining the factor analysis solution until suitable eigenvectors and their associated eigenvalues are obtained. The factor extraction process differs depending on the type of factor analysis undertaken (e.g., PCA or PAF). Typically, the majority of the variance in the items is accounted for by a relatively small number of factors. In Table 2, for example, 48.968% of the shared variance in the eight items can be accounted for by two extracted factors generated in PAF. If PCA had been used, 57.349% of the total variance would have been accounted for by the same two factors.

Selecting the Number of Initial Factors to Retain

The goal of a factor analysis is to reduce the number of factors such that the maximum amount of variance can be explained with the fewest number of factors.

While the goal is simple, there is no precise solution to determining the number of factors to extract.

Table 1 Factor-Loading Matrix Generated in SPSS for Windows From an 8 × 8 Correlation Matrix Using Principal Axis Factoring

Factor Matrix[a]

	Factor						
	1[b]	*2*	*3*	*4*	*5*	*6*	*7*
c6 Worry about diagnosis I can't do anything about	.712	–.152	.283	–.254	.064	.181	–.012
c4 Help reduce uncertainty about future	.691	–.160	–.432	–.014	.080	–.080	–.074
c2 Worry about uncertain diagnosis	.662	–.353	–.159	–.153	–.294	.026	.060
c7 Hope to make better health, lifestyle choices	.622	.577	.008	–.079	.112	–.165	.088
c3 What to do to manage risk	.607	.448	.236	.096	–.219	–.068	–.085
c5 Fear ambiguity of results	.512	–.512	.204	.106	.264	–.105	–.003
c1 Increase of personal control	.374	.428	–.184	.134	.157	.243	.005
c8 Worried about loss of health and life insurance coverage	.378	–.202	.047	.471	–.099	.044	.042

Extraction Method: Principal Axis Factoring

a. Seven factors extracted; 21 iterations required.

b. Extracted eigenvalue #1 = $(.712)^2 + (.691)^2 + (.662)^2 + \cdots + (.374)^2 + (.378)^2 = 2.725$

Table 2 SPSS for Windows Computer-Generated Output for the Total Variance Explained in Principal Components Analysis

	Initial Eigenvalues			*Extraction Sums of Squared Loadings*			*Rotation Sums of Squared Loadings*		
Factor	*Total*	*of Variance*	*Cumulative*	*Total*	*of Variance*	*Cumulative*	*Total*	*of Variance*	*Cumulative*
1	3.046[a]	38.074	38.074	2.725[c]	34.057	34.057	1.336	16.703	16.703
2	1.542	19.275	57.349	1.193	14.911	48.968[d]	1.221	15.257	31.960
3	.866	10.819	68.168	.425	5.318	54.286	1.059	13.239	45.199
4	.762	9.524	77.692	.355	4.437	58.723	.640	8.001	53.200
5	.626	7.828	85.521	.262	3.272	61.995	.622	7.772	60.972
6	.499	6.233	91.754	.144	1.795	63.790	.205	2.567	63.539
7	.354	4.425	96.180	.026	.324	64.114	.046	.575	64.114
8	.306[b]	3.820	100.000						

Extraction Method: Principal Axis Factoring

a. Initial eigenvalue #1 = 3.046

b. Initial eigenvalue #8 = .306

c. Extracted eigenvalue #1 = 2.725

d. Percentage extracted = 48.968

Given the same data set, a team of researchers might arrive at very different solutions. Several guidelines can be used to help determine the number of factors to retain: the amount of explained individual and cumulative variance, the scree plot, and factor interpretability and usefulness.

Amount and Percent of Individual and Cumulative Variance Extracted. One solution to determining the number of initial factors is to select only those factors for which the eigenvalue is > 1.0. This means only those factors will be retained that explain more than one item's worth of variance. In Table 2, this criterion would result in two factors being extracted. A second solution is called the 5% criteria: Only those factors would be retained that explain at least 5% of the variance in the items. Three factors meet that criterion with the PAF solution. A third approach is to establish a threshold for maximum variance extracted (e.g., 65%). Because only shared variance is evaluated in PAF, none of the seven extracted factors meet this 65% criterion (Table 2).

Examining the Scree Plot. A scree plot is a plot of the initial eigenvalues in descending order of magnitude. To use a scree plot for determining how many factors to retain, a straight line is then drawn through the lower values of the plotted eigenvalues. This allows the researcher to identify distinct breaks between the steep slope of the larger eigenvalues and the trailing off of the smaller ones. Figure 1 presents a scree plot of the eight initial eigenvalues shown in Table 1; two factors appear on the steep slope distinct from the slope of the smaller values and would be retained using this criterion.

Factor Interpretability and Usefulness. There is no easy solution to deciding on the number of factors to retain. This decision is based on a careful evaluation of the statistical indicators and the factors' theoretical coherence. The ultimate criteria for determining the number of factors are factor interpretability and usefulness both during the initial extraction procedures and after the factors have been rotated to achieve more clarity.

5. Rotating the Factors

Unrotated factor solutions often do not provide meaningful clusters of items. Factor rotation improves the interpretation of the generated factors. Factor rotation is the process of turning the reference axes of the factors about their origin to achieve a simple structure and theoretically more meaningful factor solution. *Simple structure* is one in which, ideally, each item has a high loading on one factor only and each factor has high, meaningful loadings for only some of the items. There are two broad classes of rotation, *orthogonal* and

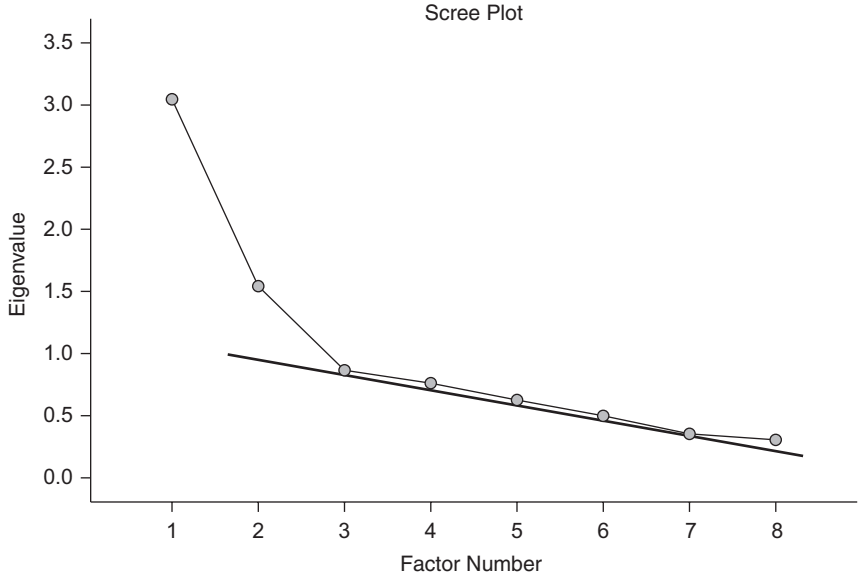


Figure 1 Scree Plot of the Initial Eigenvalues (Table 1) Plotted Against Their Factors

oblique. Each has different underlying assumptions but share a common goal of simple structure.

Orthogonal Rotation

In an orthogonal rotation, it is assumed that the generated factors are independent of each other (i.e., they are uncorrelated). The factors are fixed at 90° angles to one another and rotated to obtain an optimal fit with the distributions of the items in space. Because they are maintained at right angles to one another, these newly rotated factors are uncorrelated and, as a result, produce a single rotated factor loading matrix. The correlations among the items have not changed. Only the loadings of the items on the newly rotated orthogonal factors have changed.

Varimax is the most commonly used orthogonal rotation. This approach maximizes the variances of the loadings within the factors while maximizing differences between the high and low loadings on a particular factor (hence the name *varimax*). Higher loadings on a factor are made higher and lower loadings are made lower. Table 3 presents a two-factor unrotated and rotated solution of the factor-loading matrix in Table 2 using a PAF varimax rotation. The factor loadings have been sorted according to size for ease of interpretation. The higher factor loadings from the unrotated solution have been made higher while the low loadings are now lower in the rotated solution.

Oblique Rotation

Although orthogonal rotations often produce attractive simple solutions, these rotations rest on the critical assumption that the factors are uncorrelated.

Table 3 Unrotated and Rotated Two-Factor Solution: PAF With Varimax Rotation

Factor Matrix[a]		
	Factor	
	1	*2*
c6 Worry about diagnosis I can't do anything about	.673	.164
c7 Hope to make better health, lifestyle choices	.658	–.585
c4 Help reduce uncertainty about future	.644	.161
c2 Worry about uncertain diagnosis	.635	.354
c3 What to do to manage risk	.588	–.365
c5 Fear ambiguity of results	.483	.475
c8 Worried about loss of health and life insurance coverage	.350	.191
c1 Increase of personal control	.368	–.369

Extraction Method: Principal Axis Factoring

Rotated Factor Matrix[b]		
	Factor	
	1	*2*
c2 Worry about uncertain diagnosis	.714	.132
c5 Fear ambiguity of results	.675	–.058
c6 Worry about diagnosis I can't do anything about	.623	.303
c4 Help reduce uncertainty about future	.599	.287
c8 Worried about loss of health and life insurance coverage	.391	.076
c7 Hope to make better health, lifestyle choices	.133	.870
c3 What to do to manage risk	.220	.656
c1 Increase of personal control	.048	.519

Extraction Method: Principal Axis Factoring

Rotation Method: Varimax With Kaiser Normalization

a. Two factors extracted; 18 iterations required.

b. Rotation converged in three iterations.

This assumption is rarely met in epidemiological research since researchers are often dealing with conceptually different but nevertheless correlated dimensions of a construct.

One popular oblique rotation is *Oblimin*, which uses the parameter *delta* to control the degree of correlation among the factors. Each original factor is rotated separately by different amounts. As a result, two different factor-loading matrices are generated: a *factor pattern matrix* (a matrix of loadings that are like partial standardized regression coefficients) and a *factor structure matrix* (a matrix of simple correlations of the items with the factors).

There are differences of opinion as to which matrix, factor pattern or factor structure, should be evaluated for simple structure. Some writers argue that the *factor pattern* matrix should be used to determine the extent to which simple structure has been achieved since the difference between high and low loadings in the factor pattern matrix is more apparent than in the structure matrix. Other writers maintain that the factor structure matrix is more stable because it is unaffected by changing the size of the correlations among the factors.

The author suggests that both matrices be examined to determine which makes the most sense. Table 4 presents the two-factor structure and factor pattern matrices generated using the same data as in Table 3. This was a PAF solution using an Oblimin rotation. The items were sorted according to size, and absolute values $< .30$ were suppressed. This presentation allows the pattern of high and low item-to-factor correlations to become more distinct. The weaker item-to-factor correlations are not deleted; they remain in the background and are included in calculations of item communalities and percentage of variance explained by the factors. Their absence from the table just makes it easier to view the patterns of factor loadings. Table 4, for example, indicates that the factor pattern matrix has generated loadings that are more easily interpreted. The correlation between the two factors was $-.328$.

An advantage of oblique rotations is that the factor axes can be rotated to any position without changing the correlations among the items or their shared communalities. Kline suggests that correlated factors reflect the real world. If it is found after oblique rotation that the factors are orthogonal, the researcher can be confident that the result is not an artifact of the choice of rotation.

6. Refining the Solution

In most factor analyses, several factors will likely emerge as potential descriptors of a set of items. Ideally, each item will load strongly on a single factor following factor rotation. In reality, even with factor rotation, items will sometimes demonstrate weak loadings on all factors or will load strongly on several factors. To refine the obtained solution, both the strength of the item loadings with the factors and their consistency with the original conceptualizations before the factor analysis began need to be closely examined.

The factor pattern matrix in Table 4 indicates that five items load on the first factor and three items load on the second factor. The loadings that remain in view are strong, ranging in values from $-.882$ for Item c7 to .396 for Item c8. A similar pattern of distribution of items was also observed when a varimax rotation was generated (Table 3). The only difference was that Item c6 had loadings $> .30$ on both factors. The challenge will be to determine where best to place this item if the varimax solution is chosen.

When items load strongly on more than one factor, it is best to either delete the item entirely or place it with the factor that it is most closely related to conceptually. Reliability coefficients (e.g., Cronbach's alphas) for the group of items that load on a given factor can be used to evaluate the factor's internal consistency and to decide where to best place an item with strong loadings on several factors.

7. Interpreting and Naming the Factors

Naming factors is a poetic, theoretical, and inductive effort that should follow from the theoretical considerations that have led to the definition of the construct. There are no definitive statistical tests in factor analysis to indicate whether an item is *significant* for the purposes of factor interpretation. Usually, three or four items with the highest loadings on a factor are selected and studied. *Is there a theme or common element that these items share?* If so, then a descriptive name might be selected that is representative of these items. When the highest loadings on a factor are low (e.g., $< .60$), the researcher is faced with potentially weak interpretations.

In selecting a name for the factor, it is best that the interpretation remain simple but suggestive as to what dimension that factor represents. If the items for the factor analysis were theory driven, the researcher should return to that theory for guidance in naming

Table 4 Factor Pattern and Structure Matrices Generated From a Rotated Two-Factor PAF Solution With Oblimin Rotation: Loadings Sorted by Size and Values < .30 Suppressed

Structure Matrix

	Factor	
	1	*2*
c2 Worry about uncertain diagnosis	.727	
c6 Worry about diagnosis I can't do anything about	.666	–.397
c5 Fear ambiguity of results	.654	
c4 Help reduce uncertainty about future	.640	–.377
c8 Worried about loss of health and life insurance coverage	.398	
c7 Hope to make better health lifestyle choices		–.880
c3 What to do to manage risk	.332	–.682
c1 Increase of personal control		–.520

Extraction Method: Principal Axis Factoring
Rotation Method: Oblimin With Kaiser Normalization

Pattern Matrix[a]

	Factor	
	1	*2*
c2 Worry about uncertain diagnosis	.725	
c5 Fear ambiguity of results	.715	
c6 Worry about diagnosis I can't do anything about	.601	
c4 Help reduce uncertainty about future	.578	
c8 Worried about loss of health, and life insurance coverage	.396	
c7 Hope to make better health lifestyle choices		–.882
c3 What to do to manage risk		–.643
c1 Increase of personal control		–.532

Extraction Method: Principal Axis Factoring
Rotation Method: Oblimin With Kaiser Normalization

Factor Correlation Matrix

Factor	*1*	*2*
1	1.000	–.328
2	–.328	1.000

Extraction Method: Principal Axis Factoring
Rotation Method: Oblimin With Kaiser Normalization

a. Rotation converged in five iterations.

the factors. *How useful is the factor? Is the content too broad to be of use, or is it too specific and limited in scope? How does this identified construct fit with other identified taxonomies in the field? What is missing from this construct? In what areas do we need to direct our future factor analysis activities?* Clearly, the naming of a factor is not a simple task.

8. Reporting and Replicating the Results

The key to a successful factor analysis is the reporting and replication of the results obtained. Enough information must be reported about the development of the instrument to allow other researchers to verify the results. Although limited by the constraints of

a research journal, a report of a factor analysis should include at the very least the following information:

- a theoretical rationale for the use of factor analysis;
- detailed descriptions of the development of the items, sampling methods, and participants;
- an evaluation of the assumptions of factor analysis;
- justification for the choice of factor extraction and rotation methods;
- an evaluation of the correlation matrix;
- decisions regarding cutoffs for meaningful factor loadings;
- presentation of the structure and pattern matrices;
- descriptions and interpretation of the factors;
- report of the internal consistency of the identified factors; and
- assessment of the study limitations and suggestions for future research.

Using factor analysis involves an ongoing commitment to a research program. One must not assume that all the items that define a factor have been delineated in a single study. Many studies must be undertaken to determine if all items of the factor have been derived and correctly interpreted.

—Marjorie A. Pett

See also Likert Scale; Measurement; Pearson Correlation Coefficient; Reliability; Structural Equation Modeling; Validity

Further Readings

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis.* Hillsdale, NJ: Lawrence Erlbaum.

DeVellis, R. F. (2003). *Scale development: Theory and applications.* Thousand Oaks, CA: Sage.

Green, S. B., & Salkind, N. J. (2007). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Kline, P. (1994). *An easy guide to factor analysis.* New York: Routledge.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis.* Thousand Oaks, CA: Sage.

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use* (2nd ed.). Oxford, UK: Oxford University Press.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* Boston: Allyn & Bacon.

Family Studies in Genetics

Family studies may be considered a key entry point for research into the role of inherited genetic variation in disease. There are three major kinds of family studies: (1) evaluations of the extent to which a disease or other trait of interest aggregates or clusters within families, and how it is transmitted across the generations; (2) linkage analyses aimed at pinpointing the specific location on one of the chromosomes of a gene containing a mutation that has a major or moderate effect on disease risk; and (3) association studies aimed at finding common gene variants that have smaller but still medically important effects on disease severity or disease incidence. All three of these kinds of family studies are considered in this entry.

Analysis of Aggregation or Clustering

One of the first questions that investigators need to ask when considering genetic studies of a disease or some other trait of interest is "What is the evidence that inherited genetic variation has an important influence on the trait?" A necessary but not sufficient condition required to demonstrate the importance of genetic variation is the occurrence of *familial aggregation* of the trait. We know from the simple rules of Mendelian inheritance that family members tend to share genes in common. For example, siblings share 50% of their genes inherited identical by descent from their parents, and cousins share 12.5% of genes inherited from their common grandparents. Therefore, if genetic variation really is important for the development of a disease (incidence) or its severity, then we would expect to find the disease co-occurring or "clustering" among family members more often than among randomly drawn unrelated individuals in the population. One way this is often formally tested in research studies is to compare the frequency of disease in relatives of persons with the disease compared with the frequency found in relatives of matched healthy controls. If disease frequency is not at all elevated in relatives of cases, then it is unlikely that most cases of the disease in the population have a substantial genetic basis. On the other hand, familial aggregation of disease is

evidence that the trait has a genetic basis but still does not constitute definitive proof. This is because in addition to sharing genes, family members also usually share similar environments (diet, exposure to toxins, etc.), and it is possible that the aggregation of disease in relatives is caused by their common environments rather than by shared genes. Investigators can measure environmental exposures that are suspected to be risks for the disease in cases, in controls, and in their relatives and attempt to statistically adjust for these effects in data analyses. Alternatively, if the disease is sufficiently common, studies of monozygotic and dizygotic twin families offer a very powerful design that can provide very powerful capability to distinguish between environmental and genetic causes of variation in disease risk.

For some forms of very severe single-gene (Mendelian) disorders there may not be a positive family history if transmission is dominant and clinical symptoms onset at a young age. People with such diseases are unlikely to reproduce so patients with the disease frequently have arisen via a new mutation not present in their ancestors. On the other hand, with recessive diseases, the parents, not surprisingly, usually do not know that they are carriers for the recessive mutation and the disorder may be new to the family. An exception to this rule is for recessive diseases occurring among cultures with consanguinity (e.g., first-cousin marriages) where occurrence of the disease may not be surprising. Studies of consanguineous families can be very useful for gene-mapping studies.

It should be noted, however, that even if no evidence of familial aggregation or heritability is obtained from family studies, this does not rule out the possibility that a small subset of disease cases (e.g., 1% to 5%) might be caused by a mutation in a gene that causes a major increase in disease risk. In fact, strong familial aggregation may exist for this small genetically based subset of cases, but this is obscured by the majority (95% to 99%) of disease cases for which genetic variation has little or no influence on disease risk. For example, most cases of breast cancer lack familial aggregation, and in twin studies there is little evidence of heritability, but relatively rare mutations in BRCA1 gene and other genes have a very major effect on cancer risk in individuals who inherit these mutations. Furthermore, sometimes the same genes that are involved in the rare inherited forms of a disease are mutated somatically in nonhereditary cases. Therefore, understanding the biological mechanisms involved in rare hereditary forms may have great importance for developing improved methods of diagnosis, prognosis, and therapy for both hereditary and nonhereditary forms of the disease.

When a disorder shows familial aggregation that appears not attributable to shared environmental exposures, a statistical method called *segregation analysis* can be used in an attempt to estimate the mode of inheritance—autosomal recessive, dominant, or codominant; X-linked dominant or recessive. This technique has been successfully applied to many simple (single gene) disorders, but it has only limited value in studies of complex diseases where multiple disease susceptibility genes interact to influence disease risk. Segregation analyses usually need to assume *homogeneity*, meaning that the same type of gene is responsible for causing the disease in all families included in the study. If, in fact, some families have inherited a gene that acts dominantly while other families in the data set have inherited mutations at either the same gene or a different gene where risk is recessively transmitted, segregation analysis will be unreliable. Even with relatively simple disorders, the method has serious limitations. First, one must be aware of and appropriately adjust for the way the families and family members were selected for study (ascertainment bias). Second, there is the problem of unrecognized shared environments (noted above), and, for quantitative traits, deviations from assumptions of normality can lead to incorrect inferences about the mode of transmission. The method has been modified in recent years in an attempt to address these weaknesses, but it has nonetheless been largely supplanted in genetic epidemiological research by family studies that incorporate DNA markers.

Linkage Analysis

When a single gene has a major effect (e.g., $>$ 10-fold increase) on the risk of developing a disease, and when the disease is relatively uncommon in the population, then the method of linkage analysis can rapidly lead to successful gene identification. *Linkage-mapping* families are evaluated for the cosegregation of polymorphic DNA markers (either short tandem repeats or single-nucleotide polymorphisms) with the disease phenotype. Linkage mapping depends on the fact that recombination during meiosis occurs only rarely between markers that are located physically close (linked) to the disease gene. Recombination occurs increasingly more often as markers are located

farther away (> 10 Mbp) on the same chromosome as the disease gene or located on a different chromosome altogether. It is possible to detect linkage with as few as 12 to 16 informative individuals when the disorder is highly penetrant (i.e., nearly every person who inherits the mutation gets the disease), when there are few phenocopies (i.e., hardly anyone who does not inherit the mutation gets the disease), and if the density of markers is sufficiently high. Most often, a single sufficiently large extended family is not available, so several unrelated families may need to be combined, especially when attempting to map a recessive trait. In some special circumstances, consanguineous (inbred) families may allow investigators to use an approach called *homozygosity mapping* to localize a recessive disease gene. For example, in offspring of first cousins, about one sixteenth of the genome is expected to be homozygous. The specific homozygous genome regions would be random in affected offspring of different sets of first cousins, except for the region that contains their recessive disease gene. This region would be homozygous in the offspring of all the offspring, so by evaluating only a limited number of such offspring, a disease gene can be mapped. Linkage analysis has had some success for oligogenic diseases (i.e., those with only a few genes involved). Unfortunately, despite major investments of resources and years of effort, studies of complex disorders that are likely to involve multiple genes of smaller effect (e.g., twofold increase in risk) and potentially involving gene-gene and gene-environment interactions have usually been disappointing.

Association Analysis

Mathematical analyses and several recent disease studies have shown that *association mapping* methods can provide good statistical power for identifying genes that underlie complex diseases while requiring much smaller numbers of patients and their relatives than would be required for linkage analysis. Association mapping can be performed using either small or large families or unrelated cases and controls. The only catch is that instead of needing only a few hundred polymorphic DNA markers to cover the entire human genome, the association strategy requires several hundred thousand such markers assayed on each subject. Fortunately, molecular genetic technologies have been developed that can meet the challenge of producing these massive amounts of data, and the International HapMap Project ("HapMap" being an abbreviation of "haplotype map") has cataloged this variation in several human populations and made it freely available online for researchers wishing to tap into this rich genomic treasure chest. In the first phase of this project, the frequency of DNA variants was measured at more than 1 million locations distributed across the human genome in European, African, and Asian subjects. DNA variants that are located near each other on the chromosome often are correlated with each other, so if an investigator determines the DNA sequence for a subject at one position, the DNA bases at the neighboring variant positions often can be predicted with a high degree of confidence. This phenomenon is known as *linkage disequilibrium*, and the HapMap Project has determined where these patterns of correlation among neighboring DNA variants exist for a large portion of the human genome. Armed with this information, investigators interested in studying inherited variation at a candidate gene for their disease (or searching through all genes in the entire genome) need not undertake the large effort of conducting assays for all known variants in their clinical subjects. Instead, they can use computer algorithms on data derived from the HapMap Project to measure most of the inherited variation present in the genome at a substantially reduced cost by identifying an optimized subset of DNA variants that serve as statistical "tags" for many other variants that are not actually assayed in the laboratory. The National Institutes of Health and other organizations responsible for support of biomedical research are currently developing plans for a major expansion of whole-genome association studies to a wide range of diseases and to drug side effects and therapeutic responses (pharmacogenomics). Family studies are certain to play an important role for these exciting initiatives in the future of genomic medicine.

Ethical Issues

There are many important and complex ethical issues that arise when performing family studies. Members of families need to be carefully educated about the risks, both social and cultural, of participating in family studies, which include providing information on their relatives. Although investigators will have obtained approval from an institutional review board responsible for protecting research subjects, such review boards generally focus on possible outcomes from the genetic (biological) information that will be obtained and may

not always fully consider the possibility of altered family dynamics that may arise as a consequence of participating in the study. Anticipatory counseling of prospective families can enhance participation rates and minimize stressful effects on family dynamics.

—Scott R. Diehl and Robert P. Erickson

See also Gene-Environment Interaction; Genetic Disorders; Genetic Epidemiology; Genetic Markers; Genomics

Further Readings

Arar, N. H., Hazuda, H., Steinbach, R., Arar, M., & Abboud, H. (2005). Ethical issues associated with conducting genetic family studies of complex disease. *Annals of Epidemiology, 15*(9), 712–719.

Elston, R. C., & Spence, A. M. (2006). Advances in statistical human genetics over the last 25 years. *Statistics in Medicine*, *25*, 3049–3080.

Morton, N. E. (2006). Fifty years of genetic epidemiology, with special reference to Japan. *Journal of Human Genetics, 51*, 269–277.

Savage, S. A., & Chanock, S. J. (2006). Genetic association studies in cancer: Good, bad or no longer ugly? *Human Genomics*, *2*, 415–421.

Farr, William
(1807–1883)

William Farr is one of the major figures in the history of epidemiology. A British physician with an unusual knowledge of statistics, Farr was appointed Compiler of Abstracts at the General Register Office of England and Wales, which registers births, marriages, and deaths. He worked almost 40 years in analyzing statistics from England and Wales and pioneered the quantitative study of morbidity and mortality.

Farr developed a classification of causes of death, constructed the first English life table, and made major contributions to occupational epidemiology, comparing mortality in specific occupations with that of the general population. In a report presented in 1864, Farr addressed the disproportionate high number of deaths among miners in Cornwall, showing that at each age level, the rate of mortality attributed to pulmonary diseases among miners was much higher than among males exclusive of miners, with the difference being higher at higher ages. He concluded that pulmonary diseases were the chief cause of the high mortality rate among the miners. From the fact that excess mortality from pulmonary diseases reached its maximum after mid-age, when mine conditions had had sufficient time to produce their effect on the health of miners, Farr concluded that it might be confidently inferred that these diseases were due to labor conditions inside mines.

Being a conscious reformer, Farr opposed the gloomy Malthusian views then in fashion. Against the idea that population grows geometrically while food can grow only arithmetically, he argued that human inventiveness can increase productivity, and, moreover, that plants and animals that constitute food also grow geometrically. Against Malthus's idea that men reproduce like rabbits, without concern for consequences, Farr showed with statistics that the average age at marriage in England was 24 to 25 years, about 8 years after the onset of the reproductive age of women, and that more than 20% of men and women reaching reproductive age never married.

As the statistician in charge of analyzing mortality data, Farr argued in an official report that hunger was responsible for many more deaths than shown in the statistics, since its effects were generally manifested indirectly in the production of diseases of various kinds. Although he was a supporter of the miasmatic theory of disease and initially claimed that cholera was transmitted by polluted air, Farr was finally persuaded otherwise by John Snow, and in 1866 produced a monograph showing how cholera cases were much more frequent in Londoners receiving water from particular sources.

Being fluent in French, German, and Italian, Farr represented Britain in a number of statistical congresses, and in his later years was considered a major authority on medical statistics and public health. Today, he is considered one of the most prominent figures of the movement of social medicine in Victorian England and a major author in the history of health statistics.

—José A. Tapia Granados

See also Public Health, History of; Snow, John

Further Readings

Anonymous. (1989). Mortality of miners: A selection from the reports and writings of William Farr. In C. Buck, A. Llopis, E. Nájera, & M. Terris (Eds.), *The challenge of epidemiology: Issues and selected readings* (pp. 67–71). Washington, DC, Pan American Health Organization (PAHO/WHO).

Eyler, J. M. (2002). Constructing vital statistics: Thomas Rowe Edmonds and William Farr, 1835–1845. *Sozial- und Praeventivmedizin*, *47*, 6–13.

Hamlin, C. (1995). Could you starve to death in England in 1839? The Chadwick-Farr controversy and the loss of the "social" in public health. *American Journal of Public Health, 85*(6), 856–866.

Fertility, Measures of

In demographic terms, fertility is the actual reproductive performance of members of a population. Fertility is one of the three major components of demography, along with mortality and migration. Fecundity, on the other hand, reflects the physiological ability of a woman to reproduce (often referred to as fertility in lay conversation).

Several measures of fertility exist, of which the most frequently used include birthrate, age-specific fertility rate (ASFR), and total fertility rate (TFR).

Birthrate

Birthrates, also known as crude birthrates, represent the actual number of live births per 1,000 population per year. It is mathematically represented by the following equation:

$$\text{birthrate} = \frac{n}{p} 1{,}000,$$

where n represents the number of live births in that year and p is the population size in that year.

Similarly, the general fertility rate is the number of live births per 1,000 women of reproductive age (usually ages 15 to 44 or 15 to 49) in a given year.

Age-Specific Fertility Rate

The ASFR is a measure of the average number of births per year per 1,000 women in specific age groups (generally ages 15 to 19, 20 to 24, etc.). The following equation represents the calculation of ASFR:

$$\text{ASFR}_i = \frac{n_i}{w_i} 1{,}000,$$

where n_i represents the number of live births during a given year by women in age group i, and w_i represents the number of women in age group i.

Total Fertility Rate

The TFR measures the average number of children who would be born alive to a woman completing her childbearing years based on ASFRs in a given year. Thus, TFR is not an actual rate of live births but an estimate based on the assumption that women will continue childbearing at the ASFR in that year. Mathematically, TFR is calculated from ASFR:

$$\text{TFR} = \frac{\sum \text{ASFR}_i \times x}{1{,}000},$$

where ASFR_i represents the fertility rate for the ith age group and x is the number of years in the interval of age group i.

TFR is often referred to in the context of replacement-level fertility, the TFR necessary to sustain current population size. In industrialized nations, a TFR of 2.1 indicates a level of fertility where each couple bears only enough children to replace themselves in a population, representing equilibrium between birth and death rates. Replacement-level fertility in developing countries is less defined, particularly due to varying death rates in the context of HIV/AIDS.

Other Measures

Other measures of fertility include the following:

- *Completed Fertility Rate.* The actual number of children born to a cohort of women at the end of their childbearing years.
- *Gross Reproduction Rate.* The average number of live female births that would be born during a woman's lifetime if she passed through her reproductive years conforming to ASRF in a year.
- *Marital Fertility Rate.* The actual number of live births to married women per 1,000 women of reproductive age in a year.

—Anu Manchikanti

See also Demography; Mortality Rates; Population Pyramid

Further Readings

Population Reference Bureau. (2004). *Glossary of population terms*. Retrieved August 15, 2006, from http://www.prb.org/Content/NavigationMenu/PRB/PRB_Library/Glossary2/Glossary.htm.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes.* Oxford, UK: Blackwell.

Fetal Death, Measures of

The World Health Organization defines a fetal death as the death of a production of conception prior to complete expulsion or extraction from its mother, regardless of length of pregnancy. Factors that indicate a fetal death after separation from the mother include no breathing or evidence of life by the fetus, pulsation of umbilical cord, or voluntary muscle movement. The distinction between fetal and infant deaths is based on the place of death; a fetal death occurs in utero.

Fetal death can be measured in terms of time of death: death prior to labor, antepartum mortality, and death during labor. According to the National Center for Health Statistics, measures are operationalized in terms of gestational age: A death at or after 20 weeks is considered a fetal death. An early fetal death is at 20 to 27 weeks, while a late fetal death is at or greater than 28 weeks.

In the United States, fetal death reporting requirements vary by state. The 1992 Revision of Model State Vital Statistics Act and Regulations recommends that each fetal death of weight 350 g or more or, for unknown weight, 20 weeks of gestation or greater should be reported. However, the standards are not consistent. Thirteen areas (including states, cities, and territories) specify the requirements for fetal death as stated in the above act. Eleven areas report any expulsion or extraction of the product of human conception as a fetal death, while others base the requirements only on gestational age or use a different combination of gestational age and weight.

—Anu Manchikanti

See also Maternal and Child Health Epidemiology; National Maternal and Infant Health Survey

Further Readings

Kowaleski, J. (1997). *State definitions and reporting requirements for live births, fetal deaths, and induced terminations of pregnancy*. Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. (2005). *Fetal deaths*. Retrieved August 15, 2006, from http://www.cdc.gov/nchs/datawh/nchsdefs/fetaldeath.htm.

Field Epidemiology

The term *field epidemiology* describes the application of epidemiologic methods, usually by representatives of a public health service, to address specific health-related questions in community settings. Many, if not most, field investigations address urgent or acute health problems. However, many others are conducted as planned studies addressing less urgent needs. All field investigations have in common the aim of conducting scientifically rigorous studies, sometimes under difficult circumstances, to answer very specific epidemiologic questions with the ultimate aim of planning, implementing, or evaluating public health interventions.

Outbreak investigations are a prominent example of field epidemiology studies and serve to illustrate many of the characteristics common to all field investigations. An outbreak investigation, like other field investigations, has the following goals: (a) to determine the cause and etiology of the disease, (b) to limit the spread and severity of illness of the disease, and (c) to prevent future outbreaks. In addition, investigations of this sort can serve to identify new modes of transmission of illnesses, identify new pathogens, and monitor the effectiveness of prevention activities.

Field epidemiology also includes investigations conducted in several different types of community settings, each involving its own challenges. These settings may include health care facilities, child care settings, occupational settings, and even areas affected by natural disasters.

Investigations in health care settings, such as hospitals, rehabilitation centers, transitional care centers, outpatient settings, and long-term care facilities, differ from other field investigations in several ways. First of all, most of the infections encountered are endemic to the setting and are unlikely to be completely eradicated. Patients affected in these settings typically have additional medical conditions that render them more susceptible to adverse sequelae than healthier community-based individuals. The infectious agents involved in health-care-related outbreaks are a greater danger to the patients and health care workers in the

facility than to the general public outside the facility. Finally, an outbreak in a health care facility may require even more rapid identification and control than a community-based problem, not only because the risk of litigation is higher but also because of the greater vulnerability of the patients to the infection. These distinctions aside, the methods used to identify and control an outbreak in a health care setting are basically the same as for a community-based outbreak.

Child care settings present issues similar to those in other care facilities, though the "at-risk" population is not confined to the facility itself. Epidemiologists need to be aware of the potential spread of illness beyond the children and child care providers to family contacts and children encountered in other settings such as schools, community playgrounds, and the like.

Field epidemiology may also be conducted in areas affected by disasters, either natural or man-made. Such investigations typically involve, at least initially, the gathering and summarizing of data about the needs of the affected population: What food is available or needed? Are water sources available? What medical conditions and/or injuries are prominent or likely to occur, and what medications, treatments, and medical personnel are needed? Early data such as these can be used to guide humanitarian and relief efforts, matching available resources to the needs. In addition to assessing the needs of the populations, epidemiologists may be called on to assist in preventing further adverse health effects, to evaluate the effectiveness of relief programs in meeting the most immediate needs of the population, and to assist in contingency planning. Epidemiologists may also need to quickly establish surveillance systems, monitoring ongoing health risks and adverse events, evaluating the effectiveness of clinical interventions, and identifying potential risk factors for developing adverse outcomes. Opportunities exist as well for studying the natural history of a disaster and its potential for long-term health impacts. In the setting of a natural disaster, field epidemiologists may serve as a communication and coordination hub for medical personnel, relief effort coordinators, and other decision makers, making skills in communication, coordination, and public relations key.

Epidemiologists may be called to investigate illnesses or injuries in a workplace setting as well. There may be instances of a previously recognized work-related illness or injury occurring in a given occupational setting. Or, there may be illnesses or injuries occurring that haven't yet been identified as being work related. Epidemiologists may be asked to investigate whether a particular process or exposure causes the illness in question. Or, as in the other field investigation types described above, the investigation may involve the evaluation of intervention or prevention efforts.

Field investigations in occupational settings present several other challenges that distinguish them from field investigations in other settings. In the other types of investigations, the investigator usually has some idea what he or she is looking for with regard to the disease involved. The pathogen involved, the epidemiologic characteristics of the disease, and the clinical and laboratory methods are usually known or narrowed down to a few likely suspects. In addition, the stakeholders are generally cooperative and supportive of the investigative efforts. However, an investigation in an occupational setting may involve looking for a completely unknown agent or mechanism causing the illness or injury. Potential agents may be part of the production process, may be a contaminant, or may be the product itself. And the illnesses or injuries may be taking place in a setting where the epidemiologist has no prior knowledge or detailed understanding of the various manufacturing processes, the internal politics, and the various stakeholders' allegiances. In addition, there are unique legal issues involved dictating the circumstances and extent to which employees may be interviewed or examined and special considerations with regard to confidentiality issues. An epidemiologist may have to simultaneously deal with union representatives, management and ownership representatives, legal counsel, and the media while attempting an epidemiologic investigation in a workplace setting. Certain stakeholders may be supportive of the investigation, while others may be obstructive or uncooperative, further complicating the efforts to identify and minimize the impact of the illness or injury process necessitating the investigation.

Regardless of the precise setting of the field investigation, each has in common with the others many methods. Overall, each study has as its aim to rigorously collect valid data from which hypotheses about the cause of the illness may be made. Analytic methods are then used to test the hypotheses. Data collection can be accomplished using any of several methods. For example, data from written records such as physician or hospital notes or employer records may be abstracted onto standardized data collection

forms. Questionnaires may be developed and used to collect uniform information during subject interviews. Physical exams of affected persons and/or of a group of comparison subjects may be conducted. Biological samples may be gathered from physical exams or from existing samples, and environmental samples may also be collected. The analytic portion of the investigation may be conducted using either case-control or cohort study methods. Case-control methods are especially relevant if the investigator wishes to evaluate several potential exposures and their associations with a single outcome and are often used in infectious disease investigations. Case-control studies may also be nested within larger cohort studies in situations where conducting a case-control study with the whole population is not feasible. Cohort studies are more common in occupational investigations, where a group of workers may be followed over time to evaluate the effect of a particular exposure on subsequent health events. In all cases, the cohort of subjects to be studied or the source population of cases needs to be as clearly defined as possible, and a clear case definition should be formulated to identify affected persons as readily as possible.

—Annette L. Adams

See also Disaster Epidemiology; Environmental and Occupational Epidemiology; Epidemic; Notifiable Disease; Outbreak Investigation

Further Readings

Gregg, M. B. (Ed.). (2002). *Field epidemiology* (2nd ed.). New York: Oxford University Press.

Hartge, P., & Cahill, J. (1998). Field methods in epidemiology. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 163–180). Philadelphia: Lippincott-Raven.

Firearms

Injuries and deaths due to firearms are an important public health concern in the United States. The Centers for Disease Control and Prevention (CDC) identifies the need for improvements in four areas that have a particular focus on firearm use: suicide, homicide, physical assault, and weapon carrying among youth. While firearm-related deaths and injuries have decreased since peaking in the late 1980s, they are still a significant problem in the United States, especially among low-income African Americans living in urban areas.

Firearm Mortality

Suicide

According to the CDC's Web-based Injury Statistics Query and Reporting System (WISQARS) (2005), suicide is the fourth leading cause of death for individuals aged 10 to 14 years, the third leading cause of death for individuals aged 15 to 24 years, and the second leading cause for individuals aged 25 to 34 years, with nearly half of the suicides among 25- to 34-year-olds being completed with a firearm. Most firearm suicides by youth occur at home, and the presence of a firearm in the home is associated with an increased risk of firearm suicide. The use of a firearm for a suicide attempt results in death approximately 80% of the time.

Firearm suicide is most prevalent among males, particularly white males. However, an exception to this is among males aged 20 to 29 years. In this group, the rate of firearm suicide is highest for black males.

Homicide

Homicide is the third leading cause of death for individuals aged 10 to 14 years, the second leading cause of death for individuals aged 15 to 24 years, and the third leading cause for individuals aged 25 to 34 years, with approximately 80% of the homicides among 25- to 34-year-olds being completed with a firearm (CDC/WISQARS, 2005).

Homicide caused by firearm is most prevalent among adolescents and young adults, males, and blacks (see Figures 1 and 2). Homicide, driven primarily from death due to a firearm, grew to epidemic proportions in the late 1980s through the early to mid-1990s in the United States. Since the early to mid-1990s, homicide due to firearm has declined.

However, from 1999 through 2003, firearm homicides have stabilized, showing, if anything, a slight increase in the early 2000s. Furthermore, the changes in firearm homicides have not affected all demographic groups in the same way. For example, from 1999 to 2001, the 10- to 14-year-old males and females in each CDC race category showed a decline

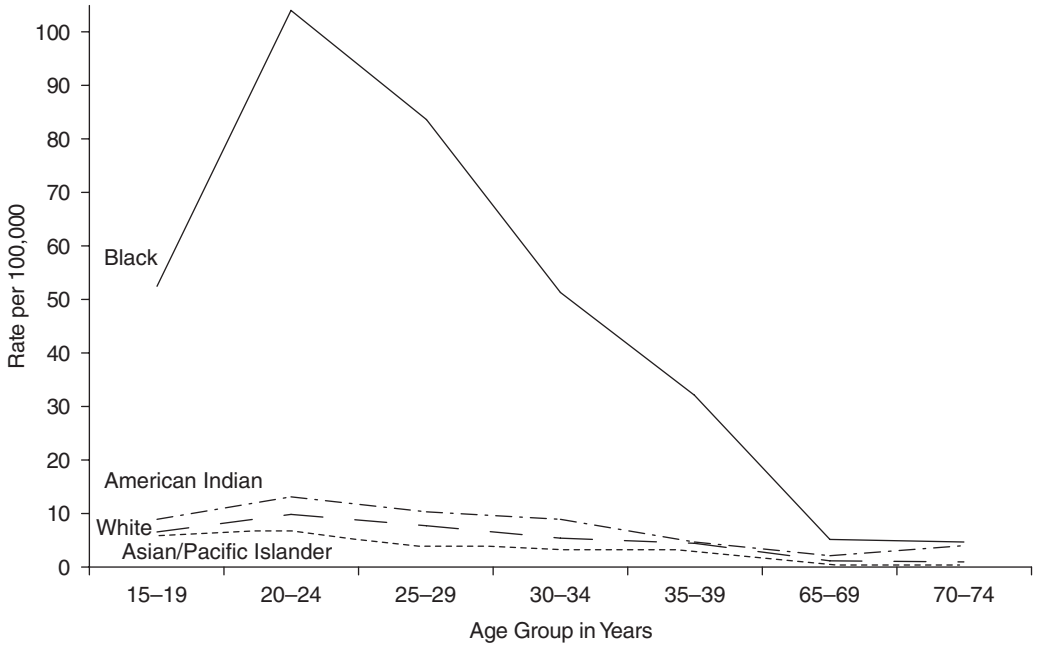


Figure 1 Average Firearm Homicide Rate (2001–2004) by Race and Selected Age Groups: Males

Source: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control (2005).

in firearm homicides from 1999 to 2001, followed by a small but steady increase in homicides over the next 2 years. A similar pattern is seen among youth between the ages of 15 and 19 years across racial groups. However, both black and white males between the ages of 20 and 25 years have shown little change in this time period, while white females and black males aged 25 to 29 years have shown a steady increase in firearm homicide between 1999 and 2003.

Firearm Morbidity

Similar to firearm mortality, firearm morbidity is most prevalent among adolescents and young adults, males, and blacks. Firearms are responsible for a considerable amount of nonfatal injury among individuals between the ages of 10 and 29. The rate of nonfatal injury due to firearms among males increases from approximately 11 per 100,000 among males aged between 10 and 14 years to approximately 101 per 100,000 when the male is aged between 15 and 19 years. This risk peaks at approximately 139 per 100,000 among males between 20 and 24 years and then steadily drops as males get older. Adolescent females experience a similar pattern with an increase in a firearm-related injury risk of 3 per 100,000 at ages 10 to 13 per 100,000 at age 24.

Possession of Firearms Among Youth

Juveniles are more likely to posses a firearm than adults, with handguns being the most popular firearms. Nationally, approximately 6% of high school students reported bringing a gun to school in the past month. For males alone, this increases to more than 10% among both white and black males (*Morbidity and Mortality Weekly Report*, Youth Risk Surveillance Survey, 2004). Most firearms owned in the United States were manufactured here. A large proportion of guns are obtained through illegal purchase, making estimates of gun possession prevalence challenging.

Firearm Surveillance

To understand the prevalence of firearm-related morbidity and mortality, the three primary public health

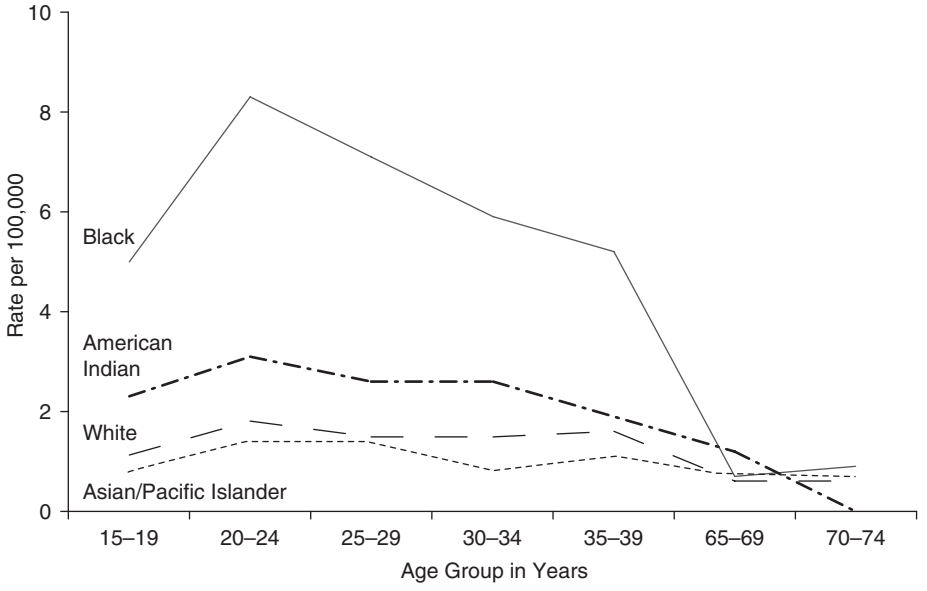


Figure 2 Average Firearm Homicide Rate (2001–2004) by Race and Selected Age Groups: Females

Source: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control (2005).

sources used are emergency room (ER)-treated and –released data, hospital admissions discharge date, and death data from medical examiner reports. Each of these types of surveillances has its strengths and weaknesses. Hospital data, both ER and discharge, use the International Classification of Disease external cause of injury codes (E-codes). Intended originally and currently used for billing purposes, E-codes have been adapted over the years to provide increasing detail regarding firearm injuries including the location, the type of bullet wound, and the intentionality of the injury. The utility of E-codes in providing an improved national picture of the extent of intentional and unintentional injuries has been recently strengthened as more hospitals and regions have created systems to aggregate E-code information.

However, there are three important limitations of hospital data as a source of information about morbidity due to firearms. First, as stated above, E-codes are created for billing purposes and not by researchers or for research purposes. This potentially threatens the validity of these codes when used for firearm research. For example, if an individual receives an additional injury that requires more serious medical attention, the more expensive treatment may only be noted or if the cause of injury is excluded on the chart by a physician who is more intent on describing the injury and treatment than the cause. Second, not all geographical areas and not all hospitals participate in either the collection or inclusion of E-code data for larger statewide and national purposes. Third, in most cases, injury related to firearms results in death, so hospital discharge data are not a useful source for these types of injuries.

Firearm-related mortality surveillance is done through death record review. Death records often provide detailed information that may include the type of firearm used, circumstances of death, and victim and offender information. However, death records are created by either a medical examiner (typically in larger areas) or a coroner (typically in smaller areas). As a result, disparities in detail and accuracy exist in records from different regions. Furthermore, data from detailed death records are difficult to obtain. What is more commonly used are vital record data that have been abstracted from death records. While vital records are more easily obtainable, they may provide less detail about the firearm-related death compared with the full death record.

Another source of firearm-related morbidity and mortality is criminal justice data. The Federal Bureau of Investigation and the Bureau of Justice Statistics

provide national information about firearm-related crimes based on arrests and victimization reports. Similar to hospital data, the consistency and accuracy of firearm data from these sources may vary across police precinct and victim, ultimately impeding a full understanding of firearm-related crime. Recently, the public health and criminal justice systems have collaborated to improve surveillance of violent deaths nationally. The National Violent Death Reporting System provides a means of linking data between medical examiners, coroners, police precincts, and crime labs to provide a better picture of the extent and risk factors surrounding violent deaths.

—Eve Waltermaurer

See also Injury Epidemiology; Suicide; Violence as a Public Health Issue

Further Readings

Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. (2005). *Web-based injury statistics query and reporting system (WISQARS)*. Retrieved May 10, 2006, from www.cdc.gov/ncipc/wisqars.

Ikeda, R., Mercy, J., & Teret, S. (1998). Firearm-related injury surveillance [Special issue]. *American Journal of Preventative Medicine*, *15*(3S).

Serpi, T. L., Wiersema, B., Hackman, H., Ortega, L., Jacquemin, B. J., Weintraub, K. S., et al. (2005). Homicide and suicide rates. National violent death reporting system, six states, 2003. *Morbidity and Mortality Report*, *54*(15), 377–380.

Fisher's Exact Test

Fisher's exact test (FET) is a nonparametric version of the chi-square test. The most common use of FET is with small data sets, in particular when at least one cell in a cross-tabulation table has an expected frequency of less than 5. Small data sets and sparse cells may render the results reached by a chi-square test invalid, because it is based on asymptotic normality. Because the FET is a nonparametric test, this assumption does not apply to it. The FET calculates the exact probability (p value) of observing the table containing the data or *more extreme* distribution of the data. This probability is calculated by looking at all possible rearrangements of the table (in the direction of the alternative). The row and column totals are held constant when determining all possible *more extreme* rearrangements. Instructions on how to calculate the FET with 2×2 tables are given below. If one has another table of other dimensions, it is best to use a computer package such as SAS.

Finding the Exact Probability of a 2×2 Table

For a 2×2 table, the probability of the table occurring is

$$\frac{\binom{r_1}{b}\binom{r_2}{d}}{\binom{n}{s_2}} = \frac{r_1!r_2!s_1!s_2!}{n!a!b!c!d!}.$$

This probability can be calculated using either combinatorial functions (the left-hand side of the equation) or factorials (the right-hand side of the equation). In the first expression, the numerator counts the number of ways that the first row can occur (expressed by the combinatorial function describing how many ways the b-cell total can be chosen from the Row 1 total) times the number of ways that the second row can occur (expressed as the combinatorial function describing how many ways the d-cell total can be chosen from the Row 2 total) divided by the number of ways that the row and cell totals can be determined with n items (expressed by the combinatorial functions describing how many ways the first column total can be chosen from the sample size). In the second expression, the relevant factorials are simply multiplied together.

2×2 Example

Table 2 contains a 2×2 table of categorical data, representing the number of times two groups of people answered "yes" to a question. The null hypothesis

Table 1 General 2×2 Table

	Column 1	*Column 2*	*Total*
Row 1	a	b	r_1
Row 2	c	d	r_2
Total	s_1	s_2	n

Table 2 Original Data

	Group 1	*Group 2*	*Total*
Yes	13	3	16
No	20	28	48
Total	33	31	64

is that the proportion of *yes* is the same for the two column groups and the alternative hypothesis is that the proportion of *yes* is greater for Group 1 than for Group 2. This is a directional alternative hypothesis.

The exact probability of Table 2, given our assumptions, is

$$\frac{\binom{16}{3}\binom{48}{28}}{\binom{64}{31}} = 0.005274.$$

However, what we want to calculate is the probability of getting a result at least as extreme as this distribution, if the null hypothesis is true. Therefore, we add to this probability the probability of *more extreme* tables under the alternative hypothesis; *more extreme* tables would have the (1, 1) cell—that is, cell "*a*" in Table 2—equal to 14, 15, or 16. These *more extreme* tables are given in Table 3 with their respective probabilities. Note that these tables are found by keeping the row and column totals the same.

The exact p value is $.005274 + .000780 + .000066 + .000002 = .006122$, so at a significance level of 5%, we would reject H_0. So we conclude from these calculations that the proportion of people answering *yes* to our question is significantly different between Group 1 and Group 2.

Table 3 More Extreme Tables With Probabilities

Table With (1,1) Cell = 14		*Probability*
14	2	.000780
19	29	
Table With (1,1) Cell = 15		
15	1	.000066
18	30	
Table With (1,1) Cell = 16		
16	0	.000002
17	31	

Nondirectional Alternative Hypothesis

Although FET can be used for a nondirectional hypothesis test, statisticians do not entirely agree as to how to calculate probabilities in this case. The p value is still defined as the probability of *more extreme* tables, but the definition of "more extreme" may vary. One method is to reverse the a and b values, find the probability of the table, increase the (1, 1) cell to find another table, find the probability of this table, and continue this method until one finds a table with a probability greater than the probability of the original table. Finally, one sums these probabilities to the other probabilities. Table 4 shows *more extreme* tables in the opposite direction of the early example. The table with the (1, 1) cell = 4 has a probability greater than the original table; therefore, the p value for a nondirectional hypothesis is $.002303 + .006122 = .008425$.

—Marjorie E. Bond

Table 4 More Extreme Tables With Probabilities, Nondirectional

Table With (1,1) Cell = 3		*Probability*
3	13	.002303
30	18	
Table With (1,1) Cell = 4		
4	12	.011821
29	19	

See also Chi-Square Test; Nonparametric Statistics

Further Readings

Samuels, M., & Witmer, J. (2003). *Statistics for the life sciences* (pp. 422–427). Englewood Cliffs, NJ: Pearson Prentice Hall.

Food and Drug Administration

The Food and Drug Administration (FDA) is a regulatory agency charged with enforcing laws pertaining to food, drugs, drink, cosmetics, and medical and therapeutic devices sold in the United States. Now within the Department of Health and Human Services, the FDA is responsible for ensuring the safety and efficacy of drugs in the health care system. Throughout its 100 years of existence, the agency has monitored, and responded to, crises that could have been avoided had the agency been given more funding or broader powers. Following almost every instance of catastrophe, usually with attendant deaths, more accountability has been given to the agency.

In its earliest years, the agency now known as the FDA was concerned with drug regulation. Various measures have been taken to keep the citizenry safe, including early laws requiring customs officials to keep adulterated drugs out of the country. The Bureau of Chemistry, founded in 1863, within the U.S. Department of Agriculture, focused more on the dangers inherent in food items than drugs, and even here, the concern about food items ran toward issues of adulteration more than foodborne illness.

The first time attention was paid to a specific drug problem was in 1902, with the passage of the Biologics Control Act. This act, which ensured the safety of serums, vaccines, and other products, was in direct response to tragedies in St. Louis, Missouri, and Camden, New Jersey. In both cities, children died of tetanus after receiving diphtheria antitoxin made from a horse infected with tetanus. The Biologics Control Act, which ultimately led to the Center for Biologics Evaluation and Research, gave the Bureau the authority to regulate biological products and ensure their safety.

The 1906 Pure Food and Drug Act addressed concerns raised by, variously, Upton Sinclair's novel *The Jungle* and Samuel Hopkins Adams's six-part series on drug adulteration in *Collier's* magazine. After almost 20 years of agitation, these two documents helped push Congress to create a new agency, the FDA, charged with enforcement of both the Pure Food and Drug Act and the Meat Inspection Act. Drug companies helped conceive of the regulations that bound them, 11 drugs were made illegal, and all product labels were required to be accurate and truthful.

Continued problems with the drug and food supply were highlighted twice in the 1930s. At the 1933 Century of Progress World's Fair in Chicago, the FDA's exhibit space focused on the shortcomings and inadequacies of the 1906 acts. In 1935, 107 people died after using Elixir of Sulfanilamide, a sulfa drug suspended in glycol ether, a deadly combination. Congress responded by passing the 1938 Food, Drug and Cosmetic Act that extended FDA powers to include cosmetics and therapeutic devices, as well as advertising of some of these products.

During World War II, new drug discoveries, as well as supply shortages, kept the agency busy. The 1941 Insulin Amendment required the FDA to test and certify the purity and potency of insulin. Since penicillin's availability had been limited during wartime to government rather than civilian use only, the batches were monitored. In 1945, the Penicillin Amendment did for penicillin what the Insulin Amendment had done for insulin: The FDA tested and certified penicillin, to guarantee its efficacy and safety. This process was discontinued in the 1970s, when the safety of the penicillin supply was unquestionable.

In the postwar years, the FDA responded to and enforced new drug and food issues, sometimes carefully and sometimes not at all. The Durham-Humphrey Amendment (1951) delineated which drugs could and could not be used without medical supervision; those that could not were available by prescription only. The designation of prescription versus over-the-counter was a joint effort of many parties. What was not being tracked was the development of the polio vaccine. After large successful test studies in 1954, more than a million children were vaccinated against polio. When more than 200 children became sick and 11 died, researchers discovered that the product from Cutter Laboratories contained live poliovirus. The trials were halted and resumed only after factory inspection and ways to test for vaccine safety were put in place.

The 1960s was a period of averting near disaster time and time again. Thalidomide, used in Europe and being tested for consideration in the United States, was kept off the domestic market through the vigilance of a medical safety officer. Following this scare, the American public, as it had in the past, pushed Congress for more regulations and greater safety. The Kefauver-Harris Drug Amendments of 1962 required drug manufacturers to demonstrate the safety and efficacy of their drugs before marketing them. Never before had there been this level of safety. A backlash against the unprotected nature of the drug market, outrage from the American public helped the

FDA expand their role and presence in the drug regulatory process. In 1966, the FDA expanded its commitment to drug effectiveness: the National Academy of Sciences and the National Research Council began to evaluate more than 4,000 drugs approved by the FDA between 1938 (the Food, Drug and Cosmetic Act) and 1962 (the Kefauver-Harris Amendments). Two years later, with the results of the study on drug effectiveness in hand, the FDA formed an organization to implement the recommendations.

Much attention was paid to drugs and drug regulations in the 1970s, from the first patient package inserts (in oral contraceptives) to the Comprehensive Drug Abuse Prevention and Control Act that recategorized drugs by addictive properties rather than therapeutic value. The Environmental Protection Agency took over some of the FDA's regulatory duties even as the FDA gained new ones with the establishment of the National Center for Toxicological Research, examining the impact and effect of chemicals on the environment. In 1972, the National Institutes of Health transferred the regulation of biologics (serums, vaccines, blood products) to the FDA. The 1976 Medical Device Amendment sharpened language, requiring all companies with medical devices and diagnostic tests to demonstrate effectiveness and safety of their products; in addition, quality control and registration with the FDA were required. Saccharine proved to be a source of continual debate in the 1970s: Initially removed from the Generally Recognized as Safe list in 1972, it was given continued life by a congressional fiat in 1977 under the proviso that products bear a warning noting that it had been found to cause cancer in laboratory animals.

In reaction to the Tylenol scare in 1982, in which several packages were tampered with and death from cyanide resulted, Congress passed tamper-resistant packing regulations. Unaware of what was to become the major focus of drug research in the mid- to late 1980s, and confident in the power of biomedicine to find cures for diseases, Congress passed the Orphan Drug Act in 1982. This allowed the FDA to encourage and support drug research for rare diseases. Three years later, the FDA approved the first AIDS blood test, primarily to prevent patients from receiving blood from infected donors. Also, in response to the AIDS crisis, the FDA changed drug trial protocols for those with serious diseases for which there were no medical alternatives. In 1991, Congress continued to support medical research for certain categories of disease by implementing an accelerated drug review process.

The end of the 20th century witnessed the consolidation and reappraisal of some FDA policies and procedures. The Board of Tea Tasters, impaneled in 1897, was disbanded in 1996. In a cost-saving measure, the agency outlined a 15-year plan in 1998 to streamline the number of laboratories. In response to 9/11, the Public Health Security and Bioterrorism Preparedness and Reponse Act of 2002 called for a coordinated and improved plan in response to public health emergencies. Continuing this public health thrust, the Project BioShield Act of 2004 gave the FDA the power to distribute treatments quickly in response to chemical, biological, or nuclear agents used in a terrorist attack.

—Gwen Kay

See also National Institutes of Health; Polio; Public Health, History of; Thalidomide; Vaccination

Further Readings

Hilts, P. (2003). *Protecting America's health: The FDA, business and one hundred years of regulation*. New York: Knopf.

Young, J. H. (1989). *Pure food: Securing the Federal Food and Drugs Act of 1906*. Princeton, NJ: Princeton University Press.

Foodborne Diseases

Foodborne diseases are caused by agents that enter the body through the consumption of food or beverages. Many cases of foodborne illness are not reported, so it is impossible to get an exact count of the incidence of foodborne illness; however, the Centers for Disease Control and Prevention (CDC) estimates that about 76 million cases of foodborne disease occur in the United States each year, resulting in 325,000 hospitalizations and 5,000 deaths. The incidence of foodborne disease is believed to be much higher in developing countries, where the World Health Organization estimates that in 2000 alone, 2.1 million people died from diarrheal diseases. Outbreaks of foodborne illness occur in both developing and industrialized countries and can affect large numbers of people. For instance, an outbreak of hepatitis

A in China, caused by the consumption of contaminated clams, affected more than 300,000 individuals, and an outbreak of salmonellosis in the United States in 1994, caused by consumption of contaminated ice cream, affected 224,000 people.

Most cases of foodborne disease are caused by microorganisms, including bacteria, viruses, and parasites. Other agents that can cause foodborne disease include mycotoxins, marine biotoxins, and the toxins occurring in poisonous mushrooms; metals such as lead, mercury, and cadium, which may contaminate food through air, water, or soil pollution; organic pollutants such as dioxin and polychlorinated biphenyls, which are by-products of some industrial processes; and other agents such as the agent causing BSE (bovine spongiform encephalopathy, also known as ''mad cow disease''), which appears to be transmissible through the consumption of tainted beef.

Common Foodborne Diseases Caused by Microorganisms

Salmonellosis is caused by the *Salmonella* bacteria, which is commonly found in the intestines of mammals, reptiles, and birds and is usually spread to humans through consumption of foods of animal origin, including eggs, meat, and milk. Symptoms of salmonellosis include fever, headache, nausea, vomiting, abdominal pain, and diarrhea; in persons with weakened immune systems or poor health, it can be life threatening. The CDC estimates that 1.4 million cases of salmonellosis occur in the United States annually, with approximately 500 fatal cases.

Campylobacteriosis is caused by *Campylobacter* bacteria. In some countries, campylobacteriosis is more common than salmonellosis, and worldwide it is the most commonly identified bacterial cause of diarrheal illness. Campylobacteriosis is transmitted mainly through drinking water, undercooked poultry, and raw milk; because *Campylobacter* bacteria live in the intestines of healthy birds, most raw poultry can be assumed to be contaminated with it. The symptoms of campylobacteriosis include fever, nausea, severe abdominal pain, and diarrhea; major health consequences may develop in 2% to 10% of cases, including neurological disorders and reactive arthritis. The CDC estimates that there are more than 1 million cases of campylobacteriosis annually in the United States, with approximately 100 fatal cases.

Escherichia coli is a type of bacteria living in the intestines of many animals, including humans and cattle. Most strains of *E. coli* are not harmful to humans; an exception is *E. coli* 0157:H7, which lives in the intestines of cattle and can have serious health effects when ingested by humans. This type of *E. coli* is usually ingested in undercooked ground beef, although it may also be transmitted through unpasteurized milk and fruit juice, contaminated water, uncooked produce, and person-to-person contact. Symptoms of *E. coli* 0157:H7 poisoning include severe abdominal cramps and bloody diarrhea; in 3% to 5% of the cases, hemolytic uremic syndrome may develop, which can result in kidney failure and death. The CDC estimates that about 73,000 people in the United States became ill from *E. coli* in 1999 and about 60 died.

Listeriosis, caused by the bacteria *Listeria monocytogenes*, is most often transmitted through milk, soft cheeses and ice cream, raw vegetables, and raw meat and poultry. Because the *L. monocytogenes* bacteria can grow at low temperatures, foods that are refrigerated for long periods of time are particularly likely routes of transmission. Symptoms of listeriosis include nausea, vomiting, diarrhea, and flu-like symptoms. Listeriosis is particularly dangerous for pregnant women, because it can cause abortion and stillbirth, and in infants and persons with a weakened immune system, because it can lead to meningitis and septicemia (blood poisoning). The CDC estimates that about 2,500 people in the United States become ill with listeriosis annually and about 500 die.

Prevention and Control

Major foodborne diseases, including salmonellosis and *E. coli* 0157:H7, are reportable diseases in most states, meaning that infections caused by those agents must be reported to the health department. However, because most cases of foodborne disease are mild and never diagnosed, the reported number of cases is assumed an undercount of the true number of cases. For instance, the CDC estimates that 38 cases of salmonellosis occur in the United States for every case that is reported. Investigation of outbreaks of foodborne disease is a common function of local health departments: Classic examples include food poisoning caused by consumption of a contaminated batch of food from a particular restaurant or supplier.

Many laws regulating the production, transport, and preparation of food are intended to prevent foodborne

disease and limit its consequences; they include laws intended to prevent the contamination of raw food, to mandate its safe preparation and storage, and, if necessary, to close restaurants or food suppliers responsible for disease outbreaks or who fail to follow safe food hygiene practices. There are many means by which raw food may be contaminated, including irrigating or washing produce with unclean water, contamination of meat and poultry with fecal matter during the slaughtering and packaging processes, preparation by food handlers who carry bacteria or viruses on their hands, or using utensils and preparation surfaces that are not clean.

Cooking at a sufficient temperature kills many microbes and parasites; for instance, the U.S. Department of Agriculture recommends cooking eggs until the yolk is firm, cooking pork and ground beef until an internal temperature of 160° F is reached, and cooking whole turkeys and chickens to an internal temperature of 180°F. Microbes may also be present in cooked food (e.g., those introduced during handling after cooking), but this may not pose a health risk if only a small number are present. Most bacteria grow rapidly at room temperature, while refrigeration or freezing keeps them from multiplying (*L. monocytogenes* is a notable exception), hence the recommendations that cooked food be promptly refrigerated to prevent the multiplication of disease-causing organisms.

—Sarah Boslaugh

See also Epidemiology in Developing Countries; *Escherichia coli;* Field Epidemiology; Government Role in Public Health; Outbreak Investigation; Parasitic Diseases; Waterborne Diseases

Further Readings

Adley, C. C. (Ed.). (2006). *Foodborne pathogens: Methods and protocols*. Totowa, NJ: Humana Press.

Centers for Disease Control and Prevention. (2005). *Foodborne illness*. (FAQ revised January 10, 2005). Retrieved February 28, 2007, from http://www.cdc.gov/ncidod/dbmd/diseaseinfo/files/foodborne_illness_FAQ.pdf.

World Health Organization. (2000). *Foodborne disease: A focus for health education*. Geneva, Switzerland: World Health Organization.

World Health Organization. (2007). *Food safety and foodborne illness. Fact sheet no. 237* (Revised March 2007). Retrieved February 28, 2007, from http://www.who.int/mediacentre/factsheets/fs237/en.

Formulary, Drug

A drug formulary is a list of pharmaceutical products; formularies are often created for the purposes of drug purchasing, dispensing, or reimbursement. Many different entities involved in the delivery of, and payment for, health care services use formularies, including health plans, institutions, and government bodies. A common reason for creating a formulary is to specify which drugs may be purchased, or which drugs' purchase will be subsidized, by the entity. The use of drug formularies is largely a reaction to increasing prescription drug costs and the multiplicity of drugs produced by different manufacturers that are available to treat common conditions. For instance, the Veterans Administration (VA) of the United States has used a National Formulary since 1997 that indicates which drugs the VA will routinely pay for. The VA will also grant exceptions to the formulary for particular patients, but the individual or his or her doctor must complete an application process and state why the nonformulary drug is necessary for the individual's care (e.g., if the patient has had an adverse reaction to the formulary drug, or if no formulary alternative exists).

In a multitier formulary, the insurer classifies drugs into tiers and grants favorable treatment to drugs in lower tiers, typically requiring consumers to make higher copayments for drugs in higher tiers. In a typical two-tier plan, generic drugs comprise the first tier and brand-name drugs the second tier. In a three-tier plan, generic drugs constitute the first tier, and brand-name drugs are split into preferred drugs (second tier) and nonpreferred drugs (third tier). A study by Strunk and Ginsberg found that in 2002, 57% of employees with prescription drug coverage had health insurance plans that included three-tier formularies. Some companies have four-tier plans, which are similar to three-tier plans with the addition of a fourth tier of drugs whose purchase is not subsidized by the company, but which are available for purchase by plan members at a negotiated price. Use of a drug formulary allows insurance companies to save money in three ways: It is a bargaining tool that allows the insurer to negotiate for lower prices from pharmaceutical companies; it encourages health plan members to reduce drug utilization and use lower-cost generic or preferred drugs; and it produces income from copayments.

Formularies are a contentious issue in the United States because most Americans get their health

insurance coverage through private insurers, many of whom have instituted multitier drug formularies to manage costs. However, the U.S. federal government does not negotiate or regulate drug prices so that a wide range of drugs at a wide range of prices are available for purchase in the United States. Some people feel that the use of drug formularies may endanger patient health by restricting the choice of therapeutics drugs, since the unsubsidized portion of a higher-tier drug may be prohibitive to the patient.

Economic and Health Effects of Drug Formularies

The primary purpose of drug formularies is to control prescription drug costs without harming the health of the insured or increasing costs in another aspect of care (e.g., by causing an increase in visits to the emergency department because a medical condition was not adequately controlled by drugs). Studies by Motheral and Fairman found that moving to a plan with three tiers from one with two tiers reduced prescription utilization and net costs for prescription drugs, without increasing utilization of office visits, hospitalization, or emergency room visits. Other studies have found that an increased number of tiers in drug formularies leads to decreased utilization and decreased company costs, but also increased costs to the insured. They have also found formularies to be effective in shaping consumer behavior, for instance, increasing the use of lower-cost alternatives such as prescriptions-by-mail plans.

It is difficult to fully evaluate the health effects of drug formularies, in part because each formulary is different, and also because health effects would probably be small and difficult to detect. In addition, due to the fragmentary nature of health and insurance databases for most Americans covered under private insurance, gaining access to both medical care and prescription records for individuals, and matching them up, is difficult in the best of cases and frequently impossible. Because of these difficulties, researchers have focused on narrower questions such as the effect of increased tiers or copayments on utilization and continuation of maintenance medications for chronic conditions such as hypertension. Results have been equivocal: Some studies have found no effects, while others have found weak effects for behaviors such as decreasing or discontinuing usage of maintenance medications, either of which could pose a health risk to the individual. However, no studies to date have established that the use of formularies per se harms the health of individuals.

—Sarah Boslaugh

See also Governmental Role in Public Health; Health Care Delivery; Health Care Services Utilization; Health Economics; Medicaid; Medicare

Further Readings

Huskamp, H. A., Deverka, P. A., Epstein, A. M., Epstein, R. S., McGuigan, K. A., & Frank, R. G. (2003). The effect of incentive-based formularies on prescription-drug utilization and spending. *New England Journal of Medicine*, *349*, 2224–2232.

Joyce, F., Escarce, J. J., Solomon, M. D., & Goldman, D. P. (2002). Employer drug benefit plans and spending on prescription drugs. *Journal of the American Medical Association, 288*(14), 1733–1739.

Motheral, B., & Fairman, A. (2001). Effect of a three-tier prescription copay on pharmaceutical and other medical utilization. *Medical Care, 39*(12), 1293–1304.

Strunk, B. C., & Ginsburg, P. B. (2004). Tracking health care costs: Trends turn downward in 2003. *Health Affairs Web Exclusive, W4*, 354–362.

U.S. Department of Veterans Affairs. (2007). *Pharmacy benefits management strategic healthcare group*. Retrieved January 10, 2007, from http://www.pbm.va.gov/NationalFormulary.aspx.

Framingham Heart Study

The Framingham Heart Study is one of the most important epidemiological studies in the annals of American medicine. Its impact on the study of cardiovascular diseases is particularly important: Much of what is now common knowledge about heart disease, such as the effects of smoking, diet, exercise, and aspirin, can be traced back to the Framingham study. Most important, the study has played a key role in influencing physicians to place greater emphasis on preventing, detecting, and treating cardiovascular disease risk factors in their earliest stages.

In 1948, Framingham, a small town in eastern Massachusetts, was selected as the site of a long-term medical study of heart disease and stroke. The project was initiated under the direction of the National Heart Institute, now the National Heart, Lung, and Blood

Institute (NHLBI). Through a contract with the NHLBI, researchers from the Boston University School of Medicine have played an important role in the Framingham Heart Study since 1971.

More than 50 years later, that follow-up study has yielded an incredible amount of information about heart diseases and the risk factors that can lead to them. The original cohort of the Framingham Heart Study included two thirds of the adult population of Framingham, with ages ranging from 30 to 62 years in 1948. The study was designed to track health information on men and women without signs of heart disease. Every 2 years, people enrolled in the study submitted to dozens of medical tests and answered detailed questions about their personal habits. Over the years, researchers recorded who got heart disease and who did not and studied the connections between disease and the data that had been collected. The two groups thus formed were statistically analyzed and compared.

Originally, researchers enrolled 5,209 Framingham residents (known as the Original Cohort), some of whom are still participating more than 50 years later. In 1970, the study added 5,124 new recruits, referred to as the Offspring Cohort, who were children of the original study group and their spouses. A Third Generation Cohort consisting of individuals who had at least one parent in the Offspring Cohort was recruited beginning in 2001, and 4,095 participants had enrolled in this cohort by June 2005. More recently, 500 members of Framingham's minority community have been recruited to participate in the Omni Study that was initiated and has continued to recruit people with the purpose of determining whether the risk factors associated with disease continue to be the same that were identified in the two previous cohorts.

The findings of the Framingham Heart Study have produced a revolution in preventive medicine and changed the way the medical community and the general public view the origins of disease. More than 1,000 papers based on the Framingham data have appeared in important scientific reviews and have inspired many clinical trials that have been crucial to understand how to manage heart disease and how to control major risk factors. In addition to its great contributions in the area of heart research, the Framingham Study has provided very useful evidence to investigate cancer, stroke, osteoporosis, arthritis, dementia, diabetes, eye disease, and the genetic patterns of many other common diseases.

Information gained from analyses of the Framingham cohort changed the views of the scientific community about heart disease. For instance, before Framingham, most physicians believed that since atherosclerosis is inherent to the aging process, blood pressure should increase with age to enable the heart to pump blood through an elderly person's narrowed arteries.

Risk Factors

For the first time, Framingham established beyond all reasonable doubt a relationship between levels of cholesterol and risk for disease. Furthermore, it established a strong positive association between low-density lipoprotein cholesterol with coronary heart disease and a powerful inverse and protective effect of high-density lipoprotein levels. Presently, researchers are working to identify the genes that regulate cholesterol metabolism to discover the mechanisms by which genes contribute to common metabolic disorders such as obesity, hypertension, diabetes, and even Alzheimer's disease. The Framingham study has collected a DNA library of blood samples from more than 5,000 individuals of two different generations, which will help researchers investigate what diseases run in families and identify the genes responsible for several serious disorders.

Research on blood pressure based on the Framingham cohort has dispelled several misconceptions. Before Framingham it was believed that women and the elderly tolerated higher pressures well. However, analyzing the Framingham data, researchers found nothing to support the contention that the elderly fare better than younger persons at a given degree of hypertension nor that women with high blood pressure are at lower risk than their male counterparts.

Framingham researchers found that an unhealthy diet, sedentary living, and weight gain can aggravate disease risk factors and influence the occurrence of cardiovascular problems. They also proved that smokers are at higher risk of having a myocardial infarction or experiencing a sudden death. Moreover, they found that the risk is related to the number of cigarettes smoked each day and that smoking cessation can halve the risk of ex-smokers compared with those who continue to smoke.

Other studies derived from Framingham demonstrated a protective effect on the heart from even low levels of exercise. Weight gain, accompanied by lack

of exercise, was also found to aggravate the effect of cardiovascular risk factors as were hypertension and diabetes.

—*Jorge Bacallao Gallestey*

See also Cardiovascular Disease; Obesity; Observational Studies; Physical Activity and Health

Further Readings

Asztalos, B. F., Cupples, L. A., Demissie, S., Horvath, K. V, Bloomfield, H. E., Robins, S. J., & Schaefer, E. J. (2004). High-density lipoprotein subpopulation profile and coronary heart disease prevalence in male participants of the Framingham Offspring Study. *Arteriosclerosis, Thrombosis and Vascular Biology, 24*, 2181–2187.

Fox, C. S., Larson, M. G., Leip, E. P., Culleton, B., Wilson, P. W., & Levy, D. (2004). Predictors of new-onset kidney disease in a community-based population. *Journal of the American Medical Association, 291*, 844–850.

Kannel, W. B., D'Agostino, R. B., Sullivan, L., & Wilson, P. W. F. (2004). Concept and usefulness of cardiovascular risk profiles. *American Heart Journal, 148*, 16–26.

Kannel, W. B., & Giordano, M. (2004). Hypertensive risk assessment: Cardiovascular risk factors and hypertension. *Journal of Clinical Hypertension, 6*, 393–399.

Kannel, W. B., & Vasan, R. S. (2004). Assessment of cardiovascular risk and choice of antihypertensive therapy. *Current Hypertension Reports, 6*, 346–351.

Kenchaiah, S., Gaziano, J., & Vasan, R. S. (2004). Impact of obesity on the risk of heart failure and survival after the onset of heart failure. *Medical Clinics of North America, 88*, 1273–1294.

Web Sites

Framingham Study: http://www.framinghamheartstudy.org/index.asp.

Frost, Wade Hampton (1880–1938)

Wade Hampton Frost was a pioneering epidemiologist. Following a distinguished career in the United States Public Health Service, he became the first professor of epidemiology at the Johns Hopkins School of Hygiene and Public Health. He made seminal observations on the epidemiology of infectious diseases and major contributions to epidemiological methods.

Frost was born in Marshall, Virginia, on March 3, 1880. He attended the University of Virginia, graduating in 1903 with both bachelor's and medical degrees. After internships in New York, he enlisted in the U.S. Public Health Service in 1905. In 1908, he was assigned to the National Hygienic Laboratory in Washington, the forerunner of the National Institutes of Health.

Frost studied water pollution at the National Hygienic Laboratory. His paper describing a typhoid outbreak in Williamson, West Virginia, in 1910 is an exemplary report of a field investigation. Between 1909 and 1912, he coupled assays of neutralizing serum antibodies with careful field investigations of polio outbreaks in Iowa, Ohio, Kentucky, and New York. From these studies, he formulated the concept that asymptomatic poliovirus infections in children were common and produced immunity. In 1913, Frost was placed in charge of a newly opened station in Cincinnati, Ohio, to study pollution of the Ohio River and other inland waterways. When the 1918 to 1919 influenza pandemic erupted, he directed the Public Health Service's Office of Field Investigations of Influenza, there working with statistician Edward Sydenstricker, who would become a lifelong friend and collaborator. Frost devised the system of tracking influenza epidemics by using reported pneumonia and influenza deaths as a surrogate in the absence of reporting requirements for influenza itself.

In 1919, William Henry Welch recruited Frost to the newly founded Johns Hopkins School of Hygiene and Public Health to become a resident lecturer, the first faculty member in the new department of epidemiology. He was promoted to professor in 1921 and was elected dean in 1931. Frost developed a curriculum based on lectures, case studies of actual epidemics, and student theses that has served as a model for curricula in epidemiology to the present. He was idolized by students for his work with them in case-study laboratories. His initial investigations at Johns Hopkins focused on acute infectious diseases; later, he worked on tuberculosis, studying the natural history of that disease in Williamson County, Tennessee. Out of this work came the first enunciation of the index case concept and the use of life tables to express data as person-years and to estimate secondary attack rates. His best-known work may be his analysis of shifting tuberculosis age profiles using historical data from Massachusetts. With Lowell Reed, he developed the first mathematical expression of the epidemic curve.

Frost died of esophageal cancer on May 1, 1938. Shortly before his death, the American Public Health Association awarded him its prestigious Sedgwick Memorial Medal.

—*Thomas M. Daniel*

See also Cohort Studies; Influenza; Polio; Tuberculosis; Waterborne Diseases

Further Readings

Daniel, T. M. (2004). *Wade Hampton Frost, pioneer epidemiologist, 1880–1938: Up to the mountain*. Rochester, NY: University of Rochester Press.

Maxcy, K. F. (1941). *Papers of Wade Hampton Frost, M.D. A contribution to the epidemiological method*. New York: Commonwealth Fund.

F Test

The F test is used for hypotheses concerning variances. It is used to test whether two variances are equal and is also used in the analysis of variance (ANOVA).

Let U and V be two independent chi-square random variables having m and n degrees of freedom (df), respectively. The ratio $F = (U/m)/(V/n)$ has an F distribution with numerator df equal to m and denominator df equal to n. Two properties of the F distribution are as follows:

- If $X \sim F_{m,n}$, then $1/X \sim F_{n,m}$, that is, the reciprocal of an F random variable is again an F random variable.
- If $X \sim t_m$ (X has a t distribution with $df = m$), then $X^2 \sim F_{1,m}$.

Test for Equality of Variances or Homogeneity of Variances

The F test is used to test whether two population variances are equal ($H_0 : \sigma_1^2 = \sigma_2^2$) versus either a one-tail alternative, upper or lower, $H_a : \sigma_1^2 > \sigma_2^2$ or $H_a : \sigma_1^2 < \sigma_2^2$, respectively, or a two-tail alternative, $H_a : \sigma_1^2 \neq \sigma_2^2$. The test statistic is the ratio of $F = s_1^2/s_2^2$, where s_1^2 is the sample variance from Sample 1 with $df = n_1 - 1$ and s_2^2 is the sample variance from Sample 2 with $df = n_2 - 1$. If H_0 is true, then the ratio would be close to 1. The farther away the ratio is from 1, the stronger the evidence for H_a. (Note that $0 < f < \infty$ since $s^2 > 0$.) One rejects the null hypothesis in the following cases, where $F_{\alpha,m,n}$ is the critical value of the F distribution with numerator $df = m$, denominator $df = n$, and a significance value α:

- $F > F_{\alpha, n_1 - 1, n_2 - 1}$ for an upper one-tailed test.
- $F < f_{1 - \alpha, n_1 - 1, n_2 - 1}$ for a lower one-tailed test.
- $F < f_{1 - \alpha/2, n_1 - 1, n_2 - 1}$ or $F > F_{\alpha/2, n_1 - 1, n_2 - 1}$ for a two-tailed test, although some texts may recommend that the F test statistic be found where the largest sample variance is in the numerator and reject H_0 if $F > F_{\alpha/2, n_1 - 1, n_2 - 1}$.

ANOVA and the *F* Test

In many experimental design situations, the data are analyzed using an ANOVA table. The plethora of problems makes it too difficult to explain each of them here. Instead, the general idea is presented using the one-way ANOVA table. In this situation, one wishes to test if the means are equal versus at least one population mean is different. A sample ANOVA table is given in Table 1. The details of where the values were derived from are not given since the interest is in the F test itself. In general, df gives the degrees of freedom and SS gives the sum of squares. The MS stands for mean squares and is calculated by SS/df. Then, the F ratio is calculated using the ratio of the MS. For example, $F = 48.8/13.6 = 3.58$. The null hypothesis is rejected when $F > F_{\alpha, df_{\text{num}}, df_{\text{denom}}}$. In this case, $F_{\alpha = 0.05, 3, 12} = 4.49$, so in this situation, one would reject H_0.

When given an ANOVA, keep in mind the following:

- Know what the null hypothesis is that the F test is testing. This will change for different situations and the ANOVA tables.

Table 1 Sample ANOVA Table

Source	*df*	*SS*	*MS*	*F*
Variable 1	3	146.4	48.8	3.58
Error	12	163.5	13.6	
Total	15	309.9		

- The F ratio is the ratio of the mean squares (*MS*).
- The numerator *df* for the critical value of the test is the degrees of freedom associated with the mean square value that is in the numerator of the F ratio; likewise, the denominator *df* is the *df* associated with the mean square value in the denominator.

—Marjorie E. Bond

See also Analysis of Variance; Degrees of Freedom; Hypothesis Testing; Study Design

Further Readings

Mood, A., Graybill, F., & Boes, D. (1974). *Introduction to the theory of statistics* (3rd ed., pp. 246–249). New York: McGraw-Hill.

Moore, D. S. (2007). *The basic practice of statistics* (4th ed., pp. 476–480, 620–644). New York: W. H. Freeman.

Functional Status

Functional status is a common health measure used in both clinical care and research. The general domains of functional status are physical, mental, and sometimes social. Numerous self-report and clinician-rated instruments for measuring functional status exist in the public health literature. Poor functional status has been shown to be associated with chronic medical conditions, comorbidities, age, and mortality. *Functional status* is a broad term that has come to encompass a number of concepts that can vary with the many different instruments developed for its measurement. Although sometimes used interchangeably, functional status can differ conceptually from health status or quality of life.

Functional status originated as a measure of disability status, but it has come to be commonly used as a health research measure. Some make a distinction between functional status (how one does something) and functional ability (the capacity to do something). Functional status is often based on assessment of individuals' performance of activities without assistance from other people or assistive technology. In addition, functional status measures can be specific to different conditions or generic. Condition-specific measures are intended to assess functional limitations caused by the specific disease/disorder of interest. Generic measures are designed to assess functional status outside of a specific medical context.

Among the most commonly used measures of physical functioning are those that assess activities of daily living (ADL) and instrumental activities of daily living (IADL). ADL are activities related to personal care and include bathing or showering, dressing, getting in or out of bed or a chair, using the toilet, and eating. IADL are activities related to independent living and include preparing meals, managing money, shopping for groceries, performing housework, and using a telephone.

Mental functional status is often based on cognitive function, psychological symptom severity, or social role function such as ability to engage in social activities or in work. Cognitive activities assessed include orientation to date and place, recall, language skills, attention span, ability to calculate, and to identify objects. Commonly assessed psychological symptomatology include feelings of depression, anxiety, phobias, stress, confusion, problems concentrating, problems making or keeping friends, and interacting with others.

Toward the end of the 20th century, and influenced by the social model of disability, assessment of functional status developed to include contextual considerations as in the World Health Organization's (WHO) *International Classification of Functioning, Disability, and Health* (ICIDH-2, hereafter ICF). The purpose of the ICF classification is to allow a unified and standard language for assessing health and health-related components of well-being. It allows a person's functioning and disability to be viewed as a dynamic interaction between health conditions and contextual factors. By integrating multiple dimensions of disability (i.e., structural and functional impairments, activity limitations, participation restrictions, and environmental factors), the ICF is intended to characterize functional status in terms of the interaction between intrinsic individual characteristics and extrinsic social phenomena. The ICF approach is considered by some to be more empowering to individuals because it does not classify people but rather the situations in which they live and experience health and health-related concerns.

—Jane K. Burke-Miller

See also Disability Epidemiology; EuroQol EQ-5D Questionnaire; Quality of Life; Quality of Well-Being Scale (QWB); SF-36® Health Survey

Further Readings

Baker, M. (1990). *Functional status measurement in primary care*. New York: Springer-Verlag.

Bombardier, C., & Tugwell, P. (1987). Methodological considerations in functional assessment. *Journal of Rheumatology, 14*, 6–10.

Cohen, M. E., & Marino, R. J. (2000). The tools of disability outcomes research functional status measures. *Archives of Physical Medicine & Rehabilitation, 81*(Suppl. 2), S21–S29.

Gray, D. B., & Hendershot, G. E. (2001). The ICIDH-2: Developments for a new era of outcomes research. *Archives of Physical Medicine & Rehabilitation, 81*, S10–S14.

G

Gay, Lesbian, Bisexual, and Transsexual Health Issues

See Sexual Minorities, Health Issues of

Gene

In the early 20th century following the rediscovery of the important work of Gregor Mendel that provided the foundation for the future field of genetics, scientists developed the classical concept of the gene as the entity responsible for many human traits and diseases. It was soon learned that genes are located on the chromosomes and each person has two copies of a particular gene, one on each of the paired chromosomes that are inherited from one's parents, one of maternal origin and the other from the father. The alternate forms of a gene that result from the process of mutation are called alleles.

The 21st century likewise began with significant discoveries in the world of genetics. The Human Genome Project (HGP) begun in the mid-1980s by an international team of researchers led by the U.S. Department of Energy and National Institutes of Health (NIH) announced in 2003 that a map of the human genome was completed. The announcement of the human genome map coincided with the 50th anniversary of the discovery of the DNA helical structure by Watson and Crick in 1953. This latter discovery gave us the basis for the structure of the gene and how it replicates and marked the beginning of the molecular revolution in genetics and in biology and medicine.

Growing knowledge from the HGP and other genomic research will make it possible to diagnose and treat disease and to identify individuals at risk for a certain disease in ways that, until recently, were inconceivable. The HGP discoveries add to our understanding of how disease mechanisms occur so that treatment can focus on the primary dysfunction or disease progressions rather than treating only the secondary manifestations or outcomes. Genetics also plays a role in the prevention of disease and in health promotion. Most diseases and our health are now thought to be the result of the interplay between multiple genes (called polygenes) and the multitude of environmental exposures to which an individual is subjected during development and over the course of life. Some geneticists now think that even single gene (monogenic) diseases and traits are really multifactorial (the result of polygenes and environmental influences) and are complex due to interactions between genes and between genes and environmental factors.

To understand the new knowledge from the map of the human genome and how this knowledge affects society, it is important to have a clear understanding of what a gene is, as we understand it today. A gene is defined as the fundamental physical and functional unit of heredity. A gene is made up of deoxyribonucleic acid, called DNA. If this DNA were stretched out, it would look like a very long circular staircase with millions of rungs. Small sections of rungs are analogous to genes. Each of these sections of DNA is called a gene—a piece of genetic information that does one particular job. That job is the encoding of a chain of

amino acids to form a polypeptide, the major components of the essential proteins of the body. Each gene is a different packet of information necessary for our bodies to grow and function. Our genes also contain the information governing physical traits such as the color of our eyes, how tall we are, and the shape of our ears and nose, and some of the traits that affect our vulnerability to diseases. In humans, genes vary in size from a few hundred DNA bases to more than 2 million bases. The HGP has estimated that humans have between 20,000 and 25,000 genes.

Genetics plays an increasingly important role in the practice of health care. The global concepts of the gene and genetics are not confined to our physical features and the development of rare diseases but rather have become key components of our understanding of most of the major common diseases, including heart disease, diabetes, cancer, and many psychiatric disorders. The classical concept of the gene has given way to the modern molecular concept, and genetics has broadened its horizons as an explanatory factor across all fields of inquiry that attempt to understand biological variation and its basis.

—Kathy Prue-Owens

See also Genetic Disorders; Genetic Epidemiology; Genotype; Human Genome Project; Icelandic Genetics Database

Further Readings

Bates, B. R., Templeton, A., Achter, P. J., Harris, T. M., & Condit, C. M. (2003). What does "A gene for heart disease" mean? A focus group study of public understandings of genetic risk factors. *American Journal of Medical Genetics, 119*, 156–161.

Jorde, L. B., Carey, J. C., Bamshad, M. J., & White, R. L. (2000). *Medical genetics* (2nd ed.). St. Louis, MO: Mosby.

McCance, K. L., & Huether, S. E. (2006). *Pathophysiology: The biologic basis for disease in adults and children.* St. Louis, MO: Elsevier Mosby.

Gene-Environment Interaction

In the field of epidemiology, there has been increasing interest in documenting the results of interaction between specific genes or genotypes and well-understood environmental exposures. Gene-environment interactions are situations in which the combination of an environmental factor, such as exposure to cigarette smoking, with a genetic factor, such as a disease-predisposing mutation, results in a greater risk or severity of the disease in question than either the genetic or environmental factor acting alone. These interactions have been studied in plants and animals for many years. While they have been hypothesized also to occur in humans, they have been documented in humans only fairly recently. In recent years, many examples of gene-environment interaction have been reported in the literature.

There is now increasing acceptance of the idea that it is likely that all human disease is the result of the interaction between the genetic susceptibility to disease and environmental exposures during the course of life. It is important to study gene-environment interactions for several reasons. One is the expectation that it will improve our understanding of the etiology of specific diseases. Second, we might be able to identify populations that are at high risk due to their possessing higher frequencies of genotypes that denote susceptibility to the disease in question. We might also be able to identify modifiable risk factors in the form of environmental exposures. Finally, there is the underlying goal of being able to prevent disease through improved understanding of the risk factors that are involved and the knowledge of the underlying disease mechanisms.

Before we discuss what an interaction is—particularly an interaction between a gene and some environmental exposure—we need some background thinking about what an effect of a genetic factor is and how we should conceptualize the environment. Genetic variation in human populations fits the label given to it by some as the "ultimate public health problem." With the ever mounting discoveries of disease-related genetic variation and the identification of specific genes related to disease, the statement that everyone is at genetic risk for some disease has become an acceptable assessment. In genetic epidemiology, the search for disease-associated genes has been continuing for many years, first through family studies and later through specific analyses involving estimation of heritability and segregation and linkage analysis. More recently, the focus has been on large-scale association studies in populations.

In research on the genetics of human diseases, one might ask what are the measurements for the genotype or, for that matter, for the genes that might

increase one's risk of a disease. Family history, that is, the presence or absence of a family history for a disease in question, is one measure of genotype. However, it is crude at best and susceptible to the misclassification of significant proportions of individuals in an analysis. If a phenotype has been adequately described and the specific features of the disease are well known and easily observed and measured, then the specific phenotype for the presence of the disease may be indicative of an underlying gene or complex of genes for that disease. But phenotypic descriptions of diseases are subject to analytical problems, especially where those diseases are heterogeneous in their expression or represent a spectrum of disorders, such as autism. In such cases, the so-called behavioral phenotype becomes extremely important in studies of the genetic factors in the disease entity in question.

We get the best measures of the influence of genes or genotypes when we can analyze the DNA directly. Even measures based on the product of genes or the biochemical profiles produced by genes are better than data at the level of the phenotype or information about the presence of a family history. The discovery of a specific gene as the cause of a disease unleashes many other research tracks. Some involve the search for the gene product and its function. Allelic variants of the disease can be discovered and described, and their associations with particular health outcomes can be investigated. Finally, there is the basic research required to investigate possible interactions between the disease gene and other genetic loci during early development and in processes that occur in the development of complex adult diseases such as cancer and diabetes.

The measurement of the environment is equally worthy of discussion. In the assessment of environment exposures, we are dealing with a vast array of aspects of the environment that can potentially be involved in disease. They include exposures to chemicals in the air, water, and soil; various nutrients in our diets; a host of infectious agents; drugs used in the treatment of disease; and, during development, the maternal environment and the multitude of chemicals to which the developing fetus is exposed, such as alcohol and drugs taken during pregnancy. Just as there are problems in measuring the genetic influences, the measurement of environmental exposures is fraught with difficulties. We would like to know the exact concentration of a substance that has interacted with specific structures or biochemical systems in the body in a way that induces disease, but we are never able to measure this. Instead, we use surrogate measures. The common approach is to collect an exposure history. This may include a detailed residential history so that exposures to known sources of pollutants can be estimated. In adults, the collection of an occupational history can be an important means of assessing potential exposures to chemicals and other environmental factors on the job. We have witnessed during the past 20 years the accelerated development of biomarkers at the chromosomal and molecular levels to assist further with the assessment of environmental exposures. Indeed, the new field of molecular epidemiology emerged initially from environmental epidemiology and attempts to improve exposure assessment through the identification of biomarkers that could potentially specify that an exposure to a particular chemical had occurred. These biomarkers are a much more accurate way of estimating the concentration of substances that has interacted with specific structures or biochemical systems in the body in a way that induces disease.

Before we discuss the meaning of gene-environment interaction in more detail, it should be explained that an interaction is really a statistical phenomenon. An interaction in statistical terms is perhaps mostly easily grasped in graphic form. In fact, it is one of those instances in data analysis when the graph tells all and the results hit the observer "right between the eyes."

In the third edition of their textbook *Biometry*, Robert Sokal and F. James Rohlf (1995) created an artificial example based on actual experimental data to illustrate the meaning of the term *interaction* in a two-way analysis of variance. The original data are the amount of food consumed by rats (in grams) during a 73-day experiment. The data are classified by two factors, the type of fat in the diet (fresh vs. rancid lard) and sex (male vs. female), with three rats in each diet and sex combination. In the artificial example, a significant interaction is shown between the two factors. Plotting the means for each combination as shown in Figure 1 shows the crossing lines that are observed in some instances when a significant interaction has occurred between the two factors (diverging lines may also be observed when there is significant interaction).

Now consider some data that demonstrate a gene-environment interaction. In the following example, taken from the volume by Bodmer and Cavalli-Sforza

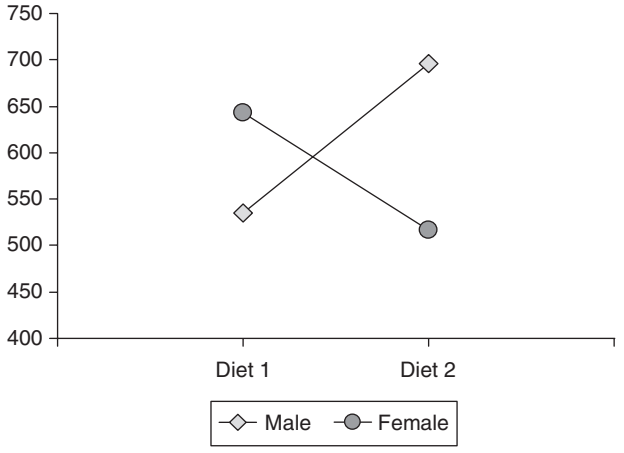


Figure 1 Interaction Between Diet and Sex for Food Consumption in Rats

(1976), we have two genotypes, G1 and G2, and two environments, E1 and E2. Table 1 shows four individuals, two of each genotype, distributed in the two environments.

If we graph these data in Figure 2, they look as follows:

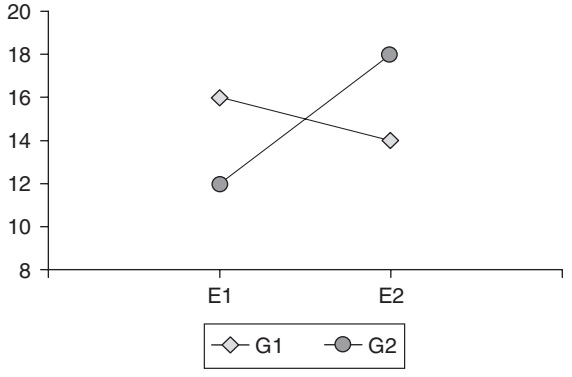


Figure 2 Illustration of Genotype-Environment Interaction

Source: Adapted from Bodmer and Cavalli-Sforza (1976).

Again, this is an example of the graph tells all. Genotype G1 has a higher phenotypic value than G2 when they are each exposed to environment E1. However, this pattern is reversed when the genotypes are exposed to E2, with G2 now having a higher phenotypic score than G1. The plotted lines are similar to what we observed in the diet experiment with rats in that they form the crossing pattern typical of an interaction effect.

Table 1 Phenotype Scores for Four Individuals by Genotype and Environment

Individual	*Genotype*	*Environment*	*Phenotype Score*
1	G1	E1	16
2	G1	E2	14
3	G2	E1	12
4	G2	E2	18

Source: Adapted from Bodmer and Cavalli-Sforza (1976).

Research on gene-environment interaction in human diseases is on the increase. An example of Medline searches conducted in two ways will demonstrate this pattern. One search was conducted for articles by the specific topic gene-environment interaction. The results of the first search specifically for the topic of gene-environment interaction were grouped by decades from 1950 to 1959 through 2000 to the present. A second search looked for articles that combined genetics and environment topics in the same article. The genetics topics included genetic predisposition to disease, genes, or genotype, while environmental topics were environment, environmental exposure, and social environment. The search results are displayed in Figure 3. From the 1980s through the present, there has been a dramatic increase in the number of publications that address gene-environment interaction per se or the combination of both genetic and environmental factors.

One category of diseases that has received much attention in the search for gene-environment interactions is birth defects. Orofacial clefts such as cleft lip and cleft palate result when there is failure to obtain complete closure of the upper lip and/or the palatal ridge inside the mouth during embryonic development. These abnormalities are known to involve both genetic and environmental factors. Early evidence of genetic influences came from twin studies in which concordance rates, the occurrence of the defect in each member of the twin pair, were much higher in identical (monozygotic) twins than in fraternal (dizygotic) twins for both types of orofacial clefts. More

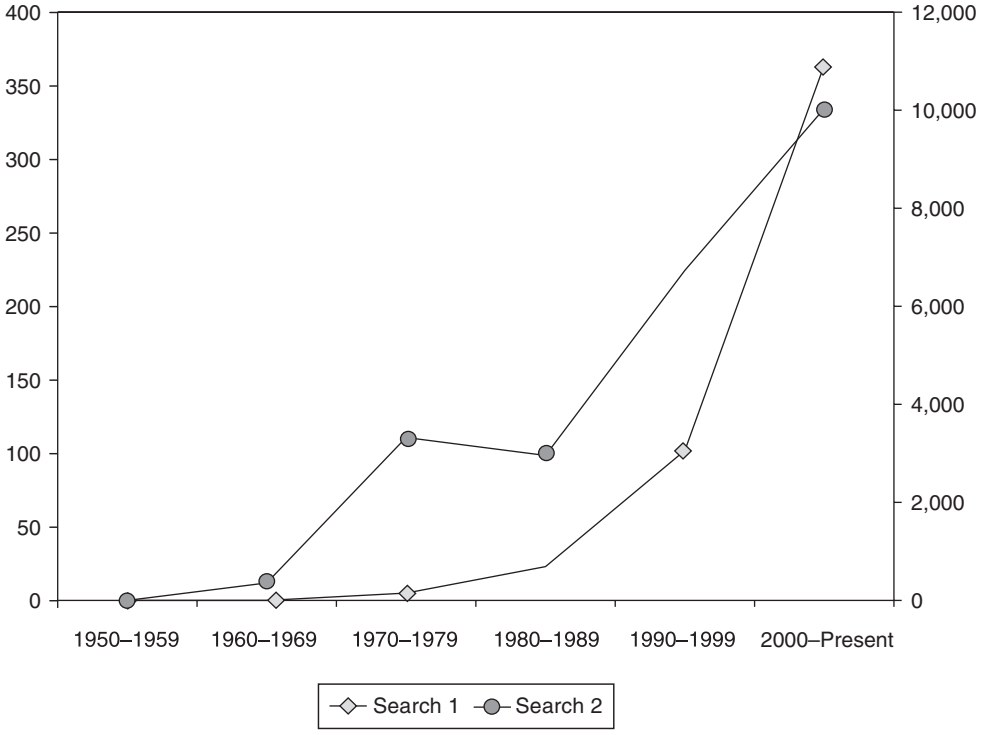


Figure 3 Publications on Gene-Environment Interaction by Decade: 1950 to the Present

Source: Medline search (2007).

recently, increased risk for cleft lip and palate has been observed in individuals with specific mutant alleles, including one for the transforming growth factor-alpha (TGFα) gene. This gene produces a protein that has been linked to oral-facial development. Evidence has also mounted that maternal smoking during pregnancy could be a risk factor for orofacial clefts in infants.

One study that was able to look at both genetic and environmental factors in orofacial clefting produced evidence of a gene-environment interaction. The study was population based and involved cases of orofacial clefts and appropriate controls. Cases were identified through a birth defects surveillance program. Telephone interviews were conducted with the mothers of both cases and controls to collect information that included smoking during pregnancy. Genotyping for the variant alleles of the TGFα gene was also conducted. One of the gene-environment interactions is shown in Figure 4.

The risk of cleft lip with or without cleft palate, as determined by the odds ratio, was the highest in those infants who had the TGFα mutation (A2) shown in previous work to be associated with clefting and whose mothers smoked at least 20 cigarettes a day. Again, the diverging lines in the graph show the interaction.

Failure to account for gene-environment interactions may be a barrier to an understanding of disease etiology. Say that we are interested in examining the relationship of an exposure, substance Q, on the risk of an outcome, disease Y. We may collect data on exposure and disease and determine that exposure to substance Q does not affect the risk of disease Y. In Table 2, column A is the 2×2 table that results from a study of a population of 100,000 of whom 50% are exposed to substance Q.

Since the probability of becoming a case is the same for those exposed or not exposed, the $RR = 1$, and we say that there is no association. If we have a population in which a specific mutation is present in 1% of the population, we might examine separately those who do or do not have the mutation. Columns B and C are the results for the group with the mutation

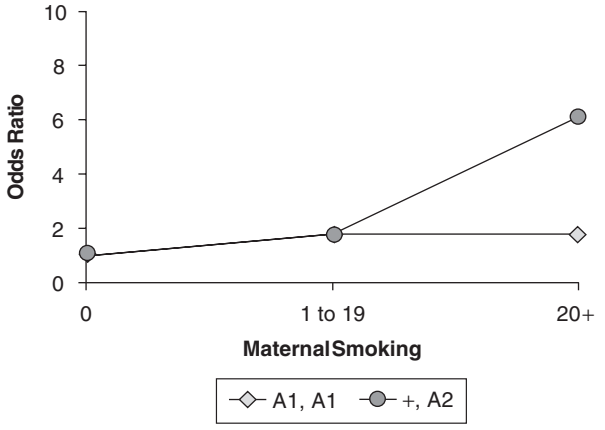


Figure 4 Cleft Lip and/or Cleft Palate Risk Interaction for Maternal Smoking and Transforming Growth Factor-Alpha Genotypes

Source: Shaw et al. (1996).

and the group without the mutation. Again, we have 50% of each group exposed, but in the mutant group the risk of the disease is 5.1 times greater in those exposed to substance *Q* than in those who were not exposed. Among the nonmutants, the probability of disease is equal for the exposed and unexposed. Column D is the sum of columns B and C. In column D, then, the population we are studying includes those with and without the specific mutation, where the frequency of the mutation is low, the likelihood of exposure is the same for those with and without the mutation, and the relationship between exposure and disease is very different for those who do or do not have the mutation. When we calculate the *RR* for this scenario, we find that it is not significantly different from 1, or no association. Therefore, we have missed the strong association of exposure and disease in the mutants because this group is a small minority of the population studied. When we are able to identify those with this specific mutation and examine them separately, we can distinguish this relationship. This would lead to opportunities for disease prevention through targeted prevention of exposure to those with the mutation. When we interpret the results of the studies, we must bear in mind that subgroups within the population who have a different genetic makeup may demonstrate a very different response to exposure providing opportunities for further understanding mechanisms of disease, prevention, and treatment.

—F. John Meaney and Sydney Pettygrove

See also Association, Genetic; Gene; Genetic Epidemiology; Genotype; Molecular Epidemiology; Phenotype

Further Readings

Bodmer, W. F., & Cavalli-Sforza, L. L. (1976). *Genetics, evolution, and man.* New York: W. H. Freeman.

King, R. A., Rotter, J. I., & Motulsky, A. G. (2002). *The genetic basis of common diseases* (2nd ed.). New York: Oxford University Press.

Ottman, R. (1996). Gene-environment interaction: Definitions and study designs. *Preventive Medicine, 25*, 764–770.

Table 2 Gene-Environment Interaction Example

		Total Populations											
		A			*B*			*C*			*D*		
		Total Population Disease			*Mutants Disease*			*Nonmutants Disease*			*Normal + Mutants Disease*		
		Yes		*Total*	*Yes*	*No*	*Total*	*Yes*	*No*	*Total*	*Yes*	*No*	*Total*
Exposure	*Yes*		49,800	50,000	10	490	500	198	49,302	49,500	208	49,792	50,000
	No		49,800	50,000	2	498	500	198	49,302	49,500	200	49,800	50,000
		400		100,000	12	988	1,000	396	99,600	99,000	408	99,592	100,000
		Relative Risk = 1.0			Relative Risk = 5.1			Relative Risk = 1.0			Relative Risk = 1.0		

Shaw, G. M., Wasserman, C., Lammer, E. J., O'Malley, C. D., Murray, J. C., Basart, A. M., et al. (1996). Orofacial clefts, cigarette smoking, and transforming growth factor-alpha gene variants. *American Journal of Human Genetics, 58*, 551–561.

Genetic Counseling

The goal of the field of medical genetics is to detect and treat individuals with genetic disorders. There are numerous single gene and chromosome disorders, but genetics also contributes to common diseases such as cardiovascular disease, cancer, and diabetes. Genetic counseling is the process of helping people understand and adapt to the medical, psychological, and familial implications of genetic contributions to disease. This process, as defined by the National Society of Genetic Counselors' Task Force, integrates the following:

- Interpretation of family and medical histories to assess the chance of disease occurrence or recurrence
- Education about inheritance, testing, management, prevention, resources, and research
- Counseling to promote informed choices and adaptation to the risk or condition

Genetic diseases due to single gene defects, sometimes referred to as Mendelian disorders, are grouped into *autosomal* (encoded by genes on one of the 22 pairs of autosomes, or non–sex chromosomes) and *X-linked* (encoded mutant genes on the X chromosome, one of the two sex chromosomes). There are two modes of expression of mutant genes: *dominant* and *recessive*. Dominant refers to those conditions expressed in individuals having only one copy of a mutant allele. Recessive refers to those conditions that clinically manifest only in individuals who carry two copies of the abnormal gene. Of the 10,000 human phenotypes known to be inherited, more than one half is classified as autosomal dominant, about one third is autosomal recessive, and about one tenth is X-linked.

The patterns of inheritance of most Mendelian or single gene traits have been learned by observing the transmission of traits within families. An important reason for studying the pattern of inheritance of conditions within families is to advise members through genetic counseling about recurrence risks or the chance that the genetic disease would be passed on to their children. Generally, dominant conditions carry a 50% risk while recessive conditions carry a 25% risk.

The term *genetic counseling* was coined in the 1940s and, as practiced today, is a multidisciplinary activity involving the provision of services by clinical geneticists, genetic counselors, nurse practitioners, laboratorians, and other health care professionals. The components of a genetic counseling interaction include the following:

- Information gathering
- Establishing or verifying the diagnosis
- Risk assessment
- Information giving
- Psychological counseling

Information gathering involves the collection of a family history from an individual knowledgeable about the family and health status of individuals within the family and is usually recorded in the form of a three-generation pedigree using conventional symbols to represent normal and affected males and females within and between generations and their relationships with other family members. The family history helps clarify relationships, identify other affected individuals, and records information of potential genetic significance (consanguinity or inbreeding, fertility problems, birth defects, ethnicity, mental disabilities). The person bringing the family to medical attention is referred to as the index case or proband.

Establishing or verifying the diagnosis is often accomplished by reviewing medical records before the family comes to the genetics clinic. Family photographs are often helpful. Sometimes, this requires physical examinations of other family members and/or laboratory testing.

Risk assessment is performed by analyzing the pedigree and the results of any laboratory testing. Mathematical calculations are often used to incorporate other information such as age of onset of the condition, population carrier frequencies, and number of affected individuals in the family. *Information giving* involves explaining to the family how the diagnosis was arrived at, what the implications are for the affected individual, what specialized services might be of help to the family, reproductive risks, options for prenatal diagnosis, and any treatment options.

Psychological counseling is often helpful because learning about any medical condition can be frightening and distressing. Genetic counseling strives to help

families make the best possible adjustment to the situation, to reach decisions about how they would prefer to deal with the risks and burdens they face, and to carry out plans. Genetic counseling provides services in a nondirective manner while painting an honest and fair picture of the challenges that a family faces.

The mode of inheritance is often established after constructing and reviewing the family history and pedigree. There are certain characteristics noted when reviewing a pedigree that indicate a specific mode of inheritance. For example, the three specific features of an autosomal dominant pedigree are as follows: (1) Males and females are equally affected, (2) more than one generation with affected members gives the impression of *vertical* transmission of inheritance, and (3) all forms of transmission between sexes are observed—that is, male to male (rules out X-linkage), female to female, male to female, and female to male. In autosomal dominant diseases, an affected parent passes either the normal gene or the disease-causing gene to the offspring, and each occurrence carries a 50% probability. There are more than 5,000 recognized autosomal dominant conditions. Some are common such as adult polycystic kidney disease (occurs 1 in 1,000 individuals) and familial hypercholesterolemia (1 in 500).

The distinctive features of autosomal recessive inheritance are the following: (1) Males and females are equally affected, (2) usually only one generation is involved (i.e., brothers and sisters affected) giving a horizontal pattern of inheritance, and (3) consanguinity. There are more than 4,000 recognized single gene diseases that require two copies of the mutant allele for an individual to be affected. The recurrence risk is 1 in 4 or 25%. There is a two thirds chance that an unaffected sibling is a carrier of the recessive gene in a family.

Three features seen in X-linked recessive inheritance are as follows: (1) Males are exclusively affected, (2) unaffected carrier females pass the genetic disease to one half of their sons giving a *diagonal* inheritance pattern (i.e., uncles and nephews are affected), and (3) affected males cannot transmit the genetic disease to their sons (fathers pass the Y chromosome to their sons and the X chromosome to their daughters). Because females have two copies of the X chromosome and males are hemizygous, most X-linked conditions are more common in males than in females.

Three features are necessary to establish X-linked dominant inheritance. These include the following: (1) Both males and females are affected, (2) the number of affected females in the pedigree may be twice the number of affected males, and females are generally less severely affected than males, and (3) affected females can transmit the genetic disease to one half of their sons and daughters but affected males can transmit the disease only to their daughters.

Thus, there are three major categories of genetic diseases: (1) single gene diseases (autosomal dominant, autosomal recessive, and X-linked recessive or dominant); (2) chromosomal (not discussed here because a particular chromosome abnormality—e.g., deletions, inversions, duplications, translocations—involving the individual pairs of chromosomes generates specific and sometimes unique recurrence risk estimates beyond the scope of this presentation), and (3) multifactorial. Multifactorial causes (combination of genetic and environmental factors) account for about two thirds to three fourths of the morbidity and mortality due to genetic diseases. Multifactorial conditions include diabetes, neural tube defects, congenital heart disease, cancer, obesity, and idiopathic mental retardation. Recurrence risk estimates have been established for most of these multifactorial conditions (e.g., 2% to 4% risk for congenital heart disease).

In addition to the typical Mendelian forms of inheritance, non-Mendelian forms exist. These include mitochondrial disorders, genomic imprinting (e.g., Prader-Willi and Angelman syndromes), and trinucleotide repeat mutations (e.g., fragile X, myotonic dystrophy). Mitochondrial diseases are due to mutations of the mitochondrial genome inherited from the mother and found in the cytoplasm of the cell. Mitochondrial, genomic imprinting, and trinucleotide repeat disorders are known to play a role in neurological and muscle dysfunction. Non-Mendelian disorders require unique disease-specific genetic testing approaches and recurrence risk estimates.

—Merlin G. Butler, Michael Begleiter, and Molly Lund

See also Chromosome; Gene; Gene-Environment Interaction; Genetic Epidemiology; Genotype; Mutation; Phenotype

Further Readings

Blum, S., Hahn, S. E., Strecker, M. N., & Williams, J. L. (2006). A new definition of genetic counseling: National Society of Genetic Counselors' Task Force Report. *Journal Genetic Counseling, 15*(2), 77–84.

Butler, M. G., Begleiter, M., Lillis, S., Lund, M., & Meaney, F. J. (2005). Patterns of inheritance: Mendelian and non-Mendelian. In M. G. Butler & F. J. Meaney (Eds.), *Genetics of developmental disabilities* (pp. 35–79). London: Taylor & Francis.

Resta, R., Biesecker, B. B., Bennett, R. L., Blum, S., Hahn, S. E., Strecker, M. N., et al. for the National Society of Genetic Counselors' Task Force. (2006). A new definition of Genetic Counseling: National Society of Genetic Counselors' Task Force Report. *Journal of Genetic Counseling, 15*(2), 77–83.

Rimoin, D. I., Connor, J. M., Pyeritz, R. E., & Korf, B. R. (Eds.). (2006). *Emery and Rimoin's principles and practice of medical genetics* (5th ed.). London: Churchill Livingstone.

Web Sites

Online Mendelian Inheritance in Man (OMIM): http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM.

Genetic Disorders

In a strict sense, genetic disorders result from mutations of single genes and include conditions such as cystic fibrosis, sickle cell anemia, and the muscular dystrophies. In a more general sense, however, chromosome abnormalities, multifactorial diseases, and single gene disorders are often thought of as genetic diseases.

The contribution of genetic disorders to human morbidity and mortality varies throughout the life cycle. Chromosomal abnormalities exert their greatest influence during prenatal and early postnatal life. Single gene disorders usually manifest in infancy and childhood. Multifactorial diseases are particularly common causes of adult health problems and include conditions such as coronary artery disease and type 2 diabetes. Most birth defects and other childhood illnesses such as asthma may also be inherited in a multifactorial fashion. As a group, chromosome abnormalities, single gene disorders, and multifactorial diseases are found in about 3% to 4% of all newborns. By adulthood, previously undetected malformations and later-onset disorders have become apparent, and up to 8% of the population is affected by genetic disease. Although a relatively small proportion of individuals in the general population are affected with one of these conditions, they are overrepresented in certain populations such as those with early mortality and those who are hospital inpatients. Genetic disorders are the most common cause of neonatal mortality, surpassing complications of prematurity. They also represent about 35% to 55% of pediatric hospital admissions and about 20% of deaths in neonatal and pediatric intensive care units. As common causes of worldwide morbidity and mortality such as infectious diseases decrease, the importance of genetic disease, especially in developed countries, has received increasing scrutiny as a target of new diagnostic tests, improved treatment strategies, and public health prevention programs.

Chromosome Abnormalities

Standard cytogenetic testing uses light microscopy to identify loss or gain of visible segments of chromosomes, including whole chromosomes and smaller segments. Although this technique readily detects abnormalities of large segments of chromosomal material, the smallest deletions or duplications are beyond the resolution of microscopic analysis. Over the past two decades, new techniques for diagnosis of smaller and smaller chromosome abnormalities have had a great impact on the practice of medical genetics. With the advent of fluorescence in situ hybridization (FISH) technology, it has become possible to detect deletions or duplication not visible through the microscope. FISH uses fluorescent molecules to bind to genes, chromosome regions, or whole chromosomes and detects missing or extrachromosomal material. More recently, fluorescence technology has been used to construct chromosome microarrays, which can detect literally thousands of abnormalities in a single reaction. Among patients who are evaluated for a chromosome abnormality and who have already had a normal cytogenetic analysis, 5% to 7% have a chromosome deletion or duplication when analyzed by chromosome microarray. Such technical advances will increasingly be used to improve genetic diagnosis and to identify mechanisms by which chromosome imbalance leads to human disease.

Although chromosome abnormalities may be found in all age groups, they are particularly common in spontaneous abortions. Among recognized pregnancies, between 10% to 20% are spontaneously aborted; and of these, chromosome abnormalities are present in about half. The great majority, more than 95%, of these abnormalities, involve an entire missing or extra chromosome, which is referred to as aneuploidy.

Aneuploidy is the result of nondisjunction, which is failure of homologous chromosome pairs to disjoin and assort to different germ cells during the first phase of meiosis. The only well-characterized risk factor for chromosome abnormalities is advanced maternal age, which is associated with aneuploidy. Environmental factors such as parental drug use, cancer chemotherapy, and radiation exposures have been investigated, and none of these appears to increase the relative risk for having an affected child. Although chromosome abnormalities may be present in between 5% and 10% of conceptions, they are found in about 0.5% of newborns. The most common abnormalities in liveborn infants are trisomies for chromosomes 21, 18, and 13, while other trisomies such as trisomy 16 and trisomy 22 are rarely, if ever, encountered at birth. The decreased viability of fetuses with these chromosome abnormalities is believed to result from an increased amount of genetic imbalance, as compared with those with viable chromosomal abnormalities. Clinically, those conditions with the greatest quantity of missing or extrachromosomal material also have the most severe phenotypic manifestations. As a general rule, abnormalities of the autosomal chromosomes (the chromosomes not involved with sex determination) are characterized by significant developmental disabilities and birth defects, and sex chromosome abnormalities (involving the X and Y chromosomes) are characterized by abnormalities in sexual differentiation, maturation, and fertility.

The most common chromosome abnormality in humans, and the most studied, is Down syndrome. Down syndrome is caused by an extra chromosome 21 and is often referred to as trisomy 21. Clinically, individuals with Down syndrome have characteristic facial features, mild to moderate developmental disabilities, and variable birth defects such as cardiac abnormalities in 40% to 50%, gastrointestinal abnormalities in 1% to 2%, and other abnormalities of most organ systems. About 95% of individuals with Down syndrome have 47 chromosomes, and the remaining 5% either have a chromosome translocation (an extra copy of chromosome 21 attached to another chromosome, usually a chromosome 14) or mosaicism (the presence of two cell lines: one with the normal chromosome complement and another with trisomy 21). As with other autosomal trisomies, its relative incidence increases with older maternal age. While the general population probability of having a child with Down syndrome is about 1 in 700, the age-specific risk varies from about 1 in 1,500 in women below age 20 years to about 1 in 50 for women at age 45 years. Although the relative incidence of Down syndrome is increased in older mothers, about 80% of affected children are born to women younger than 35 years of age. This is because the sheer number of pregnancies in women younger than 35 years is so much greater than those of older women. As a result of advances in obstetrical practice beginning in the 1970s, women ages 35 years and older have been offered amniocentesis for prenatal detection of Down syndrome. Amniocentesis involves withdrawing amniotic fluid from around the developing fetus and culturing fetal cells for chromosome analysis and potentially for other analyses. Subsequent advances in detection of maternal serum markers during pregnancy have brought about additional changes in obstetric practice. For women in developed countries, maternal serum screening and prenatal ultrasonography are often used to identify pregnancies at increased risk for Down syndrome, trisomy 18, and birth defects such as open spina bifida, anencephaly, and defects of the abdominal wall. Prenatal ultrasonography is also used routinely and in high-risk pregnancies for detection of birth defects and birth defect syndromes. When chromosome abnormalities or serious birth defects are detected antenatally, therapeutic abortion may be elected. There are currently no primary prevention strategies for chromosome abnormalities or for most birth defects.

Single Gene Disorders

Single gene disorders are those conditions that follow the rules of inheritance first identified by Mendel in the mid-19th century. They comprise many familiar medical conditions such as cystic fibrosis, sickle cell anemia, Tay-Sachs disease, hemophilia, and Duchenne muscular dystrophy. Single gene disorders are individually very rare, and as a group, they still affect less than 1% of the general population. Nevertheless, they are the targets of long-standing and intensive study for several reasons: (1) Because their inheritance patterns are readily recognizable, they were among the first human disorders known to have their origins in single genes; (2) early detection and medical intervention have resulted in virtual cures or dramatic improvements in the well-being of some affected individuals; and (3) knowledge of how single gene abnormalities lead to human disease has served

as a paradigm for studying interactions between susceptibility genes and environmental factors that lead to more common diseases.

At each autosomal gene locus, there are two alternative copies, known as alleles, one inherited from each parent. When mutation of only one of these alleles produces disease, the disorder is said to be an autosomal dominant trait. Although most people with autosomal dominant diseases inherit the mutated allele from an affected parent, spontaneous mutation may also produce an autosomal dominant disease in the offspring of unaffected parents. Advanced paternal age, usually considered to be 40 years old or greater, is associated with an approximately 1% risk for a spontaneous mutation that results in the offspring having an autosomal dominant genetic disorder. The reason for this increased risk among older fathers and not mothers is unknown, but its manifestation may be striking. For example, almost all spontaneous mutation for many well-characterized genetic disorders occurs only in the copy of the gene contributed by the father. Advanced paternal age is also a risk factor for spontaneous mutation of genes carried on the X-chromosome. In this circumstance, spontaneous mutation of a gene contributed by an older father to his daughter results in her being an unaffected carrier. Her sons are then at risk for inheriting the mutation and manifesting an X-linked genetic disease.

In autosomal recessive disorders, the disease arises only in people who are homozygotes, which means they possess two mutant alleles. People who have only one mutant allele show no disease manifestations and are commonly referred to as carriers, or heterozygotes. It has long been recognized that certain genetic conditions are more common in consanguineous matings, wherein the parents are related through a common ancestor. The mechanism by which a homozygote receives two mutant alleles from the common ancestor is known as homozygosity by descent. This increased risk of autosomal recessive disease has given rise to a taboo against marrying one's blood relatives in many, though by no means all, cultures. For cultures who promote consanguineous matings, the most common pairing is among first cousins. Although the worldwide burden of genetic disease is relatively small as a result of consanguinity, in families who have a high percentage of carriers for a mutant allele, autosomal recessive disease may be particularly deleterious. General estimates suggest that among all first-cousin matings, there is an approximately 1% increased risk for autosomal recessive disease, when compared with the general population.

Genetically isolated populations are also at increased risk for autosomal recessive genetic disease. Isolated populations are groups that tend to mate exclusively within their group because they are bound by geography, culture, religion, and/or language. The genetic disease that runs within such populations is usually the result of a founder effect, in which a mutation is introduced into the group by a single founder and is propagated throughout the population. Because there is a high carrier rate in the population, the risk of that disease is much higher than in the general population. Founder effects are known to exist for conditions such as Tay-Sachs disease in Ashkenazi Jews and French Canadians, gyrate atrophy in the Finnish population, and variegate porphyria in Afrikaners from South Africa. The high prevalence of genetic disorders in some ethnic groups has led to the development of carrier screening programs to identify people who are at risk for having an affected child. These programs provide members the opportunity to make informed reproductive choices, based on specific knowledge about the individual's carrier status.

Of the approximately 30,000 genes in the human genome, almost 2,000 human diseases are known to be the result of mutations in single genes. The past decade has seen a dramatic increase in knowledge regarding the genetic origin of many clinically recognizable syndromes, the phenotypic spectrum of specific genotypic abnormalities, and the diagnosis of these syndromes through molecular techniques. From a pathogenetic perspective, most of these genetic mutations produce abnormalities in the quantity, structure, and/or function of encoded proteins. The role of the affected protein determines the disease phenotype. Abnormalities of structural proteins are most often associated with birth defects or poor tissue integrity, whereas abnormalities of enzymes or transport proteins usually produce disease through accumulation of deleterious precursors or a deficiency of the end products. These enzyme abnormalities are the focus of newborn screening programs and represent the targets of some of the world's most successful treatment programs for genetic disease. The standard medical treatment for enzyme defects has historically involved dietary or pharmacologic intervention to limit accumulation of abnormal precursors or to provide supplementation of the deficient end product. There are, however, an increasing number of enzyme abnormalities that are being treated by enzyme

replacement therapy (ERT). This type of treatment relies on the intravenous infusion of an artificially produced but bioactively equivalent enzyme preparation that replaces the defective enzyme. The greatest role for ERT has been in the treatment of disorders in which there is abnormal storage of enzyme precursors and progressive organ dysfunction. Beginning with the 1991 approval of a modified form of the enzyme beta-glucosidase for the treatment of Gaucher disease, the number of disease targets for ERT has expanded to include Fabry disease, mucopolysaccharidosis type 1, and Pompe disease. ERT has the potential to eliminate the deleterious effects of any disease that results from enzyme dysfunction.

Multifactorial Conditions

Multifactorial diseases result from a combination of both genetic and environmental factors and represent some of the most common human diseases such as atherosclerotic cardiovascular disease, adult onset diabetes, and birth defects. For most of these diseases, the genetic basis is inferred from studies of familial clustering, but the causative genetic loci are generally unknown. One of the most common methods for calculating the relative contributions of genetics and environment for a multifactorial trait is the relative risk ratio. This ratio is derived by dividing the risk for a disease in a relative of someone with the disease by the risk for a disease in someone from the general population. If the disease is completely unrelated to genetic constitution, the relative risk ratio is 1; and, for increasingly greater genetic contribution, the ratio increases proportionally. For example, a disease with a minor genetic contribution such as juvenile onset diabetes has a sibling relative risk ratio of about 10, whereas a condition with a substantial genetic contribution such as childhood autism has a ratio of > 100. There are a number of other methods by which the heritability (an estimate of the amount of phenotypic variation that is genetic) of a disorder can be quantified. Although these methods are generally useful for quantifying the relative genetic contribution to the disorder, one should be cautious about inferring a genetic causation from relative risk ratios and similar calculations, because in addition to sharing a similar genetic constitution, related individuals often share a similar environment.

Multifactorial disorders are excellent targets for primary prevention strategies aimed at modifying environmental risk factors that contribute to their causation. Some of the great public health successes of the previous century focused on the link between environmental factors and specific multifactorial diseases. Decreases in smoking have resulted in decreased rates of lung cancer and cardiovascular disease; and folic acid fortification and dietary supplementation have been accompanied by dramatic decreases in the occurrence of spina bifida and anencephaly. Identification of genetic risk factors may also be useful for designating a high-risk group that might benefit from general prevention strategies, surveillance, and early treatment. For example, individuals who are homozygous for an abnormality of a gene encoding the protein alpha-1 antitrypsin are particularly susceptible to the pulmonary effects of smoking and can develop a severe form of emphysema if they smoke. Smoking avoidance greatly reduces these risks. Much of our current knowledge about the relative contribution of genetic factors to multifactorial disease has arisen from association studies. These studies examine the frequency of a disease among two different groups: those who carry a genetic variant and those who do not. Such studies have identified an increased risk for Alzheimer's disease in people who carry an allele for apolipoprotein E, for spina bifida in the offspring of women with a thermolabile variant of the methylene tetrahydrofolate reductase enzyme, and for asthma in people with alleles for more than a dozen different loci. The next public health challenge for primary and secondary prevention of multifactorial diseases will be finding ways to influence the expression of these genes or their gene products.

Conclusion

Many genetic disorders are individually rare, but taken as a group, they contribute greatly to the human disease burden. For chromosome and single gene disorders, modifiable risk factors are few, and primary prevention represents a great challenge. Established carrier detection programs will continue to provide informed choices in at-risk populations; and new treatments such as ERT hold promise for secondary prevention of selected genetic diseases. For multifactorial disorders, the identification of complex interactions among environmental and genetic determinants provides opportunities for prevention or amelioration of some of the most common human diseases. As

morbidity and mortality from historically common disorders such as infectious diseases decrease, the treatment and control of genetic diseases will become increasingly important.

—Christopher Cunniff

See also Association, Genetics; Birth Defects; Chromosome; Family Studies in Genetics; Gene; Gene-Environment Interaction; Genetic Epidemiology; Genetic Markers; Genotype; Molecular Epidemiology; Mutation; Phenotype

Further Readings

Baird, P. A., Anderson, T. W., Newcombe, H. B., & Lowry, R. B. (1988). Genetic disorders in children and young adults: A population study. *American Journal of Human Genetics, 42*, 677–693.

Crow, J. F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics, 1*, 40–47.

Khlat, M., & Khoury, M. J. (1991). Inbreeding and diseases: Demographic, genetic and epidemiologic perspectives. *Epidemiologic Reviews, 13*, 28–41.

Khoury, M. J., Beaty, T. H., & Cohen, B. H. (1993). *Fundamentals of genetic epidemiology.* New York: Oxford University Press.

Mantipragada, K. K., Buckley, P. G., de Stahl, T. D., & Dumanski, J. P. (2004). Genomic microarrays in the spotlight. *Trends in Genetics, 20*, 87–94.

Nussbaum, R. L., McInnes, R. R., & Willard, H. F. (2004). *Genetics in medicine* (6th ed.). Philadelphia: W. B. Saunders.

Rimoin, D. L., Connor, J. M., & Pyeritz, R. E. (2001). *Emery and Rimoin's principles and practice of medical genetics* (4th ed.). Philadelphia: W. B. Saunders.

Roizen, N. J., & Patterson, D. (2003). Down's syndrome. *Lancet, 361*, 1281–1289.

Stevenson, R. E., Allen, W. P., Pai, G. S., Best, R., Seaver, L. H., Dean, J., et al. (2000). Decline in prevalence of neural tube defects in a high-risk region of the United States. *Pediatrics, 106*, 677–683.

Genetic Epidemiology

Genetic epidemiology is an emerging field that developed initially from population genetics, specifically human quantitative genetics, with conceptual and methodological contributions from epidemiology. One of the early proponents of genetic epidemiology, Morton, defines the field as one that addresses the etiology, distribution, and control of disease in groups of related individuals and the inherited causes of diseases in populations. This definition has by necessity been broadened to include the role of the environment by others who emphasize the role of genetic factors as they interact with environmental factors in the occurrence of diseases in human populations. Khoury, Little, and Burke (2003) have recently coined the term *human genome epidemiology* to encompass a system of investigations that use the methods of epidemiology in population-based studies of the influences of genetic variation in both health and disease.

It should be noted that the field of molecular epidemiology stands in contrast to genetic epidemiology, as the former grew out of environmental epidemiology. The rationale for the emergence of molecular epidemiology was the need to identify biomarkers of environmental exposures as an application of molecular biology in epidemiology.

There are a number of aspects that distinguish genetic epidemiology from other areas of genetics from which it developed. The first is the population-based nature of the research, which, together with shared methodological approaches, is one of its key links with epidemiology. Second, newer ways of conceptualizing the field stress the search for combined and interactive effects of genetic and environmental factors. Finally, genetic epidemiology includes the consideration of the biological basis of the diseases into developing models of causation for diseases.

The goals of modern genetic epidemiology have been broadened to include all diseases, whether they are common and complex or supposedly simpler, such as the so-called monogenic disorders. There has been a tendency in recent years in genetic epidemiology to place almost exclusive emphasis on the complex diseases, but some of the best advances in the epidemiology of genetic diseases have indeed involved simple inherited disorders, which are increasingly seen as more complex as we come to a better understanding of epigenetics and interactions of disease genes with environmental factors. In addition, any definition of genetic epidemiology arguably needs to encompass all aspects of the epidemiology of genetic diseases, including studies of prevalence, clinical epidemiology, genotype-phenotype relationships, and disease outcomes and progression. We also need more comprehensive studies of environmental factors that influence outcomes in genetic diseases.

Historically, the field of genetic epidemiology has some of its roots in the interests of medicine

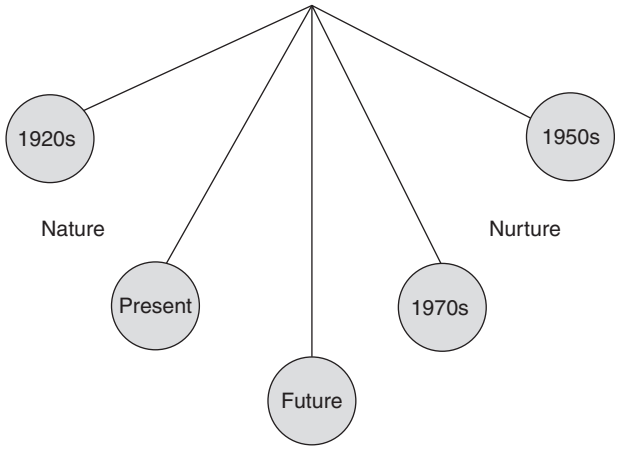


Figure 1 Swings of the Nature-Nurture Pendulum

Source: Adapted from Plomin and Petrill (1997). Copyright Elsevier; used with permission.

concerning the causes and heredity of disease. In the days prior to the field actually having a label (i.e., prior to the 1950s), scientists who perhaps would now be labeled as early genetic epidemiologists were trying to unravel the issues of nature and nurture with respect to human diseases. These activities were in contrast to the early practitioners of medical genetics who tended to be oriented toward the clinical and descriptive aspects of what were recognized as diseases with potential genetic involvement and to genetic counseling based on what was known then about the inheritance patterns of some diseases. Early practitioners of genetic epidemiology often looked for associations between diseases such as stomach ulcers and well-studied genetic traits of the day such as blood groups. One of the authors (F. J. M.) took a genetics course in the early 1960s that had the name "Heredity and Eugenics." Interestingly, a faculty member with a medical degree (MD) taught the course, most of which was spent examining what was known then about the genetics of human diseases. This was a time period when the field of medical genetics was in its childhood and genetic epidemiology had barely taken hold, and the legacy of eugenics was still alive in the names of journals and organizations.

The early history of attempts to solve the problem of heredity versus environment, that is, nature versus nurture, in causing human diseases was subject to numerous swings in the so-called nature-nurture pendulum over the 20th century (Figure 1). Both scientific and societal concerns have driven these swings over time. The swings reflect periods of time during which the scientific community and society viewed either nature or nurture as the more important determinant in causing diseases.

In modern times, the controversy about nature versus nurture has been gradually replaced by a view that includes roles for both nature *and* nurture in human disease in that both genetic and environmental factors influence (frequently according to an interactive model) disease susceptibility.

Genetic epidemiology is becoming an increasingly important field in science today. While epidemiologists increasingly recognize the importance of genetic determinants of human diseases, human geneticists are learning about the importance of environmental exposures in all human diseases. There is increasing acceptance of the assertion by Khoury et al. (2003) that most human disease, if not all, is the result of the interaction between underlying genetic susceptibility and exposures to various components of the environment, including chemical, dietary, infectious, physical, and behavioral. The latter would encompass the influences of cultural factors in human behavior and their interaction with other environmental factors.

While human geneticists and epidemiologists continue their search for optimal strategies to identify disease genes and overcome some of the present methodological limitations in dissecting the genetic components of complex diseases, there is little doubt that genetic association studies will become increasingly important in the translation of the results of genomic research into public health. Genetic epidemiology in this sense will continue to increase in its importance in both human genetics and epidemiology.

One of the most important contributions that epidemiology has made to genetic epidemiology is probably the population-based approach. There are obvious advantages of epidemiological methods that can be applied in genetic epidemiology research, such as case-control studies. However, there are some aspects that have yet to be fully appreciated. Some of these include the knowledge base created by prevalence studies done prior to analytical or interventional research. Knowing the distributional characteristics of socioeconomic and other demographic variables provides an effective basis for the evaluation of the samples of cases obtained for clinical description and interventional research. In addition, population-based data on clinical features and outcomes in genetic

disorders have obvious advantages over data gathered on the basis of attendance in specialized clinical centers that may not reach all strata in a population of individuals with the disorder of concern.

Some of the important approaches used today in genetic epidemiology are covered in more detail elsewhere in this *Encyclopedia* and therefore will not be covered in detail here. For genetic disorders inherited in a Mendelian fashion, suffice it to say that linkage studies have been the method of choice for many years. Unfortunately, to date, the same methods have not shown the same overall success for multifactorial diseases in which multiple genes and environmental factors are each assumed to play roles in the disease etiology. These complex diseases account for the vast majority of all human diseases and therefore have received increased attention during the past two decades.

Some have touted association studies as being more powerful than other methods in the detection of susceptibility genes for multifactorial diseases. However, it has not always been possible to replicate the findings of such studies in a consistent manner. The major problem is that in any given multifactorial disease, we may be dealing with genetic loci in the thousands, thereby decreasing the probability of finding an association with a specific gene.

There is some agreement among genetic epidemiologists that there should be evidence for a role of genetic factors in a disease before launching into large-scale studies to identify genes. Twins studies have been a tried-and-true method for establishing such evidence in many multifactorial diseases. Family studies of disease risk in first-degree relatives are also considered preliminary to further research. These should include comparisons of familial risk to risk in the population (i.e., prevalence) and, if controls are available as in case-control studies, to the relatives of control individuals. This is done by computing λ, which is the ratio of disease risk in the relatives of cases (those affected with the disease in question) to the prevalence for the disease in the population.

Another important contribution of epidemiology to the development of genetic epidemiology has been the introduction of traditional epidemiological methods to the mix of study designs. Human quantitative genetics used studies of twins, sibs, half-sibs, and in some cases adoptees to investigate the genetic and environmental sources of variation in human traits and diseases. Application of epidemiological methods such as case-control and cohort studies has further extended these methods for examining genetic and environmental risk factors simultaneously. In any of the methods used, the emphasis on a population-based approach becomes important, whether it is the collection of prevalence data in surveillance programs or enhancing our ability to test how representative a specific sample is through an understanding of the characteristics of the larger population. Many genetic epidemiologists are promoting the use of prospective cohort studies such as the National Children's Study in the United States and other longitudinal research worldwide for their potential to contribute to studies of associations between genes and disease and gene-environment interactions.

If we take into consideration the broad view of genetic epidemiology that includes all genetic diseases, there have been major achievements in the identification of the genetic causes of many genetic disorders during the past 25 years, especially with improving techniques in cytogenetics and molecular genetics. Following the discovery of restriction fragment length polymorphisms in the late 1970s, there came a period of identification of the genes for many genetic disorders. Further advances in molecular genetic techniques have rapidly increased our knowledge of the genes for most monogenic disorders. What remains is the ongoing research on interactions of these major disease genes with other genes and environmental factors as they influence the multitude of disease outcomes in these disorders.

To date, we have not seen the important promised breakthroughs in establishing definitive risk factors for complex traits and diseases, although there are some exceptions to this statement such as the relationship of alpha-1-antitrypsin deficiency to chronic obstructive pulmonary disease. The genetic dissection of complex diseases is proving much more challenging for a number of reasons. The magnitude of any specific genetic effect is usually very small. The involvement of multiple genes alone makes the task more complicated. The increasing evidence for gene-gene and gene-environment interactions complicates the task even further. In addition, the possible modifying effects of phenotypic heterogeneity and developmental processes on genetic associations in complex diseases may not have been taken into account sufficiently in most association studies to date.

Khoury et al.'s (2003) definition of human genome epidemiology increases the scope of traditional

genetic epidemiology by emphasizing the application of epidemiological methods to evaluate the impact of genetic variation on human health and disease. It reflects the new vision of genomics and its expected influence in public health. This will likely have an important impact on the public health burden of common complex diseases. However, the epidemiology of chromosomal and Mendelian genetic disorders also needs to be given continued emphasis under the umbrella of genetic epidemiology. Before we had genomics, there were many important contributions of epidemiological studies of genetic disorders to public health. One of these involved newborn screening, which is likely the most successful public health genetics program to date. A clinical intervention study conducted during the 1980s on sickle cell disease paved the way for expansion of newborn screening programs in the United States to include this genetic disease in the programs of more states. The study was a randomized trial to determine the prophylactic effects of penicillin administration in children diagnosed with sickle cell disease. A reduction in both the incidence of pneumococcal infections and the mortality in those infants who received penicillin compared with the untreated group led to the conclusion that children should be screened in the newborn period for sickle cell disease and receive penicillin early in the first year of life to prevent the infections. This is one of the few instances of the outcome of a clinical epidemiological study driving changes in genetic health services delivery. It suggests that one goal of genetic epidemiology ought to be to produce data that will drive changes in health care services and public health to reach the population at large.

The goals of genetic epidemiology need to be flexible, because new discoveries in genetics and genomics bring different ways of viewing and analyzing the data. For example, some now have proposed that in effect, there are no actual monogenic diseases. This view is supported by the increasing evidence of the effects of interacting genes and environmental factors on the phenotypic outcomes in single gene diseases. If we consider these complications in phenotypic expression, all diseases are complex, although some might be complex and not all that common. The message is that increasing complexity means increasing need for the approaches of genetic epidemiology and the accompanying statistical methods that have traditionally been applied to deal with these complexities. It also means that as the genetic components of human diseases are unraveled, there will be a need for effective ways of translating genetic findings into public health improvements, accomplishing the counseling to incorporate genetic risk assessments, increasing public understanding of the role of genetics in disease, and in turn, developing culturally competent ways of implementing these programs and delivering these messages.

—F. John Meaney and Stefano Guerra

See also Association, Genetic; Eugenics; Gene-Environment Interaction; Genotype; Linkage Analysis; Molecular Epidemiology; Phenotype

Further Readings

Burton, P. R., Tobin, M. D., & Hopper, J. L. (2005). Key concepts in genetic epidemiology. *Lancet, 366*, 941–951.

Davey Smith, G., Ebrahim, S., Lewis, S., Hansell, A. L., Palmer, L. J., & Burton, P. R. (2005). Genetic epidemiology and public health: Hope, hype, and future prospects. *Lancet, 366*, 1484–1498.

Khoury, M. J., Little, J., & Burke, W. (2003). *Human genome epidemiology: A scientific foundation for using genetic information to improve health and prevent disease.* New York: Oxford University Press.

Merikangas, K. R. (2003). Implications of genomics for public health: The role of genetic epidemiology. In *Cold Spring Harbor Symposia on Quantitative Biology* (Vol. 68, pp. 359–364). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Plomin, R., & Petrill, S. A. (1997). Genetics and intelligence: What's new? *Intelligence, 24*, 53–77.

Genetic Markers

Genetic epidemiology aims to identify genetic variation related to risk for disease. Because it is currently not feasible to fully sequence the genomes of every person in a sample, the field has traditionally relied on genetic markers with known locations to act as surrogate information for the surrounding sequence. These markers are typically called polymorphisms to reflect the concept that they are locations in the genome with variability within and across individuals (i.e., they have multiple forms or "spellings"). The ability for markers to act as surrogates for surrounding sequence is a function of a genetic property called linkage and a related concept of linkage disequilibrium, which results in correlation between

polymorphisms and surrounding sequence. Because markers are often simply proxies for unmeasured sequences that can influence risk for disease, marker-based approaches are often termed *indirect association* studies. Emerging technology has greatly increased the catalogue of such variable sites in the human genotype and the ability to accurately and affordably genotype individuals at these markers, such that marker-based genetic epidemiology is now the paradigm for most studies.

The field of genetic epidemiology aims to identify genetic variation that is related to disease. This can be a daunting task, considering that the size of the human genome is around 3 billion base pairs. Finding a single genetic variant that influences risk for a disease would be like finding a single misspelled word among 3 billion letters. This is often compared with trying to find a single misspelling in an entire encyclopedia. While one could carefully read the entire encyclopedia to identify misspellings, this could take an enormous amount of time and many misspellings may simply be overlooked. Furthermore, technology has traditionally limited the ability to sequence ("read") the entire genome of each participant in a study. Instead, the field has relied on genetic "markers" located at known locations in the genome to represent the surrounding sequence. Continuing the encyclopedia example, this would be like marking the first three sentences of every entry, so that the specific location in the context of the encyclopedia is known.

Genetic markers are "polymorphisms," meaning they contain variable sequence (literally "multiple forms") within and across individuals of a population. The most common types of markers in the human genome are single nucleotide polymorphisms (SNPs), simple tandem repeat polymorphisms (STRs), and insertion/deletions (indels) (see Figure 1). SNPs are defined by the existence of more than one nucleotide at a particular position in the genome. For example, at a genomic location with the sequence **ACCTGA** in most individuals, some may contain **ACGTGA** instead. The third position in this example would be considered an SNP with either a C or G allele. Because each individual inherits one copy of their genome from each parent, each person has two copies, and therefore three types of individuals can be distinguished based on this polymorphism: those with two copies of the C allele (homozygous CC

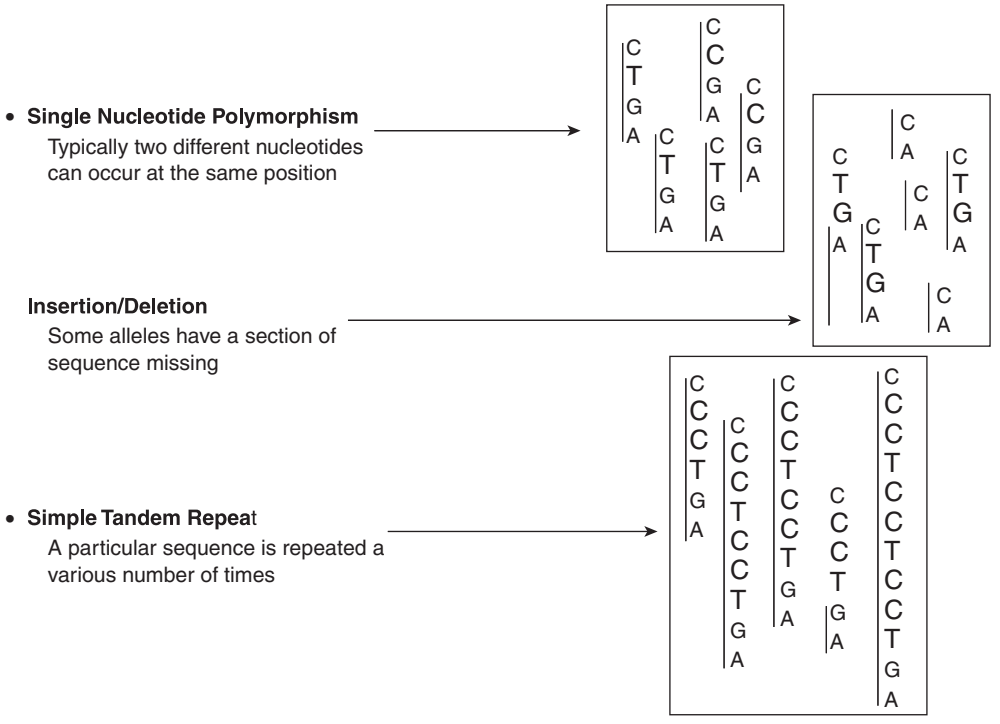


Figure 1 Common Genetic Markers

genotype), those with one C and one T allele (heterozygous CT genotype), and those with two T alleles (homozygous TT genotype). The three genotype groups of this marker can be used as "exposure" categories to assess association with an outcome of interest in a genetic epidemiology setting. Should such an association be identified, researchers may investigate that "marked" genomic region further to identify the particular DNA sequence in that region that has a direct biological effect on the outcome of interest.

STRs are markers where a piece of sequence is repeated several times in a row, and the number of repeats (considered an allele) is variable within and across individuals. For example, a CCT pattern may be repeated up to 10 times, such that individuals in the population may have genotypes at that locus representing any combination of two repeat alleles of sizes 1 to 10 repeats (e.g., $10(10+1)/2 = 55$ different possible genotypes). Indels refer to polymorphisms where a piece of DNA sequence exists in some versions (insertion allele) and is deleted in others (deletion allele) in the population. While each of these additional types of markers are used in genetic epidemiology, the SNP marker is the most common (approximately 10 million in the human genome), is arguably the most affordable to genotype in large samples, and is therefore the most widely used type of marker.

The marker-based approach to genetic epidemiology has often been considered an "indirect" approach, since markers are used as surrogate information for unmeasured sequence variation that directly influences disease risk. While "direct" approaches, in which the polymorphism with known biological consequence is directly measured and related to disease, would be ideal, it is often not possible. Even for genes with well-known function, the particular sequence variation with biological consequences may not be known. Polymorphisms that change the amino acid sequence or the rate of transcription are great candidates, but knowledge about all functional consequences for most sequence variation does not yet exist. Therefore, to maximize the detection of disease-related sequence variation, the field relies on marker-based approaches that allow markers to represent the rest of the genome that is not directly measured.

This "indirect association" approach assumes that markers will provide adequate surrogate information for surrounding sequence that may contain unmeasured disease-related alleles. This would not be possible if correlation did not exist between alleles located close to each other in the genome. Fortunately, such spatial correlation does exist naturally, through the Mendelian property of genetic linkage and the population genetic consequence of linkage that is often called linkage disequilibrium (LD). Under linkage equilibrium, alleles should be independent of each other (no correlation), but for very closely located polymorphisms, correlation does exist due to population genetic

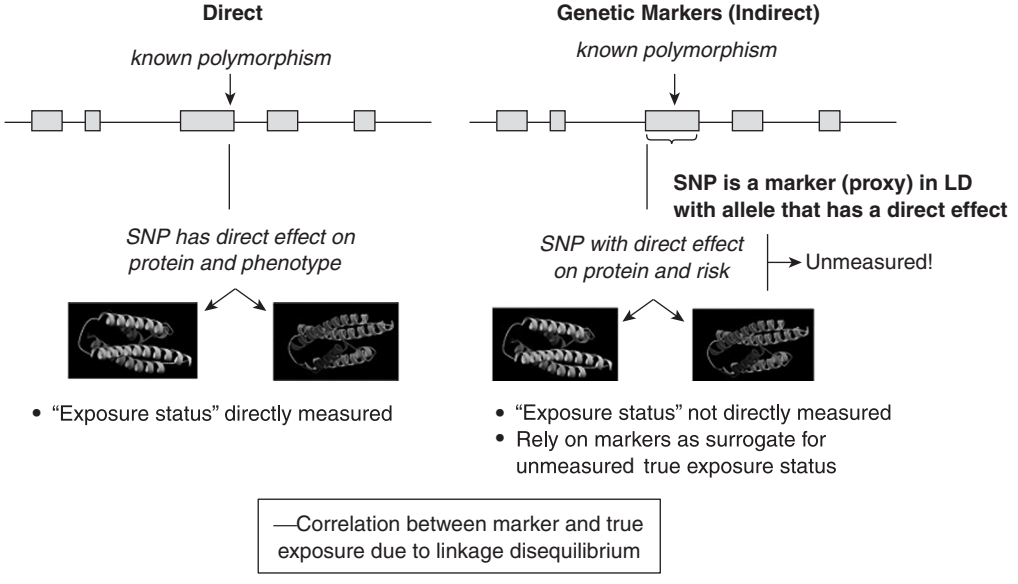


Figure 2 Direct Versus Indirect (Marker-Based) Association

history, and this departure from equilibrium is called linkage disequilibrium. LD patterns have been characterized across the genome in several populations and show wide variability in the length of DNA that shows correlation. However, on average, LD exists in "blocks" of DNA of about 22,000 base pairs in length in populations of European descent. This pattern depends on the genome region and population, but once characterized, genetic markers can be chosen to represents the surrounding sequence within a correlated LD block. The set of markers that most efficiently represents the entire correlated set of sequence (often SNPs) is called "tagSNPs." Various calculations have been proposed to determine the most efficient set of SNPs to "tag" an area of the genome, but in general, these should be chosen based on LD information of the genomic region and population relevant to a particular study sample, with an aim to maximize the proxy information available when using just the tagSNPs compared with all polymorphisms known in the region.

—Margaret Daniele Fallin

See also Gene; Genetic Disorders; Hardy-Weinberg Law; Linkage Analysis

Further Readings

Cordell, H. J., & Clayton, D. G. (2005). Genetic association studies. *Lancet, 366*(9491), 1121–1131.

The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature, 437*, 1299–1320.

Genocide

Genocide is the intentional destruction of human groups, in whole or in part, by mass killing and other methods. Such, at least, is a shorthand definition based on the United Nations Genocide Convention of 1948. The Convention capped two decades of scholarly work and activist endeavor by Raphael Lemkin (1900–1959), a Polish-Jewish jurist troubled by the failure of international society to suppress atrocities inflicted by states against their own minority populations. Lemkin had a vision of cultural bonds and collective identities as essential to human civilization, and thus his framing of "genocide"—combining the Greek *genos* (race, tribe) and the Latin *cida* (killing)—downplayed the physical killing of individuals, highlighting instead the destruction of communal integrity and identity. This emphasis survives in contemporary conceptions of "ethnocide" and "cultural genocide." It is also reflected in the UN Convention, which defined genocide as "any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such":

1. Killing members of the group
2. Causing serious bodily or mental harm to members of the group
3. Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part
4. Imposing measures intended to prevent births within the group
5. Forcibly transferring children of the group to another group

The Convention definition is notable, indeed notorious, for its lack of specificity in key areas. How to define the groups covered "as such" by the Convention (national, ethnic, racial, and religious), and why are other groups excluded (e.g., those united by political belief, social class, or gender)? Can groups be "destroyed" by means other than mass killing, for example, by the infliction of "mental harm"? What "part" of the group must be destroyed to qualify as genocide? And how might a genocidal "intent to destroy" be ascertained?

In part because of these ambiguities, the concept of genocide languished for over two decades after the Genocide Convention entered into force in 1951. The renewal of interest in genocide can be traced to two main factors. First, the trial and execution of the captured Nazi Adolf Eichmann, in Israel in 1961 to 1962, spawned a flood of research and commentary on the Jewish Holocaust—for many, still the paradigmatic case of genocide. The analyses gradually assumed a comparative bent, as scholars became interested in other cases of genocide, such as the destruction of the Armenian population of Ottoman Turkey during World War I. Following the publication of Leo Kuper's seminal 1981 work *Genocide: Its Political Use in the Twentieth Century*, a field of "comparative genocide studies" gathered steam, hitting full stride in the latter half of the 1990s and into the 2000s. In part, this reflected the second key factor in the renewed

prominence of genocide for governments and publics: the continued prominence of the phenomenon itself in the post–Cold War era. This was brought devastatingly home by the apocalyptic slaughter of nearly 1 million Tutsis in Rwanda in 1994, and by the less destructive—but European-centered, and hence heavily publicized—mass atrocities against Bosnian Muslims following Yugoslavia's collapse in 1991. Both genocidal outbreaks prompted the formation of international criminal tribunals to try alleged perpetrators. Subsequent initiatives included "mixed tribunals" of national and international judges to preside over tribunals for Sierra Leone, and most recently for atrocities in Cambodia under the Khmer Rouge regime (1975–1979). The new International Criminal Court (ICC) also includes genocide in its jurisdiction.

From the outset, genocide scholars and activists have been motivated by a perceived need to *prevent* genocide as well as to *punish* it. "Early-warning" mechanisms have been devised, along with interventionist strategies aimed at nipping genocide in the bud. Emphasis has often been placed on interventions by individual nations or international (usually regional) alliances. In fact, all major cases of genocide suppression have featured military interventions by national armies or regional bodies. These include the Allies' defeat of Nazi Germany; India's intervention in East Pakistan/Bangladesh in 1971; Vietnam's in Cambodia in 1978 to 1979; Tanzania's in Uganda in 1979; the NATO and EU-led initiatives in Bosnia and Kosovo in the 1990s; and the Australian-led peacekeeping force (under the UN aegis) that brought an end to genocidal violence in East Timor in 1999. Attempts to globalize such interventions have had some success, notably through UN-sponsored peacekeeping and postconflict peace building. But more ambitious mechanisms, such as the "international peace army" proposed by genocide scholar Israel Charny (1999, p. 650), have so far foundered on a lack of political will.

From an epidemiological perspective, several core issues arise in the study and prevention of genocidal outbreaks. A central challenge is to decide whether a large death toll (even if only relative to group population) is by definition an element in genocide, or whether nonfatal and nonphysically injurious phenomena, such as cultural genocide, should also be included. For the most part, cases of alleged genocide have failed to arouse much interest among governments, international institutions, and modern publics, if mass killing is absent. An exception is the forcible transfer of Aboriginal children to white families in Australia, which a commission of inquiry found to be genocidal under the UN Convention definition. Another example, more amenable to an epidemiological framing, is the growing acceptance in international law of sexual violence against women as not only a concomitant of genocide but genocidal in itself. This interpretation was buttressed by the outbreaks of mass rape in the Balkans in the early 1990s—and in Rwanda in 1994, where women and girls were not only physically violated but often infected with the HIV virus by their attackers.

To the extent that mass killing *is* considered essential to genocide, debate inevitably arises over the scale of the killing. To cite just one instance, estimates of Bangladeshis killed in 1971 range from a low of 200,000 to a high of 3 million. Population data—notably censuses carried out before and after a genocide—can help to narrow the range of estimates. Such data have been crucial to determining the number of victims of Joseph Stalin's regime in the USSR (1928–1953) and Mao Zedong's in China (1949–1976). Where genocidal regimes have collapsed, such as the Nazis in Germany and communist rule in the Soviet Union, the documentary record left by the perpetrators has also been vital to evaluating death tolls and degrees of intention in the killing.

An especially significant epidemiological issue is the place of disease, and the structured undermining of public health, in genocidal outbreaks. Many scholars and other commentators have been reluctant to apply the term *genocide* to the destruction of indigenous communities, in the Americas and elsewhere, following Western colonial invasion and occupation. The key sticking point has been the prominence of disease as a mechanism of demographic collapse, which is sometimes presumed to rule out genocidal intent. More recent analyses, however, have stressed the interaction of disease with other factors, such as colonial assaults on the indigenous land-based and nutritional resources, as well as debilitating psychological trauma caused by the destruction of long-established social networks and cultural practices.

A similar conundrum arises when we consider cases of mass famine and material privation. The cases most commonly deployed in the genocide studies literature include (in chronological order) Ireland in the 1840s; the great famines in British India in the late 19th and early 20th centuries; the deaths of

hundreds of thousands of Germans under Allied blockade during, and especially after, World War I; the massive mortality suffered in Ukraine and Soviet Central Asia following the imposition of agricultural collectivization under Stalin (1932–1933), which Robert Conquest alleged was employed by the Soviet regime as a means of crushing Ukrainian nationalism; the famine, probably the largest in history, associated with Mao Zedong's "Great Leap Forward" in China (1958–1961); and repeated waves of starvation in Ethiopia under the dictatorial Dergue regime of the 1970s and 1980s.

In all these cases, debate centers on the degree of intention of the alleged perpetrators and the power they possessed, but chose not to employ, to remedy conditions of famine, pervasive malnutrition, and/or disease epidemics. While many scholars and policymakers are reluctant to stretch the genocide framework this far, it is notable that the hundreds of thousands of Jews who died from starvation and disease in the ghettoes and concentration camps of Nazi-occupied Europe are routinely numbered among the roughly 6 million victims of the Jewish Holocaust.

A final issue worth considering from an epidemiological perspective is structural violence as a form of genocide. Various commentators have held that the structure of global society—in which extreme poverty and curtailed life spans are the lot of hundreds of millions of people—constitutes a form of genocide against the poor. This is implicit in the 2005 statement by the UN Special Rapporteur on the Right to Food, Jean Ziegler, that "every child who dies of hunger in today's world is the victim of assassination" (Ziegler cited in "UN Expert Decries 'Assassination'," 2005, Online); Ziegler referred to the *daily* death of 100,000 people from starvation as a "massacre of human beings through malnutrition." Specific institutions, such as the World Bank and International Monetary Fund (IMF), stand accused of imposing "structural adjustment" programs that inflicted mass debility and mortality on vulnerable populations in Latin America, post-Soviet Russia, and elsewhere. Since the consequences of such measures were fairly predictable, and in any case were observable by those who imposed and then retained them, a *prima facie* case for genocide might be made—although such thinking remains quite far from the mainstream of genocide scholarship. In a similar vein, some have contended that the reluctance to supply lifesaving antiretroviral drugs to developing-world populations afflicted by AIDS constitutes intentional mass killing—hence genocide, or at least a crime against humanity. The UN Special Envoy for HIV/AIDS in Africa, Stephen Lewis, has stated that "those who watch [the AIDS pandemic] unfold with a kind of pathological equanimity must be held to account. There may yet come a day when we have peacetime tribunals to deal with this particular version of crimes against humanity" (cited in Mann, p. 61).

—Adam Jones

See also Health, Definitions of; Health Disparities; Violence as a Public Health Issue; War

Further Readings

Charny, I. W. (1999). An international peace army: A proposal for the long-range future. In Charny (Ed.), *The encyclopedia of genocide* (p. 650). Santa Barbara, CA: ABC-CLIO.

Gellately, R., & Kiernan, B. (Eds.). (2003). *The specter of genocide: Mass murder in historical perspective.* Cambridge, UK: Cambridge University Press.

Jones, A. (2006). *Genocide: A comprehensive introduction.* London: Routledge.

Kuper, L. (1981). *Genocide: Its political use in the twentieth century.* New Haven, CT: Yale University Press.

Mann, M. (2005). *Incoherent empire.* London: Verso.

Totten, S., Parsons, W. S., & Charny, I. W. (Eds.). (2004). *Century of genocide: Eyewitness accounts and critical views* (2nd ed.). New York: Routledge.

UN expert decries "assassination" by hunger of millions of children. (2005, October 28). *UN News Center*. Retrieved August 8, 2007, from http://www.un.org/apps/news/story.asp?NewsID=16407&Cr=food&Cr1=.

United Nations, Convention on the Prevention and Punishment of the Crime of Genocide. (1948). Retrieved August 8, 2007, from http://www.preventgenocide.org/law/convention/text.htm.

Genomics

An organism's genome could be likened to an enormous set of encyclopedias, and genomics is the science of reading and interpreting all the entries in the set. The entries in the encyclopedia could be equated with the genes encoded in an organism's genome. The genes contained within an organism's genome control every aspect of its life. They control growth, maintenance, and development, and for humans, the genome also includes information that controls

behavior, physiology, and the susceptibility to some diseases. Disease susceptibility or resistance is of great concern and interest in the areas of medicine and public health. Determining the interplay of genetic information and the detection, prevention, and treatment of disease is the combined job of epidemiologists and genome researchers.

A Genomics Primer

Genomics is a field that encompasses many areas of science, simply because it involves the smallest building block of life, DNA. Almost every cell within an organism contains its unique genome. Genomes are inherent in an individual and a species. Every individual or organism within a species has certain derivations of that species' genome. Genomes of individuals within a species are quite similar; for example, all humans are at least 99% genetically identical. Because of this close similarity, it is enough to study one or a few genomes within a species to glean information for the entire population.

Genes contain DNA, and chromosomes are made up of genes. Broken down even further, DNA contains the information to make RNA. RNA contains the information to make proteins. All these different levels of genetic structure make up a genome. DNA is made up of four similar chemicals called bases. These bases are repeated in a different order many times and, in the case of humans, enough times to make up approximately 3 billion base pairs. It is the difference in this order that accounts for the diversity within humans and between humans and other organisms. Within the 3 billion base pairs that make up the human genome, there are approximately 20,000 to 25,000 genes. Genes contain the information to make proteins, and RNA is the template that serves as the mediator in creating proteins from genes.

The cell uses the components of RNA as the building blocks of proteins. Proteins are composed of 20 different elements called amino acids. Proteins are involved in many different facets of the working human body. Proteins act as enzymes that are involved in metabolism. They can also be hormones that are involved in a process known as cell signaling. Cell signaling is a type of communication between and within cells as a response to the environment within the body. This communication guides cell actions.

Proteins also act as transporters. Examples of transporter proteins are antibodies that transport or bind antigens, foreign substances in the body, to the immune system for destruction. Hemoglobin, the protein found in red blood cells, transports oxygen from the lungs to other parts of the body. Proteins are also part of the structural components of the cell. They make up structural tissues such as that in muscles, collagen, hair, fingernails, and much more.

Because not every gene in a genome is recognizable, it is important to derive the entire sequence of a genome. Having the whole sequence of a genome will help scientists recognize what sequence is a gene and how or if that gene has a role in the function of other genes in the genome. Many areas of the genome do not contain genes. Some scientists claim that these areas are "junk" DNA, simply because these regions do not have a designated function or one is not yet known. Studies are being conducted to determine whether there is function associated with "junk" DNA and to propose reasons for why this sequence is retained in the genome.

Branches of Genomics

The advent of genomics and the Human Genome Project has broadened the scope of genomics by creating offshoots into new branches of learning. These other "-omics" disciplines include proteomics, transcriptomics, metabolomics, pharmacogenomics, and many more disciplines are developing as research continues and data are published. Proteomics is the study of the structure and function of proteins. Transcriptomics is the study of the degree of expression of all known genes in an organism, and metabolomics is the study of metabolites, or end products, that result from cellular processes. Pharmacogenomics studies the interaction of genetic variation on drug response and efficacy. The aim of pharamcogenomics is to develop personalized drug therapy based on an individual's specific genome to reduce adverse side effects and improve the response of the body to specific drugs. Apart from creating new disciplines, the Human Genome Project has influenced other scientific areas such as forensics, agriculture, anthropology, and epidemiology.

Genomic scientists analyze the function of genes within an organism to determine how genes create or influence phenotypic outcomes. These scientists have also begun the difficult task of examining the complexities of how the interaction of genes and environmental exposures is involved in the occurrence of disease. Epidemiologists, and especially molecular epidemiologists,

study how genetic as well as environmental factors influence disease etiology. Epidemiologists will apply this knowledge to determine the distribution of disease as well as to develop prevention plans to stem the occurrence of disease within families and whole populations. This knowledge will also help epidemiologists determine better methods for detection of disease and better treatments for disease conditions.

So how does epidemiology factor into the world of genomics? Genomics along with epidemiology will play an important role in deciphering the issues affecting the health of human populations. There are limitations in applying genomic data to human population health. Problems have resulted in translating the findings from the study of genomic data and applying those findings to medicine and public health practices. This is where epidemiology, which is the science and logic behind the institution of public health practice, will be essential. Many epidemiologists believe that epidemiological practices will be necessary to fulfill the promises of the Human Genome Project.

Four small chemicals in DNA contain the information controlling all the cellular actions in the body. Problems that arise during cellular actions as well as the effects that result from gene-environment interactions can lead to disease. The genome of human individuals differs by 1%. This difference or variation in single bases between individual genomes is called a single nucleotide polymorphism, or SNP.

To be considered an SNP, a variation must occur in that gene in at least 1% of the population. SNPs can occur anywhere in the genome, and they show up every 100 to 300 bases in the 3 billion base pairs human genome. These single base changes affect how individuals react to certain diseases; how they react to factors in their environment such as pathogens, chemicals, and toxins; and how individuals respond to treatment. A change in a person's genetic code does not necessarily cause disease in an individual, but it can give information as to the likelihood that an individual will develop a certain disease. Research is still being conducted regarding the effects of SNPs because when it comes to complex diseases such as cancer, diabetes, and heart disease, it is not one gene that is involved but many.

Another facet of genomic research is in the area of genetic disorders. Genetic disorders can result from mutations in a gene, abnormal chromosome number, multiple repeats of three bases, or the defective gene can be inherited from a parent. Mutations can occur at a single base location with an exchange of one base for another, insertions of one or more bases, or deletions of one or more bases. Mutations can also occur at the chromosome level with gene duplications or deletions of large regions of the chromosome, or portions of one chromosome can be translocated or displaced to another chromosome. Mutations, or changes in the sequence of genes, can cause errors when genes are translated into proteins, which in turn can create proteins that are partially functional or not functional at all. Improperly functioning proteins can play a part in the development of disease.

Mutations and SNPs all result in gene variants. Gene variants are essentially the same as other genes, but there are slight differences. Many studies are being conducted on gene variants within families and populations. Case-control studies on the incidence of disease in populations often provide a basis for further research on the discovery of new genes. Application of that data can be used to develop assumptions on how specific genes are related to disease in underlying populations.

Because of the huge mass of published data on genomics and molecular epidemiology, a network called HuGE Net has been established. HuGE Net, or Human Genome Epidemiology Network, is a global collaboration that aims to provide information on human genomic variation and health and on the quality of genetic tests for screening and prevention. The HuGE Net includes journal clubs, journal reviews, fact sheets, case studies, newly published articles, and links to informative Web sites. HuGE Net will be a useful tool for the public as well as scientists in making sense of all the information available on genomics and health.

Epidemiology and the Future of Genomics

Epidemiology will help measure the impact that gene variants have on the risk of disease, and it is hoped that it will be able to draw associations between modifiable risk factors and their interaction with gene variants. Epidemiology will also have an effect on the future of genomic studies. In the field of gene discovery, epidemiological practices will play a role in study design with regard to participant selection and the ability to generalize results to different populations. Epidemiology will also be needed to determine the efficacy and safety of newly developed genetic tests in different populations. Genetic testing will need to be evaluated based on sensitivity and specificity.

Biomarkers, which are substances that are used to indicate the state of the biological environment within the body, will become useful tools in determining measures of exposure. Biomarkers can be specific chemicals present in the body or they can be by-products of the breakdown of chemicals in the body. In the scope of genetics, biomarkers can present themselves as functional or dysfunctional proteins. These proteins could be the result of mutations or SNPs in an individual's genome. Research involving transcriptomes, proteomes, and metabolomes may show correlations with future risk of disease. This research may help bridge the gap between early exposure and the development of disease later in life.

Another promise of genomics is that in the future an individual will be able to go to their doctor and have genetic testing done to determine the risk of disease given their genetic variants. This concept is called genomic profiling. Epidemiological data along with clinical trials will be needed to determine and develop risk estimates and to assess whether genomic profiling will have validity and value.

Genomic profiling has also raised questions on cost-effectiveness. Using genetic information to develop interventions may be no more effective at disease prevention than epidemiologic studies that have developed generic population-wide interventions such as smoking cessation, increased exercise, and institution of a proper diet. Interventions based on genetic information may also be more costly. Other cost-effective issues are raised in the field of pharmacogenomics. Epidemiological parameters will be necessary to determine the value of completing genetic testing prior to the selection of a drug therapy.

Genomics will also affect study design, the analysis of data, and how those data will be interpreted. The effects on study design will include the appearance of large-scale cohort and case-control studies and the development of new study designs such as the case-only study design. The case-only study design has been used to study gene-environment interactions as well as how multigene interactions relate to the cause of disease.

New and emerging methods for complex data analysis are also a result of genomics. Sample size will greatly increase as a result of genomic data. However, within this vast amount of data, only a fraction of the information will likely have any relevance to the aims of the study. Sorting this large amount of information will require the development of new statistical algorithms. These new algorithms must be tested before they can be used to deem them practical in epidemiologic studies. The test will be in proving reproducibility of outcomes. Emerging analytical methodologies include hierarchical regression modeling and Bayesian methods. Interpretation of data will be strengthened by genomic research in the sense that the statistical power of association between environment and human disease will be enhanced. Associations between genetic variants and environmental factors will be extremely helpful in clarifying the complex issue of disease etiology.

Genomics is influencing and will continue to influence epidemiology and epidemiology will have the same effect on genomics. These two disciplines exist in quite separate worlds. Training for scientists in both disciplines will need to be undertaken to keep abreast of the rapidly changing worlds of both genomics and epidemiology. Collaboration will also be necessary to ensure a minimization of bias and to eliminate confounding factors in research being undertaken. Deciphering the genetic code to determine its role in disease etiology will be crucial to understanding health and developing disease preventions.

Applying genomic concepts to epidemiology will have to be done in the context of social and ethical beliefs with an end goal of improving human health. The race to publish the entire human genome is over, and now begins the painstaking review, analysis, and application of the information contained within the code. The generalized population-wide interventions developed by epidemiologists thus far have provided a strong basis for the next stage of epidemiology, examining the complex interplay of genes and environment. It is hoped that this examination will lead to the development of more specific plans for prevention, detection, and treatment of the diseases that plague the world today.

—Jennifer Currie

See also Gene; Gene-Environment Interaction; Genetic Disorders; Genetic Epidemiology

Further Readings

Brown, T. A. (2002). *Genomes.* Hoboken, NJ: Wiley.
DeSalle, R., & Yudell, M. (2005). *Welcome to the genome: A user's guide to the genetic past, present, and future.* Hoboken, NJ: Wiley.

Khoury, M. J., Little, J., & Burke, W. (2004). *Human genome epidemiology: A scientific foundation for using genetic information to improve health and prevent disease.* New York: Oxford University Press.
Smith, G. (2005). *The genomics age: How DNA technology is transforming the way we live and who we are.* New York: AMACOM.

Genotype

The term *genotype* was introduced in the developing field of genetics following the rediscovery and further application of Gregor Mendel's work during the early 1900s. The genotype is the genetic constitution or makeup of an individual. It can be used to refer to the overall assemblage of genes that an individual possesses at all genetic loci, the positions occupied by genes on a chromosome, or more typically, in the specific sense to refer to the genetic constitution at a precise genetic locus in an individual. In the second usage, we refer to the alleles, the alternate forms of a gene, which are present at a specific locus on each member of the chromosome pair on which the gene is located. The introduction of new genetic variation in the base pair structure of the alleles is the result of the process of mutation. Mutations, or mutant alleles, are alterations in the genetic material of plants and animals that can have positive, deleterious, or neutral change effects. Epidemiological studies and medical genetics typically focus on mutations having deleterious or disease-causing effects.

The genetics of the ABO blood group is familiar to most people and provides a useful tool for explaining genotype in relation to phenotype. There are four phenotypic expressions in this trait. Some people have type O blood because their blood does not react with the antibodies for either A or B antigens, the proteins on the red blood cells that cause antibodies to be made. Other individuals have type A or B blood because their blood reacts with antibodies for the antigens in these respective blood types. Finally, some people react to both type A and B antibodies and are deemed type AB. At the genetic locus for ABO on the chromosome number 9 (Chr 9) pair, there are a number of possible pair combinations of alleles A, B, and O that yield any one of six genotypes that account for the four phenotypic expressions. Type AB individuals have the genotype AB, which means on one member of the Chr 9 pair, they have an A allele, and on the other, a B allele. Type O individuals are genotypically OO (an O allele on each Chr 9). People with type A blood may be genotypically AA or AO, as type A is a dominant trait with respect to O. Type B is also dominant, yielding two possible genotypes, BB or BO. A and B are said to be codominant, as both traits are expressed in individuals with this genotype.

Some genetic disorders occur when an individual inherits a single disease-causing allele. These disorders are said to exhibit dominance. For most of these disorders, the disease gene is located on one of the autosomal, or non-sex-determining, chromosomes, and the disorders are labeled autosomal dominant (AD). This means that the mutant allele needs to be located on only one member of the chromosome pair for the disorder to occur (assuming no other complications that sometimes occur in the expression of the disease). Achondroplasia, a genetic disorder causing short stature predominantly due to problems converting cartilage tissue to bone, is an example of a disorder that is inherited in an AD fashion. This disorder is caused by a mutation in the fibroblast growth factor receptor 3 gene.

A second large group of genetic disorders involving the genes located on the autosomal chromosomes is labeled autosomal recessive (AR), due to the fact that for these disorders to manifest, there must be a mutant allele on both members of the chromosome pair. An example of an AR disorder is sickle cell disease, a disease involving abnormalities in hemoglobin, a complex protein in the red blood cells used to carry oxygen. Patients with sickle cell disease have mutations in both the alleles of the Beta Hemoglobin gene.

Finally, there are some disorders that result from mutations in genes located on the X chromosome, one of the two sex-determining chromosomes (the other is the Y chromosome). These disorders may manifest in a recessive or dominant manner and are referred to as X-linked recessive or X-linked dominant disorders, respectively. In X-linked recessive disorders, if a male has the mutation on his only X chromosome (males have a genotype of one X chromosome and one Y chromosome while females have two X chromosomes), he will manifest the disorder. However, for a female to have this X-linked recessive disorder, she must have a mutant allele on both X chromosome. The occurrence of mutations in the same gene on both X chromosomes in female individuals occurs less frequently in populations, such that the majority of individuals affected with an X-linked recessive disorder are males.

With the major advances in techniques of human molecular genetics in the past 20 years, we have witnessed a renewed focus on the genetic factors influencing disease and the potential for treatment, diagnosis, and prediction of diseases through genotype information. Current research focuses on the relationship of the genotypes in individuals affected with a genetic disorder to the phenotypic (chemical, behavioral, and physical expression of the underlying genotype and environmental factors) outcomes, such as clinical findings found by physical examination in patients with the disorder. New methods for detecting mutations in genes have identified mutant genotypes not previously observed in genetic disorders. The Human Genome Project has led to the development of hundreds of new genetic tests designed to identify the genotypic information of individuals thought to have specific genetic diseases. The success of this program and similar initiatives, along with continuing scientific advances in molecular medicine, have moved us even closer to the day when physicians will be equipped with detailed genetic data to assist them in the prediction and prevention, and the diagnosis and treatment, of a broad array of genetic diseases in human populations.

—F. John Meaney, Jennifer Andrews, and Timothy Miller

See also Association, Genetic; Chromosome; Gene; Gene-Environment Interaction; Mutation; Phenotype

Further Readings

Butler, M. G., Begleiter, M., Lillis, S., Lund, M., & Meaney, F. J. (2005). Patterns of inheritance: Mendelian and non-Mendelian. In M. G. Butler & F. J. Meaney (Eds.), *Genetics of developmental disabilities* (pp. 35–79). London: Taylor & Francis.

Nussbaum, R. J., McInnes, R. R., & Willard, H. F. (2001). *Thompson and Thompson genetics in medicine* (6th ed.). New York: W. B. Saunders.

Scriver, C. R., & Waters, P. J. (1999). Monogenic traits are not so simple: Lessons from phenylketonuria. *Trends in Genetics, 15*, 267–272.

Geographical and Social Influences on Health

There is abundant and growing evidence that places matter to health, over and above the characteristics of individuals. The relevance of places to health has been considered across a wide range, from the national scale down to the neighborhood level. This entry highlights some of the key geographical and social factors (including socioeconomic status, social capital, and income inequality) and empirical evidence that have been explored at the country, state/regional, and neighborhood levels, and that have advanced existing knowledge in the field.

Country-Level Contexts

Geographic contexts at the country level have long been recognized as related to variations in health. Life expectancy and mortality rates from particular diseases are known to vary widely across countries. A half century ago, evidence was gathered that suggested that part of these differences may stem from variations in lifestyle behaviors, which may in turn be socioculturally determined. In the 1950s, Ancel Keys initiated the Seven Countries Study, to examine the relations between diet and cardiovascular diseases within and across cohorts of men in Finland, Greece, Italy, Japan, the Netherlands, the United States, and Yugoslavia (Keys et al., 1967). Using dietary records, considerable variations were found in dietary intakes between countries. Marked variations in mean serum cholesterol levels across countries were furthermore identified. These variations were thought to, in part, account for corresponding differences in incidence and mortality rates of coronary heart disease.

Epidemiologist Geoffrey Rose (1985) made the key distinction between the *causes of cases of disease* within a population and the *causes of incidence of disease*. Rose contrasted the population distribution of serum cholesterol levels in Japan (a country in which heart disease was uncommon) to that in eastern Finland (where the incidence rate of heart disease was relatively high) and observed that the entire population distribution of cholesterol in Finland was to the right of the distribution in Japan. In other words, most of the population in Finland was "hypercholesterolemic" relative to the distribution in the Japanese population. It was this contrast (reflecting the causes of incidence of disease) that could potentially explain the much higher rates of cardiovascular disease in Finland compared with Japan.

Migrant studies have also provided evidence to support the important influences of geographical and social contexts on health, with these differences

arising from members of the same racial or ethnic group. For example, Japanese populations in Japan, Hawaii, and San Francisco have been observed to have very different levels of saturated fat intakes, as well as serum cholesterol levels, body weight, and age-adjusted coronary heart disease rates, which are all respectively higher with closer proximity to the mainland United States. These patterns are consistent with effects of context on dietary intakes, since genetic factors would be relatively homogeneous across these populations of Japanese descent.

Moreover, the vast differences in life expectancy and mortality across countries globally have been attributed to the socioeconomic attributes of these societies, including levels of economic development and economic inequality. Economic development may be related to health through differences in the availability of food and other local material resources and related effects of urbanization. For instance, there is some evidence to support that populations in countries at higher levels of economic development (as measured by the gross domestic product, GDP, per capita) have higher levels of body mass index and serum cholesterol. Among rich countries, variations in average life expectancy are not explained by their GDP levels, and this observation gave rise to income inequality as a possible explanatory factor. Evidence in the 1990s among selected rich countries showed a strong positive correlation between a higher share of total income going toward the least well-off proportion of the population and higher life expectancy. Mechanisms put forth for this relationship have included negative health effects resulting from individuals' feelings of relative deprivation, declines in social cohesion and trust, and underinvestments in public goods such as education and health care, as interests of the rich move away from those of the poor. Nevertheless, since the initial evidence, a number of ecologic studies of the effects of income inequality at the country level have been conducted that have been relatively mixed in their findings.

State and Regional Contexts

State and regional contexts have been proposed to have independent effects on one's health, primarily through policy-related mechanisms. Examples of two such characteristics that have become prominent in the epidemiologic literature by way of demonstrated state-level associations with health in the United States are *social capital* and *income inequality.*

Social capital has been defined as the resources within social networks that can be mobilized for purposeful actions. Alternatively, it has been defined as both the social structures and the associated cognitive resources such as trust and reciprocity. While social capital has been conceptualized at the neighborhood and individual levels, its significant associations with health were first demonstrated at the state level in the United States. U.S. researchers analyzed state-level data from the National Opinion Research Center's General Social Surveys on interpersonal trust, norms of reciprocity, and membership in voluntary associations and determined that these social capital variables accounted for significant proportions of the cross-sectional variations in mortality rates across the U.S. states. Lower levels of trust were found to be associated with higher rates of most major causes of death, including heart disease, cancer, infant mortality, and homicide. Similar associations were observed between death rates and norms of reciprocity (the proportion of residents agreeing that "most of the time, people try to be helpful") and per capita membership in civic associations. These associations further remained after accounting for state differences in median income and poverty rates. Since these initial ecologic studies, exploration of the associations between social capital and health outcomes has rapidly expanded to encompass multilevel study designs, with adjustment for individual-level factors such as one's socioeconomic status (SES); investigate other specific forms of social capital (e.g., bonding and bridging social capital, informal and formal social capital) and health outcomes (e.g., cardiovascular and infectious diseases, and obesity and physical inactivity); and measure social capital at more local levels, that is, the neighborhood and community levels.

The associations for state- and regional-level social capital with individual health are thought to be largely mediated by policy-related mechanisms. For example, the collective efforts of state populations built on mutual cooperation and trust may plausibly bring about the implementation of statewide policies such as greater access to high-quality education and health care, and statewide funding of local resources for physical activity, which in turn could affect the health of state residents.

Building on earlier cross-national studies, studies of income inequality at lower spatial units of analysis have been performed within countries, particularly at the state or regional level. Early investigations were

ecologic in design, whereas more recent investigations have applied a multilevel analytic framework, which have taken into account the spatial correlations between individuals in the same area while controlling for individual-level SES to reduce confounding bias. Notably, most of the studies among developed nations that have found an association between income inequality and health have involved populations in relatively inegalitarian countries, conducted primarily at the state level in the United States, whereas findings at more local levels have been generally less significant. Findings at the regional and local levels in more egalitarian countries such as Japan and Sweden have generally been null and possibly may be due to the lack of sufficient variations in income inequality in these settings.

Neighborhood Contexts

There is growing evidence to support the notion that the levels of socioeconomic resources and amenities across neighborhoods, as well as the levels of social resources such as social capital, may affect the health of individuals, after taking into account the socioeconomic characteristics of individuals. Relevant material resources and amenities include the availability of nutritious foods and green spaces; the quality of housing and of health services; the presence or lack of "incivilities," such as graffiti and litter; and environmental hazards, such as air pollution and noise.

A variety of studies have identified moderate yet statistically significant associations between neighborhood SES (typically measured by aggregating individual-level measures of SES from surveys or censuses) and one's risk of dying from cardiovascular disease and from any cause, with 1.1 to 1.8 times higher observed risks of these outcomes after controlling for one's SES. Other studies have reported significant inverse associations between neighborhood SES with chronic disease risk factors, including smoking, diet, physical activity, and hypertension, and with the incidence of heart disease.

Evidence is also emerging on the effects of specific neighborhood material resources on health. For example, in one multilevel study, adjusting for individual SES and other types of food service places, the presence of a supermarket within one's census tract was found to be significantly associated with a 1.5 times higher relative risk of meeting the dietary guidelines for fruit and vegetable intakes, as well as with higher risks of meeting the guidelines for total fat and saturated fat intakes among African Americans.

Neighborhood social resources, particularly social capital, have also been increasingly investigated as a predictor of health outcomes. Using multilevel study designs, these analyses have suggested protective effects of specific forms of social capital (e.g., bonding and bridging social capital) on health behaviors (e.g., physical inactivity) and self-rated health. Such health effects may potentially occur through the promotion of the diffusion of knowledge about health promotion, by influencing health-related behaviors through informal social control, by enabling greater access to local services and amenities, and/or by contributing to psychosocial processes that provide support and mutual respect.

An additional key factor at the neighborhood level relevant to health is residential segregation, particularly by race or ethnicity. *Residential segregation* by race or ethnicity refers to the segregation of racial or ethnic groups along subunits of a residential area. Such segregation, which in the United States has historical roots for African Americans (e.g., through racial discriminatory practices of federal housing policies and bank lending practices), may affect health by influencing the levels of socioeconomic resources (e.g., levels of health care resources, and educational and job opportunities) available to its residents. Some evidence suggests that low-income African American and other minority neighborhoods have been "targeted" with fast-food restaurants and, prior to tobacco legislation, with the advertising of cigarettes. Furthermore, residential areas that are predominantly African American have been associated with higher rates of infant and adult mortality, although such evidence has primarily been drawn from ecologic (rather than multilevel) studies.

—Daniel Kim

See also Health Disparities; Multilevel Modeling; Social Capital and Health; Social Epidemiology

Further Readings

Diez Roux, A. V., Merkin, S. S., Arnett, D., Chambless, L., Massing, M., Nieto, F. J., et al. (2001). Neighborhood of residence and incidence of coronary heart disease. *New England Journal of Medicine, 345*, 99–106.

Kawachi, I., & Berkman, L. F. (2003). *Neighborhoods and health*. New York: Oxford University Press.

Kawachi, I., Kennedy, B. P., Lochner, K., & Prothrow-Stith, D. (1997). Social capital, income inequality and mortality. *American Journal of Public Health, 87*, 1491–1498.

Keys, A., Aravanis, C., Blackburn, H. W., Van Buchem, F. S., Buzina, R., DjordjeviÌ, B. D., et al. (1967). Epidemiological studies related to coronary heart disease: Characteristics of men aged 40–59 in seven countries. *Acta Medica Scandinavica, 460*(Suppl.), 1–392.

Kim, D., Subramanian, S. V., Gortmaker, S. L., & Kawachi, I. (2006). State- and county-level social capital in relation to obesity and physical inactivity: A multilevel, multivariable analysis. *Social Science & Medicine, 63*, 1045–1059.

Rose, G. (1985). Sick individuals and sick populations. *International Journal of Epidemiology, 14*, 32–38.

Geographical and Spatial Analysis

All events have space and time coordinates attached to them—they happen somewhere at some time. In many areas of epidemiology, recording the place of individual events and exposure is vitally important. The recent surge in the availability of desktop computing power, geographical information systems (GIS) software, and interest in the effect of neighborhood conditions on development of disease have caused a resurgence of interest in spatial data analysis.

Types of Spatial Data

Spatial data consist of measurements or observations taken at specific locations or within specific spatial areas. In addition to values for various attributes of interest, spatial data sets also include the locations or relative positions of the data values. There are three main types of spatial data. The first type of data, geostatistical data, is measurements taken at fixed locations. In most cases, the locations are spatially continuous, that is, data locations are available in close proximity to each other. An example of geostatistical data would be measures of the concentration of pollutants at monitoring stations. The second type of spatial data is lattice data, which are area-referenced data with observations specific to a region or area. An example of lattice data is the rate of specific types of cancer deaths by state from the National Cancer Institute's cancer mortality atlas. The areas can be regularly or irregularly spaced. Areas are often referenced by their centroid. The third type of spatial data is point pattern data, which arise when locations themselves are of interest. Spatial point patterns consist of a finite number of locations observed in a spatial area. Examples of point pattern data include the locations of women diagnosed with breast cancer on Long Island.

Spatial Scale

The spatial scale or resolution is an important issue in the analysis of spatial data. Patterns observed in spatial data may be the result of different processes operating at different scales. This is known as the modifiable areal unit problem (MAUP). The MAUP consists of both a scale and an aggregation problem. The concept of the ecological fallacy is closely related to the MAUP. The scale problem refers to the variation that can occur when data from one scale of areal units are aggregated into more or fewer areal units. For example, much of the variation among counties changes or is lost when the data are aggregated to the state level. The choice of spatial areas is often arbitrary in nature, and different areal units can be just as meaningful in displaying the same base-level data. Clearly, it is more meaningful to use "naturally" defined areas (e.g., neighborhoods or hospital catchment areas) rather than arbitrary administrative units.

Frequentist Versus Bayesian Analysis

Traditionally, epidemiologists have used the frequentist approach for the analysis of data, including spatial data. However, fluctuations in disease rates may occur because of the sparseness of the data in certain areas. In addition, data are often spatially correlated, meaning there is a tendency of adjacent areas to have similar rates of disease incidence. In many cases, there are no valid or accepted frequentist methods for tackling these problems, or the frequentist methods are complex and difficult to interpret.

In contrast, Bayesian models can easily incorporate spatial correlations. The ability to consider such correlations serves to stabilize estimates of relative risk, making them less vulnerable to random fluctuations in the observed number of events in each area. The Bayesian approach begins by considering our knowledge regarding the parameters being estimated, prior to the collection or inspection of any current data. This

information, termed *prior information*, or simply the "prior," is usually represented by a probability density function that quantifies the investigator's beliefs about the likelihood of any particular value of the parameter being the true value, before we know the current data. This prior knowledge is then revised using Bayes's theorem, based on the new or current data, to obtain posterior information represented by a posterior probability density function. Bayesian analysis is increasingly being used in spatial epidemiology. While there is no debate about the mathematical truth of Bayes's theorem, questions regarding the interpretation of Bayesian results, however, largely center on methods used to define the prior information.

Analysis of Spatial Point Pattern Data

In spatial epidemiology, one pattern of particular interest is the presence or tendency of locations to cluster together, that is, for disease to occur more frequently in cases in close proximity to one another than would be expected from a set of cases with no common environmental exposure. Complete spatial randomness defines a situation where the disease of interest is equally likely to occur at any location within the study area, regardless of the locations of other disease events. To test the presence of a cluster of disease, we determine if disease occurrence follows a uniform distribution across the study area. There are numerous examples of identifying clusters of disease occurrence.

A second approach to the analysis of spatial point pattern data is the use of multilevel models when examining the association of area-level exposures on individual-level disease occurrence. There is increasing interest in the use of area-level characteristics as having an effect on various health outcomes over and above individual characteristics. In multilevel models, an individual, including his or her exposures, covariates, and disease occurrence, is considered to be nested within a spatial area. For example, individuals (Level 1) may be nested within census tracts (Level 2), which are, in turn, nested within counties (Level 3). In multilevel models, the random components are assessed at the individual level and at the area level(s). Multilevel models are able to calculate the variance between areas as a percentage of the total variance thereby providing an indication of the spatial variation. Recent advances in multilevel models also allow for calculation of the median odds ratio that represents the extent to which a person's probability of disease occurrence is a function of the area in which he or she resides. An additional measure, the interval odds ratio, provides for comparison of the importance of the area-level characteristic relative to the variation remaining among areas. Although the multilevel approach is typical of most studies examining area-level effects, it ignores frequently the spatial adjacency between areas and does not effectively incorporate any notion of space.

A third method, the geostatistical approach, considers individuals and areas to be distributed across continuous space that allows for estimation of the spatial scale of the variation in disease occurrence unlike multilevel models and examination of the effect of weighted area-level characteristics on the same outcome. Individuals are positioned at their geocoded location using latitude and longitude and nested within a prespecified spatial area. Area-level characteristics can then be used in circular space around the location of the individuals. This area may exceed the boundaries of the area of residence. Weights are then obtained by means of a decay function to reflect the magnitude to which area-level characteristics at distant locations from individuals have a lower impact than those that were closer to their location.

Analysis of Lattice Data

Lattice data are observations from a random process observed over a countable collection of spatial areas supplemented by a structure describing the location of each area relative to its surrounding areas (i.e., adjacency matrix). Data observed can be continuous (e.g., mortality rate) or discrete (e.g., participation in physical activity).

Examining clusters of disease occurrence using lattice data has been done frequently. The goal for methods based on scanning local disease rates is often to identify areas with unusually high or low local disease rates. There are several approaches to the analysis. First, many studies have used a spatial scan statistic. The statistic uses a window of variable angles and elliptical shapes that moves across the map. The null hypothesis is that event rates are the same in all windows. Clusters are defined as areas having either a lower or higher rate of disease than expected. The process of cluster detection uses Monte Carlo permutations of the data set.

A second method uses a GIS to define a set of grid points covering the study area and calculate

local incidence proportions within circles centered at the grid points. An area falls within the circle if the area's centroid falls within the circle. To assess statistical significance of the local incidence rates, Rushton and Lolonis (1996) use Monte Carlo tests at each location based on an overall constant risk hypothesis where cases are assigned to areas according to the incidence rate observed for the entire study area.

A third method is the use of area-level data for the exposure, covariates, and disease rates as part of an ecological study design. Bayesian methods are frequently used to analyze these data. Area-level relative risks are estimated by integrating local information consisting of the observed and expected number of cases in each area, prior information on the overall variability and/or spatial structure of the relative risks, and the potential effect of spatially defined covariates. The expected number of events is frequently based on the age and sex distribution of the area and a reference set of disease probabilities. The hierarchical statistical model consists of two levels. The first level represents the local variability (within area) and is modeled using a Poisson distribution. The second level consists of a random effect structure that accounts for the extra-Poisson variability due to the aggregated effect of unknown confounders via a spatial ''clustering'' component and unstructured ''heterogeneity'' component. In such studies, a statistical challenge is to account both for potential errors in the numerator and/or denominator of the rates and for unequal population sizes inducing differential variability of the disease rates between areas. Care needs to be taken not to extrapolate to the individual level based on these aggregate data (ecological fallacy).

—Mario Schootman

See also Bayesian Approach to Statistics; Ecological Fallacy; Geographical and Social Influences on Health; Multilevel Modeling

Further Readings

Chaix, B., Leyland, A. H., Sabel, C. E., Chauvin, P., Rastam, L., Kristersson, H., et al. (2006). Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmo, Sweden, in 2001. *Journal of Epidemiology and Community Health, 60*, 427–435.

Kawachi, I., & Berkman, L. (2003). Introduction. In *Neighborhoods and health* (pp. 1–19). New York: Oxford University Press.

Kulldorff, M., Huang, L., Pickle, L. W., & Duczmal, L. (2006). An elliptic spatial scan statistic. *Statistics in Medicine*, *25*(22), 3929–3943.

Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., et al. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health, 60*, 290–297.

National Cancer Institute. (n.d.). *Long Island Geographic Information System (LI GIS)*. Retrieved December 6, 2006, from http://www.healthgis-li.com.

Rushton, G., & Lolonis, P. (1996). Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine, 15*, 717–726.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling.* London: Sage.

Gestational Age

Gestational age is the time period in which the fetus grows inside the uterus. Measured in weeks, gestational age has implications for the fetus's growth, as well as cognitive and physical development. The gestational age of a fetus is particularly important when determining the potential negative effects of a fetal exposure to toxins or infection and has a direct impact when planning appropriate medical treatment for such situations.

Gestational age is divided into two periods: embryonic and fetal. Preceded by the embryonic period, the fetal period begins at the gestational age of Week 10 and continues until birth. Prenatal development benchmarks are linked to gestational ages. For example, at the gestational age of 7 to 8 weeks, all the vital organs have begun to form along with the formation of bones and cartilage. By the gestational age of Weeks 9 to 13, the genitalia have formed and the entire fetus weighs about 1 oz. By Weeks 21 to 23, the fetus's eyes have developed and the fetal heartbeat can be heard by stethoscope.

A normal pregnancy has a gestational age ranging from 38 to 42 weeks with a full-term pregnancy considered to be 40 weeks. Infants born at a gestational age of less than 38 weeks are considered premature and are susceptible to increased risks of morbidity

and mortality. For example, the vast majority of infants born at 24 weeks will experience respiratory distress syndrome as the air sacs in the lungs have just begun forming.

Gestational age can be calculated before and after birth. Before birth, gestational age is calculated as the time from the first day of a pregnant woman's last menstrual cycle to the current date. Although often used to determine gestational age, health professionals recognize the potential for inaccuracies using this method due to variations in ovulation dates. Therefore, there are a number of other methods employed to accurately determine gestational age. One such method is the use of an ultrasound whereby growth can be determined through measurements of the head and abdomen. Another method is to determine the date of conception as per the mother's knowledge.

After birth, a newborn's gestational age can be measured using the Ballard Scale or the Dubowitz Exam. The Ballard Scale involves an examination of the neuromuscular and physical maturity of the newborn, while the Dubowitz Exam focuses on the neurological maturity of the newborn.

It is important to note that although gestational age may be accurately determined by the methods mentioned above, developmental growth at each week may vary from fetus to fetus. An estimated 3% to 10% of all newborns are determined to be small for gestational age (SGA) because their birthweight or length was determined to be less than the 3rd percentile. SGA newborns have higher incidence of learning disabilities, autism, and attention deficit disorder (ADD).

—April L. Winningham

See also Child and Adolescent Health; Maternal and Child Health Epidemiology; Preterm Birth

Further Readings

The Magic Foundation. (2007). *Small for gestational age*. Oak Park, IL: Author. Retrieved March 2, 2007, from http://www.magicfoundation.org/www/docs/113/small_for_gestational_age.html.

U.S. National Library of Medicine and National Institutes of Health. (2007). *Gestational Age. Medline Plus: A service of the U.S. National Library of Medicine and National Institutes of Health*. Bethesda, MD: Author. Retrieved March 2, 2007, from http://www.nlm.nih.gov/medlineplus/ency/article/002367.htm.

Global Burden of Disease Project

The Global Burden of Disease Project (GBDP) draws on a wide range of data sources to produce consistent estimates of the cost of morbidity and mortality worldwide, as defined by the years of healthy life lost due to injury and illness. The GBDP is conducted by the World Health Organization (WHO) and updates the Global Burden of Disease Study for the year 1990, which was commissioned by the World Bank and carried out by the WHO and Harvard University in 1992. Both the original study and the GBDP quantify the burden of disease using the concept of the disability-adjusted life year (DALY), which allows an estimation of the total years of healthy life in a population by combining two types of information. Years of life lost (YLL) represent the cost of premature mortality, calculated as the projected years a person was expected to live; years lost due to disability (YLD) represent the years of healthy life lost due to diminished quality of life caused by illness or disability. One DALY is equivalent to the loss of 1 year of life in full health, so that the DALY cost would be equivalent for a person in perfect health dying 1 year prematurely, or a person living 2 years in a health state judged to be half as desirable as the state of full health.

The primary advantage of using a common metric such as the DALY is that it allows comparison of the costs of disparate medical conditions and the benefits that might be gained by different interventions. It is particularly important that the DALY approach assigns a value to the cost of ill health as well as to death, allowing that an estimation of the societal cost of conditions such as hearing loss are seldom fatal but may significantly reduce an individual's quality of life. This approach is commonly used today but was an innovative concept when included in the 1990 report. The DALY approach allows governments and other policymakers to evaluate the costs of disparate medical conditions, make informed judgments about where to target resources in order to produce the greatest improvement in health for a given investment of resources, and evaluate the comparative effectiveness of interventions targeted at different conditions. Many countries have adopted the global burden of diseases (GBD) approach to guide national priorities in health research as well.

For 2001, the GBD found that, as in 1990, the most important causes of disability in all regions of

the world were neuropsychiatric illnesses, which accounted for more than 37% of YLD for adults above 15 years worldwide. The greatest amount of loss of healthy life due to premature mortality was caused by noncommunicable conditions, including cancers, diabetes, and heart disease: These conditions accounted for almost 50% of DALYs lost in low- and middle-income countries and more than 85% of the loss in high-income countries.

—*Sarah Boslaugh*

See also Chronic Disease Epidemiology; Disability Epidemiology; Quality of Life, Quantification of; World Health Organization

Further Readings

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. R., & Murray, C. J. L. (2006). *Global burden of disease and risk factors*. New York: Oxford University Press. Retrieved November 24, 2006, from http://www.dcp2.org/pubs/GBD.

Murray, C. J., Lopez, A. D., & Jamison, D. T. (1994). The global burden of disease in 1990: Summary results, sensitivity analysis, and future directions. *Bulletin of the World Health Organization, 74*(3), 495–509.

World Health Organization. (2007). *About the Global Burden of Disease Project*. Retrieved November 24, 2006, from http://www.who.int/healthinfo/bodabout/en/index.html.

Goldberger, Joseph
(1874–1929)

Joseph Goldberger was the medical detective and pathbreaking epidemiologist in the U.S. Marine Hospital Service (later the U.S. Public Health Service [PHS]) responsible for discovering that pellagra was a nutritional disease. Goldberger's research and controversial experimentation revealed that pellagra could be cured and prevented by eating a balanced diet rich in animal protein or augmenting the diet with brewer's yeast. Later research following Goldberger's death revealed that insufficient nicotinic acid or niacin in the diet caused pellagra.

Born in Giralt, Hungary, on July 16, 1874, Goldberger emigrated to the United States with his parents and three older siblings in 1883 after a plague wiped out the herd of sheep on which the family depended for its living. Among the millions of Jewish immigrants to arrive in the United States in this era, Joseph's parents, Samuel and Sarah Goldberger, opened a small grocery store on New York's Lower East Side. Joseph studied at the City College of New York but soon transferred to Bellevue Hospital Medical College where he received his medical degree in 1895. After several unsatisfying years in private practice in Wilkes-Barre, Pennsylvania, Goldberger passed an entry examination and was commissioned an Assistant Surgeon in the U.S. Marine Hospital Service. He began his career as a federal physician examining immigrants at the Barge Office in lower Manhattan while a scorched Ellis Island was being reconstructed.

Between 1902 and 1906, Goldberger battled epidemics, gaining valuable experience in both field and laboratory. He fought yellow fever in Mexico, Puerto Rico, Mississippi, and Louisiana. While on duty in Louisiana, a colleague, Farrar Richardson, introduced him to Mary Farrar, the daughter of a wealthy Episcopalian New Orleans family and the grandniece of Confederate president Jefferson Davis. Although both families initially opposed the marriage on grounds of religious difference, Joseph and Mary married in 1906, deciding that while they eschewed traditional religion, rationalism, humanitarianism, research, and science would guide their lives together. They had four children.

Shortly after his marriage, Goldberger was assigned to the Hygienic Laboratory in Washington (precursor of the National Institutes of Health) to study typhoid fever. Soon, he was off to Texas after an outbreak of dengue fever and back to Mexico to battle typhus fever. For the third time, Goldberger was felled and almost lost his life to the pathogens he studied. Previously he had suffered from yellow fever and dengue while battling epidemics. From his work in the field, Goldberger observed that the diseases individuals contracted were often related to the conditions in which they lived. In his almost daily letters to his wife, Goldberger frequently shared his observations that poverty, ignorance, and poor sanitation made some populations significantly more vulnerable to disease than others.

Dr. Joseph Goldberger was earning a reputation as a hardworking, bold epidemiologist. His study of Schamberg's disease and discovery that its characteristic red rash was actually the result of tiny mite bites earned him the reputation for being clever. Collaborating with PHS physician John Anderson, Goldberger

demonstrated that Brill's disease was identical to typhus. In their work on measles, they were able to infect monkeys and learned that the disease was caused by a virus small enough to pass through a Berkefeld filter and that the virus could be identified in a victim's buccal and nasal secretions, but that in patients, infectivity decreased as convalescence proceeded.

In 1912, Goldberger's fine work earned him promotion to the rank of Surgeon. Two years later, he was in Detroit investigating an outbreak of diphtheria when he was summoned back to Washington by Surgeon General Rupert Blue, who requested that he direct the government's investigation of pellagra, which was begun in 1912 but was faltering.

First identified among Spanish peasants by court physician Don Gaspar Casal, pellagra was known as *mal de la rosa.* It killed and caused chronic illness among populations in various parts of Europe, the Middle East, Africa, and Asia. The disease was rampant in parts of southern Italy, where it was called *mal del sole* because it seemed to peak as spring arrived. Italian physician Francesco Frapolli dubbed it "pellagra" referring to the "rough or dry skin" that became the basis for diagnosis and for distinguishing the disease from other ailments. By the early 20th century, pellagra was taking thousands of lives annually, especially in the South Atlantic states. Some called it the "scourge of the South," others dubbed it the disease of the four D's: dermatitis, diarrhea, dementia, and death. By 1912, South Carolina reported 30,000 with a mortality rate of 40%. In the years prior to 1940, 100,000 Americans died of the disease.

Goldberger's notes and underlines in state public health reports from Illinois and southern states, as well as in the published essays of Italian physician Cesare Lombroso, suggest that even before Goldberger headed south, this expert on infectious diseases speculated that dietary deficiency and not a germ was pellagra's cause. His pencilings reveal that Goldberger thought that better diets, especially those with ample milk and fresh meat, might make a difference.

Observing pellagrins (persons with pellagra) in the South's orphanages, mental hospitals, and mill towns, Goldberger became increasingly convinced of his dietary hypothesis. He observed that pellagrins in the South often ate a corn-based diet deficient in animal protein. Staff, on the other hand, often had access to eggs, meat, and milk and rarely contracted the disease. The distinction between inmates and staff was not usually made by pathogens.

Goldberger tested his dietary theory in various venues beginning in 1914. He requested shipments of federal food and fed children in two orphanages, one Baptist and one Methodist, a balanced diet. All those stricken with pellagra recovered, and there were no new cases. At the Georgia State Asylum at Milledgville, Goldberger and his assistant, Dr. George Wheeler, performed dietary research using female inmates. They isolated 36 white female pellagrins in one ward and 36 female African American pellagrins in another ward. Both groups were fed wholesome diets of fresh meat, milk, eggs, and vegetables. A control group of 32, 15 nonpellagrous women, including 17 black and 15 white females, continued the normal diet. Fifteen of the control group developed pellagra symptoms. None of the women on the new diet got the disease. Although demonstrations were suggestive, they were not the kind of controlled double-blind experiments that might have persuaded even the most ardent skeptics that pellagra was a dietary disorder. Nor did Goldberger's efforts dissuade some critics from their belief that there was a pellagra germ.

Goldberger designed a bold assault on the germ theory of pellagra to demonstrate that there was a substance the absence of which from the diet induced the disease. In April 1915, Goldberger, with the assent and cooperation of Mississippi's progressive governor Earl Brewer, fed the corn-based diet ubiquitous among poor southerners to 11 volunteers at Mississippi's Rankin State Prison Farm. Brewer granted pardons to inmates who participated. Six of the eleven showed pellagra lesions when the experiment was ended in November. Goldberger had succeeded in producing pellagra symptoms through dietary change alone.

To persuade critics who continued to insist that pellagra was an infectious disease, Goldberger sought to transmit the disease to 14 healthy volunteers plus himself and his wife in April 1916. Calling his experiments his "filth parties," Goldberger injected into himself and his volunteers the blood of pellagrins. Some also swabbed the secretions from pellagrins' noses and applied it to their own and swallowed capsules containing pellagrins' excrement. None of the volunteers got the disease.

In his effort to determine which nutrients were missing from pellagrins' diets, Goldberger and PHS statistician Edgar Sydenstricker conducted epidemiological

investigations pathbreaking in their thoroughness in seven South Carolina mill villages from 1916 to 1919. Data on diet, housing sanitary conditions, and income among other variables offered an unparalleled view of living conditions that contributed to dietary deficiencies possibly causing pellagra. Some still doubted Goldberger's hypothesis, and he could not identify the missing nutrient.

A downturn in cotton prices in 1921 resulted in southern economic hardship and Goldberger predicted a rise in the number of pellagra cases. He was critical of sharecropping and an agricultural system that impoverished families and discouraged growing diversified food crops. He was equally vocal in criticizing mill owners for paying low wages that left workers unable to afford a balanced diet. Southerners objected to the notion that there would be hunger and more pellagra in the South. However, President Warren Harding stood by the PHS.

By 1925, Goldberger learned that small amounts of dried brewer's yeast could prevent and cure pellagra and at a much lower cost than a regular diet of meat, milk, and eggs. Learning that black tongue was the canine equivalent of pellagra, Goldberger and his associates now worked with dogs in the laboratory searching for the nutritional factor that when deficient in a diet could cause pellagra. During the Mississippi flood of 1927, Goldberger repeated his criticism of the southern economic system as he and Sydenstricker urged the Red Cross to send brewer's yeast to the affected areas and then toured the flooded counties. Again, Goldberger attributed the pellagra he saw to a grinding poverty rooted in the tenant system of agricultural labor and landowners' preference for cotton crops because of their profitability. Seasonal fluctuations in tenant income made a steady balanced diet uncertain at best.

Dr. Joseph Goldberger fell seriously ill in 1928. His last public address was at a meeting of the American Dietetic Association, where he reminded listeners that pellagra was primarily a matter of poverty and that medical science alone could never remedy the social conditions at the root of the disease. On January 17, 1929, he succumbed to hypernephroma, a rare form of cancer. Less than a decade after his death, Conrad A. Elvehjem and his associates discovered that a nicotinic deficiency caused black tongue in dogs. Dr. Tom Spies used nicotinic acid to treat cases of pellagra at the University of Cincinnati College of Medicine and at the University of Alabama, Birmingham Medical School. Nicotinic acid is part of the vitamin B complex. Later, biochemists at the University of Wisconsin learned that corn consumption depressed the level of nicotinic acid retained in the body. The corn-based diet did indeed, then, contribute to the prevalence of pellagra in the South. During World War II, the War Food Administration required all commercially produced white bread to be fortified with niacin. And after the war, state legislatures began to mandate flour enrichment.

—Alan M. Kraut

See also Governmental Role in Public Health; Health Disparities; Nutritional Epidemiology; U.S. Public Health Service

Further Readings

Bollet, A. J. (1992). Politics and pellagra: The epidemic of pellagra in the U.S. in the early twentieth century. *Yale Journal of Biology and Medicine, 65*, 211–221.

Carpenter, K. J. (Ed.). (1981). *Pellagra.* Stroudsburg, PA: Hutchinson Ross.

Etheridge, E. (1972). *The butterfly caste: A social history of pellagra in the South.* Westport, CT: Greenwood Press.

Kraut, A. M. (2003). *Goldberger's war: The life and work of a public health crusader.* New York: Hill & Wang.

Marks, H. M. (2003). Epidemiologists explain pellagra: Gender, race and political economy in the work of Edgar Sydenstricker. *Journal of the History of Medicine and Allied Sciences, 58*, 34–55.

Roe, D. A. (1973). *A plague of corn: The social history of pellagra.* Ithaca, NY: Cornell University Press.

Terris, M. (Ed.). (1964). *Goldberger on pellagra.* Baton Rouge: Louisiana State University Press.

Governmental Role in Public Health

Public health is the organized societal effort to protect and improve the health of communities. Public health focuses on the health of the entire population rather than on the health of individuals. The protection of population health is a key governmental responsibility at the federal, state, and local levels. Traditional activities include collecting information on the population, making rules to restrict individual activities that threaten the health of the community, and providing health services and educating people in healthy

behaviors that promote health. While health services and education can be provided by private sector organizations, only governmental authorities are authorized to collect official data and to create and enforce rules to restrict individual and corporate behaviors to protect and improve health. Governments also ensure access to health services for vulnerable and underserved populations.

History of U.S. Governmental Public Health

Organized governmental public health activities in the United States date from colonial times. Officials of the Massachusetts and Plymouth colonies mandated the recording of all births and deaths as early as 1639. The Massachusetts Bay Colony passed a law prohibiting the pollution of the Boston Harbor in 1647; other localities similarly created ordinances restricting activities perceived as dangerous to the health of the population, including laws to isolate people with infectious diseases and to quarantine ships in the harbors. However, there were often no effective governmental organizations to enforce the laws that were put in place.

Permanent boards of health were established in cities of the new United States of America as early as 1780, and scientific discoveries about the nature and transmission of disease provided a framework for their decisions to improve community health. However, for many years these boards were voluntary committees with no administrative staff to carry out their mandate to oversee the protection of water supplies and elimination of health hazards. Local health departments with staff began to appear at intervals in the larger cities (Baltimore in 1798, Charleston in 1815, Philadelphia in 1818, Providence in 1832, Cambridge in 1846, New York in 1866, Chicago in 1867, Louisville in 1870, Indianapolis in 1872, Boston in 1873). County health departments, with responsibility for the rural as well as the urban areas, did not begin to appear until 1911.

In 1850, a committee headed by Lemuel Shattuck published the *Report of the Sanitary Commission of Massachusetts*, calling for the establishment of state boards of health staffed by professional sanitary inspectors. The report also recommended laws and systems to collect data on the population and control the sources of disease transmission, programs to address social problems that affected community health, and improved training of medical personnel. Massachusetts then established the first permanent State Board of Health and State Health Department in 1869.

The earliest federal efforts in public health focused on preventing the transmission of disease via the nation's seaports through a system of Marine Hospitals and a National Quarantine Convention. A National Board of Health was created in 1879, but was only authorized for 4 years and was allowed to expire in 1883. The national Public Health Service was permanently established in 1902, and evolved over the years into the current U.S. Department of Health and Human Services.

Governance Foundation of Public Health in the United States

The organizational structure of governmental public health in the United States is driven by the unique structure of the U.S. government. The use of a federal system of governance provides for multiple layers and multiple players in the delivery of public health services. All three levels of the U.S. government (national, state, and local) are involved in governmental public health. The responsibilities, authority, and funding of those levels vary and sets into play a relational dynamic full of many challenges.

A federal governance system is based on power sharing between a central national government and subnational (i.e., state) governments. This system represents a compromise between centralized and decentralized authority with powers derived from a written constitution as opposed to a unitary system of governance such as the one found in Great Britain.

The U.S. Constitution limits federal powers to 14 enumerated powers among which is not found any public health authority. The last enumerated power, the authority of the federal government "to make all laws which shall be necessary and proper," has been used as a grant of implied power to do many things, including the conduct of public health programs.

States have inherent power as sovereign governments. Since the original states preceded the formation of the federal government and granted power to establish the national government, state governments have historically and legally been the central unit of the U.S. government. The 10th Amendment of the Constitution reserves to the states all powers that are neither given to the federal government nor prohibited by the Constitution. Thus, the police powers of government have primarily resided in state governments,

police power being the fundamental sovereign right of a government to promote order, safety, security, health, and general welfare within its territory.

In the troika of the U.S. government, local governments bear the most responsibilities for direct service to citizens but have the least authority. All local government organizations are creations of their state. They are legal entities only under state law with their power derived from state law and, in most cases, only political subdivisions of the state government.

The three levels of government involved in public health constitute a massive network of organizations involved in governmental public health. At a minimum, one must recognize one federal government with multiple agencies and 50 state governments with a collection of various departments involved in public health and related services. And there are no less than 3,000 local public health governmental entities participating in the system.

Federal Role in Public Health

The oldest element of the federal government public health activities is the U.S. Public Health Service, which traces its origins to 1798. Under the administration of President John Adams, a network of hospitals was established for the care of merchant seamen. This loose network was organized in 1870 into a centrally managed and professional medical care entity based in Washington, D.C., under the supervision of what became the Surgeon General of the United States. Today, the Public Health Service is the focal point of federal public health activities emanating from the U.S. Department of Health and Human Services (DHHS). Within DHHS are more than 300 programs covering a wide spectrum of health activities. Many of those program activities focus on preventing diseases, improving maternal and child health, assuring food and drug safety, analyzing health data, preparing for and responding to health-related emergencies, and conducting health and social science research.

One of the primary agencies within DHHS conducting public health activities is the Centers for Disease Control and Prevention (CDC). Founded in 1946 to help control malaria, the CDC remains at the forefront of governmental public health efforts to prevent and control infectious and chronic diseases, injuries, workplace hazards, disabilities, and environmental health threats. The CDC is globally recognized as the nation's premiere epidemiologic agency as well as a public health preparedness leader in the post-9/11 era.

The federal government, by far, represents the largest source of funding for health-related services. A primary role of the federal government is the direct provision of, as well as the payment for, both curative and preventive health services. Through programs such as the Indian Health Service, the Health Resources and Services Administration, and the Centers for Medicare and Medicaid Services, some $12.2 billion were appropriated for health services in 2006. In the area of health research, the National Institutes of Health within DHHS was provided $29 billion in 2006 to conduct medical and behavioral research in pursuit of basic knowledge about the nature and behavior of biological systems. Through various grant programs, DHSS is also a major funder of state and local health agencies.

State Role in Public Health

The primary constitutional responsibility for public health within the United States rests with the 50 state governments. Each state has some agency responsible for public health. No single standard exists for what constitutes a state health agency. In some states, the health agency is a freestanding cabinet department answerable to the governor. In other states, the health agency is part of a larger ("super") social services agency managing an array of health, social, and welfare programs. Mental health authorities and environmental protection services are sometimes assigned by the states to public health agencies. Medicaid administration as well as medical professions regulation may be located in the state health authority. These various organization combinations are based on the constitutional and statutory constructs of each state and reflect specific political cultures and historical developments of a state's region. The structure and placement of state health agencies within these variations will have a direct impact on the specific responsibilities of the agencies as well as the size of the agency's budget as a share of total state resources.

State health agencies play a primary role in determining the nature of services provided as well as the means of service delivery through local health departments. The policy-setting agenda for public health in any state is subject to state-level health agency influence. Most local health ordinances are based on enabling state legislation. State health agencies are

deeply involved in the development and drafting of legislation and even more involved in the promulgation of regulations to implement that legislation.

Functions that state health agencies across the United States have in common include disease surveillance and control, vital statistics, food safety, health facility regulation, public health laboratories, and environmental health. Most of the state health activities are funded by state general revenue and federal grants. One common program administrated by 98% of state health agencies is the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) funded by the U.S. Department of Agriculture.

Local Role in Public Health

The 3,000 local health agencies in the United States are the front line of governmental public health service delivery. As with organizations of state health agencies, no single organization standard exists for local health departments. In some states, the local health department is a direct agency of the state health department. In other states, local health departments can be part of the general-purpose local government within either a city or a county government. Local health departments can also exist as freestanding special-purpose governmental entities with their own tax base and elected board independent of any municipal or county government. Municipal local health departments have a long history, predating the existence of state health departments; however, the long-standing responsibility of local health departments to protect the community's health is often trumped by the legal supremacy of state government over local government.

Local health departments in 21st-century United States exist to carry out responsibilities embodied in state law and in local ordinances that are derived from state laws. Local health departments share responsibility with state health departments under the federal system of governance. Along with that sharing of responsibility comes the obligation of the state government to financially support the local heath department. Again, variations exist across the 50 states as to the extent of financial support provided to local health departments. Most local governmental public health depends on property taxation as its principal revenue source, supplemented by state subsidies and limited federal grants.

The structure of the U.S. governance has allowed the evolution of a rather complicated governmental system of public health. There exist many overlapping roles in this system. Although many functions are shared in this system, several significant problems exist. Decision making is fragmented between levels of government, creating problems of coordination leading to administrative confusion. There is a lack of congruence between organizations and functions. The system leaves the local health department, which is the entity with the most contact with citizens and the greatest service responsibility, with the least authority and the fewest resources.

The Future of Public Health

In 1988, the Institute of Medicine's Committee for the Study of the Future of Public Health published its report, *The Future of Public Health.* This succinct 159-page report concluded that the public health system was in disarray, and it called for a complete overhaul and refocus of the governmental public health functions and organization.

Key among the report's recommendations was an articulation of the core functions of public health as assessment, policy development, and assurance, including recommendations that

- every public health agency regularly and systematically collect, assemble, analyze, and make available information on the health of the community, including statistics on health status, community health needs, and epidemiologic and other studies of health problems;
- every public health agency exercise its responsibility to serve the public interest in the development of comprehensive public health policies by promoting the use of the scientific knowledge based in decision making about public health and by leading in developing public health policy. Agencies must take a strategic approach, developed on the basis of a positive appreciation for the democratic political process;
- public agencies assure their constituents that services necessary to achieve agreed-on goals are provided, either by encouraging actions by other entities (private or public sector), by requiring such action through regulation, or by providing services directly; and
- each public health agency involve key policymakers and the general public in determining a set of high-priority personal and community-wide health services that governments will guarantee to every member of the community. This guarantee should

include subsidization or direct provision of high-priority personal health services for those unable to afford them.

Specific activities were recommended for agencies at the federal, state, and local levels of government, as well as a series of strategies for linkages with other governmental authorities, building system capacity, and educating professionals. Since that report was published, substantial efforts are under way in public health agencies at all levels to strengthen performance in the core functions.

A subsequent report, *The Future of the Public's Health in the 21st Century*, was issued in 2003, calling for strengthening of the governmental public health infrastructure, and the creation of a broader system of public-private collaboration to ensure the health of the population. Steps in that direction are critical to the health and well-being of the population of the United States.

—Margret O'Neall and H. Douglas Adams

See also Centers for Disease Control and Prevention; Council of State and Territorial Epidemiologists; Surgeon General, U.S.; U.S. Public Health Service

Further Readings

Institute of Medicine. (1988). *The future of public health.* Washington, DC: National Academies Press.

Institute of Medicine. (2003). *The future of the public's health in the 21st century.* Washington, DC: National Academies Press.

Novick, L. F., & Mays, G. P. (2005). *Public health administration: Principles for population-based management.* Gaithersburg, MD: Aspen.

Pickett, G., & Hanlon, J. J. (1990). *Public health: Administration and practice.* St. Louis, MO: Mosby.

Schneider, M. (2006). *Introduction to public health* (2nd ed.). Boston: Jones & Bartlett.

Turnock, B. J. (2003). *Public health: What it is and how it works* (2nd ed.). Gaithersburg, MD: Aspen.

Graphical Presentation of Data

Graphs, tables, text boxes, and sentences can all be used to communicate quantitative information in epidemiology. Graphs present the big picture; they show patterns and trends. Tables contain the details, so they are useful for looking up specific values. Although sentences can provide a small amount of numerical data clearly, the numbers are easily lost in a page of text. Therefore, text boxes are useful for highlighting these numbers. This entry discusses how to choose between tables and graphs, the advantages of each, and guidelines for effective tables and graphs.

Choosing Tables or Graphs

Graphs are preferable for some situations and tables for others. The following list points out the advantages of each:

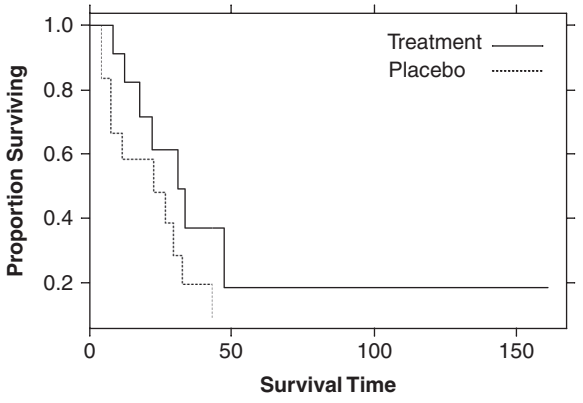
- Graphs show the big picture: patterns, trends, correlations, and the general shape of the data, while tables show exact values and offer precision.
- It is easy to detect extreme values such as the maximum and the minimum in a graph, while it is more difficult to do so with a table.
- Outliers, which are data points far from the rest of the data, are easier to spot on a graph, while they are more difficult to detect on a table.
- Graphs help discover data errors since problems such as an average value greater than a maximum value can be spotted easily. Tables are less effective in highlighting data errors.
- A great deal of information can be shown in a small space with a graph, while this is not true with tables.
- Graphs are appropriate for paper documents, Web documents, computer screens, or projector screens, while large tables are not appropriate for projector screens since the audience cannot see the details.
- Tables are useful for looking up values, while graphs offer only approximate values.
- Tables can accommodate a number of variables with different units of measures more easily than can graphs.
- The reader can use the data from tables for other purposes such as calculations and drawing other figures. This is more difficult with data from graphs.

Designing Graphs

Good graphs are powerful tools to visualize and understand data. Unfortunately, graphs can also confuse, mislead, or even deceive. This section provides principles of graph design to enable the reader to design effective graphs and to avoid common mistakes.

The data stand out in an effective graph. The designer should emphasize the data and de-emphasize

Graph



Table

Subjects With Adverse Events Reported

Adverse Event	*Treatment A # N = 247*	*Placebo # N = 238*
Headache	12	3
Nausea	10	4
Skin Rash	7	2
Dizziness	5	2
Fatigue	2	3

Sentence

The number of patients who received Treatment A was 247, while 238 patients received the placebo.

Text Box

Treatment A	247
Placebo	238

Figure 1 Examples of a Graph, a Table, a Sentence, and a Text Box (Not All the Same Data)

everything else. Grid lines, if used, should be in the background. This can be achieved by making them a pale gray or by using a dotted line. The plotting symbols and lines should be visually prominent; care must be taken so that the symbols are not hidden by tick marks, axes, grid lines, other data points, or other graphical elements. Clutter has no place in a graph. Too many tick marks or tick mark labels are a form of clutter. So are too many decimal places. The number of decimal places in labels should be appropriate for the data.

One form of clutter is adding a pseudo third dimension to bars, pies, lines, or other graphical elements. The unnecessary dimension often distorts the data. It always adds clutter. If bars are drawn with depth, the reader does not know how to read the bar. Is the value read from the front where the arrow on the X bar points or from the back where the arrow on the Y bar points? It turns out that the way to read these bars depends on the software that was used to create them, but the reader rarely knows what software was used. Two-dimensional bar charts are unambiguous. If the designer chooses to use a bar chart and knows the

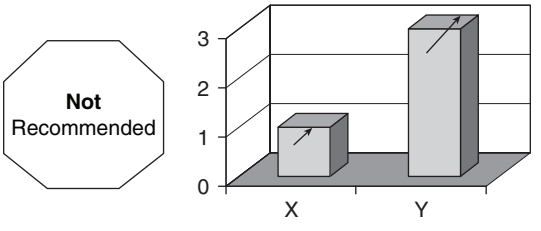


Figure 2 The pseudo third dimension distorts data.

categories and values of the data, then a two-dimensional bar chart should always be the choice.

Another common mistake is using equally spaced tick marks for unequal intervals. For example, if patients are seen at baseline and then after 1 month, 3 months, 7 months, and 1 year, then the figure should reflect these times, and the bars or symbols should not be equally spaced on the horizontal axis.

Color can be a powerful means of distinguishing groups of data, but it can also be a form of clutter. Consistency is important for color as well as for other graphical elements. It is distracting to view a series of graphs when red is used for Treatment A and blue for Treatment B on the first graph and then these colors are reversed in the next figure or completely different colors are used. It is also distracting if each bar or symbol is plotted with a different color for no apparent reason.

One graph can be acceptable by itself but confusing when in a group of charts. When graphs are shown in a group, they should have consistent scales if possible. Sometimes the same scale on all graphs of the group hides the data of some of the figures since the range of data is so different. In that case, the different scales should be emphasized either in the caption or by showing relative scales graphically.

Creating More Effective Graphs by Naomi Robbins (2005) and *The Elements of Graphing Data* by William Cleveland (1994) provide useful graph forms for presenting data and principles for drawing effective graphs. *Visualizing Data* by William Cleveland (1993) discusses graphs useful for statistical analyses.

Designing Tables

Tables are useful tools for storing data, looking up data values, and presenting data to readers who require precision. This section provides a few tips for designing effective tables. *Show Me the Numbers: Designing Tables and Graphs to Enlighten* by Stephen Few (2004) provides excellent advice about spacing rows and columns, using or avoiding grid lines, and other aspects of designing tables.

It is important to round data when presenting them in a table. Too many decimal places clutter the table and distract the reader from the values of interest.

Order the rows and columns in a sensible fashion. Determine the purpose of the table before designing it. If the purpose is data lookup, then alphabetical order is often best. If, on the other hand, the reader needs a sense of the distribution of the data, then ordering by size makes more sense. The order of the rows and columns has a major impact on the understanding of the data and the ease of making comparisons from the information in the table.

Eliminate nonessential information. Clutter interferes with understanding data from tables just as it does with graphs.

Highlight key data. Presentations often include tables without any attempt to direct the reader's attention to the particular numbers that support the point the presenter is trying to make. Highlighting those numbers can help.

Differentiate summary data such as means or totals from the rest of the data. This can be accomplished with spacing or by including vertical or horizontal lines.

Combining Tables and Graphs

It is often useful to show the data in more than one way. Each presentation adds different insights into the data. Including both a table and a graph shows the details as well as the big picture. Richard Heiberger and Burt Holland (2007) propose microplots to help

treatment	*min.*	*m – sd*	*mean*	*m + sd*	*max.*
25.cycle	54.13	61.13	69.83	78.58	86.08
60.cycle	46.31	57.73	65.22	72.71	79.71
faradic	38.89	51.91	57.68	63.45	67.96
galvanic	45.07	52.86	60.13	67.39	73.54

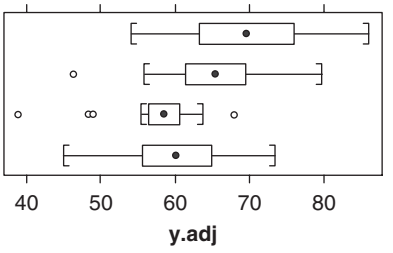


Figure 3 The microplot (on right) helps visualize the results of the table.

Source: Heiberger and Holland (in press). Used with kind permission of Springer Science and Business Media.

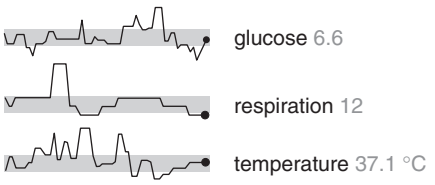


Figure 4 Edward Tufte's sparklines integrate text, numbers, and graphics.

Source: Tufte (2006, p. 47). Used with permission of Edward Tufte.

visualize the results of the tables often included with the results of statistical analyses.

Edward Tufte (2006) suggests integrating text, numbers, and graphics. One way to do that is with sparklines, which are intense, simple, word-sized graphics. The shaded areas represent the normal range of glucose, respiration, temperature, and white blood count readings.

Conclusion

Both tables and graphs are useful for presenting data. It is helpful to include both when time and costs permit since graphs show the forest while tables show the trees. Think about the data and the purpose in displaying them before designing a table or graph so that your message is clear and you emphasize the comparisons that are the most important.

—Naomi B. Robbins

See also Bar Chart; Box-and-Whisker Plot; Causal Diagrams; Histogram; Pie Chart; Population Pyramid; Receiver Operating Characteristic (ROC) Curve; Scatterplot; Stem-and-Leaf Plot

Further Readings

Cleveland, W. S. (1993). *Visualizing data.* Summit, NJ: Hobart Press.

Cleveland, W. S. (1994). *The elements of graphing data* (Rev. ed.). Summit, NJ: Hobart Press.

Few, S. (2004). *Show me the numbers: Designing tables and graphs to enlighten.* Oakland, CA: Analytics Press.

Heiberger, R. M., & Holland, B. (in press). Structured sets of graphs. In C. Chen, W. Hardle, & A. Unwin (Eds.), *Handbook of data visualization*. New York: Springer.

Robbins, N. B. (2005). *Creating more effective graphs.* Hoboken, NJ: Wiley.

Tufte, E. (2006). *Beautiful evidence.* Cheshire, CT: Graphics Press.

Graunt, John
(1620–1674)

John Graunt was a storekeeper from London, England, who is credited with being the founding father of the science of vital statistics. Before Graunt's time, public health surveillance was unprecedented, and no method was available to quantify disease patterns in the population. Graunt provided a statistical and analytical approach to examine the population's health status. He collected data from the *Bills of Mortality* to study the disease patterns in London. The *Bills of Mortality* were documents that were published weekly by the London parishes and offered information on the number of births, deaths, and cause of death in each parish.

The *Bills of Mortality* were an untapped data source for Graunt, who was able to organize the data to compile mortality tables, allowing him to compare the trends in mortality and natality by the season, year, and geographic area. On the basis of his work, he published a book in 1662 entitled *Natural and Political Observations Mentioned in a Following Index and Made Upon the Bills of Mortality.* This 79-page book provided the first example of descriptive statistics. For example, Graunt was able to determine the number of deaths due to acute or chronic illnesses, the number of maternal deaths during childbirth, and the number of deaths from the plague. He scrutinized the data available and reported on problems with disease classification, reporting bias with regard to some diseases such as syphilis, irregular data collection intervals, and other inconsistencies in data collection and reporting.

Graunt's work was important to public health because he developed principles of epidemiology and demography. He made inroads by discovering patterns of disease and was able to identify diseases afflicting a geographic area or gender. He described and quantified disease occurrence in London at that time, and although he was not trained in mathematics, he was able to interpret the statistics. Graunt reported that more boys were born than girls, women had a longer life expectancy than men, the ratio of boys to girls was stable over time, mortality rates were

highest among infants and older adults, and death rates were higher in urban than in rural areas. He also identified variation in mortality rates by season and year and calculated the population of London at the time.

Graunt was also a councilman and politician. He served as a member of the Common Council of London. More important, Graunt was also a member of the Royal Society. King Charles II was so much in awe of Graunt's work that he recommended Graunt for membership in the newly created Royal Society, which was a forum for the nation's leading scientists. Many of the scientists at the time were opposed to Graunt's membership because Graunt did not have a formal education. However, King Charles overruled the objections, allowing Graunt to become a Fellow of the Royal Society. Graunt died of jaundice at the age of 53.

—Britta Neugaard

See also Applied Epidemiology; Demography; Life Tables; Mortality Rates

Further Readings

Rothman, K. J. (1996). Lessons from John Graunt. *Lancet, 347*(8993), 37–39.

Wain, H. (1970). The bookkeeping of life and the birth of vital statistics. In *A history of preventive medicine* (pp. 121–125). Springfield, IL: Thomas.

Wynder, E. L. (1975). A corner of history: John Graunt, 1620–1674, the father of demography. *Preventive Medicine, 4*(1), 85–88.

Gulf War Syndrome

The term *Gulf War syndrome* or *Gulf War illness* refers to a number of symptoms reported by American veterans who served in the 1991 Persian Gulf War, including chronic fatigue, headaches, dizziness, memory problems, gastrointestinal problems, and skin problems. Several studies have shown that self-reported health conditions are higher among veterans deployed to the Persian Gulf than to veterans serving elsewhere, but establishing higher rates of clinically defined diseases has proven elusive. The wide range of symptoms reported, the fact that many of the symptoms could have multiple causes, the lack of objective verification for some of the symptoms, and the wide range of potential causes of the symptoms have made research into Gulf War syndrome difficult.

Military personnel serving in the Gulf War were potentially exposed to numerous health hazards, including dust and sand particles, smoke from oil well fires, insecticides, vaccinations, depleted uranium, and psychological and physiological stress. However, linking exposures with specific health conditions in individuals has proven difficult for several reasons, the most important of which is that data quantifying the exposure of particular individuals to specific health threats are not available. Most estimates of exposure are based on self-report, which is subject to recall and other types of bias. Another difficulty is that only limited medical information is available for the veterans before and after deployment, making it impossible to establish a baseline for health status that would serve as a standard of comparison for health status after deployment.

Numerous studies have been conducted investigating the health of Gulf War veterans and evaluating the effects of the hazards they were exposed to. In 1998, the Institute of Medicine (IOM) began a series of congressionally mandated reports to evaluate and summarize the available scientific and medical literature regarding these issues. Volume 4 of the IOM reports, *Health Effects of Serving in the Gulf War*, summarizes all scientific and medical peer-reviewed literature available in 2006 that addresses the health status of veterans deployed in the Persian Gulf. This report found no evidence for the existence of a unique "Gulf War syndrome," although it did find that Gulf War veterans were at increased risk for anxiety disorders, depression, and substance abuse problems, and found evidence for a possible elevated risk of amyotrophic lateral sclerosis (ALS). No evidence was found for increased cancer rates among Gulf War veterans, although there was some suggestion that brain cancer rates might be higher, and the IOM recommended follow-up studies to examine this connection further. The IOM report found that self-reported multisymptom illnesses were higher for Gulf War veterans both from the United States and from other countries that sent troops to the Persian Gulf. Gulf War veterans were more likely to be injured or die in a traffic accident in the first few years following their return but not in later years. Rates of hospitalization were similar among deployed and nondeployed veterans, and there was no evidence of an increase in

cardiovascular disease or peripheral neuropathy. Evidence on respiratory illness was inconsistent, although some studies found a link between exposure to smoke from oil well fires and asthma. Evidence of the presence of birth defects in children of Gulf War veterans was also inconsistent.

—*Sarah Boslaugh*

See also Bias; Environmental and Occupational Epidemiology; Exposure Assessment; Government Role in Public Health

Further Readings

Institute of Medicine. (2006). *Gulf War and health: Vol. 4. Health effects of serving in the Gulf War*. Washington, DC: National Academies Press.

Web Sites

Centers for Disease Control and Prevention, Gulf War Studies: http://www.cdc.gov/nceh/veterans/default2.htm.

H

Halo Effect

The halo effect refers to an error in reasoning where an impression formed due to a single trait is allowed to influence multiple judgments or ratings of unrelated factors. For instance, the halo effect operates when a rater forms a general impression of another person, based on one outstanding trait, and that general impression is allowed to influence judgments or ratings that should instead be based on specific aspects of the person. For example, people who are attractive may also be judged to be good workers without regard to their actual work performance; the positive impression of their attractiveness clouds the rater's ability to judge the actual quality of their work, although these two traits are not related. These overall impressions misrepresent the specific traits a person may have because they are based on a small amount of information.

The American psychologist Edward L. Thorndike (1874–1949) was a pioneer in studying the phenomenon of the halo effect. For instance, it was clearly present in an experiment conducted with servicemen in 1920, which was reported in his article "A Constant Error on Psychological Rating." In the experiment, the commanding officers were asked to rate their subordinates on their intelligence, physique, leadership, and character without having spoken to them. Thorndike noted a correlation between unrelated positive and negative traits. The service members who were found to be taller and more attractive were also rated as more intelligent and as better soldiers. Thorndike determined from this experiment that people generalize from one outstanding trait a person has to form a favorable view of the person's whole personality.

Another psychologist who studied the halo effect was Polish American Solomon Asch (1907–1996). In his 1946 article "Forming Impressions of Personality," Asch delineated how people form impressions of one another. Asch found that impressions of others were formed by a "primacy effect." First impressions were established as more important than subsequent impressions in forming an overall impression of someone. Participants in the experiment were read two lists of adjectives that described a person. The adjectives on the list were the same but the order was reversed; the first list had adjectives that went from positive to negative, while the second list had the adjectives in reverse order, from negative to positive. How the participant rated the person depended on the order in which the adjectives were read. Adjectives presented first had more influence on the rating than adjectives presented later. When positive traits were presented first, the participants rated the person more favorably; when the order was changed to introduce the negative traits first, that person was rated less favorably.

—Britta Neugaard

See also Bias; Measurement; Questionnaire Design

Further Readings

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258–290.

Thorndike, E. L. (1920). A constant error on psychological rating. *Journal of Applied Psychology, 4*, 25–29.

Hamilton, Alice
(1869–1970)

Alice Hamilton's groundbreaking epidemiologic studies investigating the occupational exposures of workers to poisonous substances used in manufacturing were critical to the development of the field of industrial toxicology and epidemiology. Hamilton's work contributed to the development of regulations enforcing healthier conditions in America's workplaces. Her career powerfully illustrates how epidemiologic research can stimulate and inspire scientific inquiry to serve the public good.

Hamilton was born in 1869 to a close-knit, patrician family in Fort Wayne, Indiana. Trained as a physician in the medical department of the University of Michigan, she joined the faculty at the Woman's Medical School of Northwestern University in Chicago as professor of pathology in 1897. For 22 years, Hamilton was a resident of Hull House, a famous American settlement in Chicago founded to connect the privileged classes and the socially disadvantaged, and this experience cultivated her interest in service and activism. She found her life's work in 1910, when she was invited to lead a state-funded study of industrial diseases, the first large-scale study of this type. She went on to conduct epidemiologic research of industrial diseases as a special investigator for the federal Bureau of Labor from 1911 until her final report in 1940 (see Table 1).

In 1919, Hamilton became Harvard University's first woman professor when she was invited to join the newly created Department of Industrial Hygiene at Harvard Medical School, a position she held until her retirement in 1935. Her professional work in public health continued until she was 80, and a new edition of her textbook *Industrial Toxicology* was published in 1949. Three months after Alice Hamilton died at age 101 in 1970, Congress passed the Occupational Safety and Health Act.

A Research Example: Lead Poisoning in Pottery Trades

Hamilton's monograph, "Lead Poisoning in Potteries, Tile Works, and Porcelain Enameled Sanitary Ware Factories," exemplifies her use of many modern epidemiologic principles, including the influence of case ascertainment and methods of comparing susceptible

Table 1 Selected Monographs Authored by Alice Hamilton for the *Bulletin of the U.S. Bureau of Labor Statistics*

Year Published	*Industry*	*Principal Exposure(s) Studied*
1912	Potteries, tile, and porcelain	Lead
1913	Painting	White lead
1914	Lead smelting and refining	Lead
1914	Battery manufacture	Lead
1915	Rubber industry	Lead, antimony, benzene
1917	Explosives industry	Nitrous fumes, TNT
1917	Printing trades	Lead poisoning, wood alcohol
1918	Stonecutters	"Dead fingers syndrome"
1921	Coal tar dyes and dye intermediates	Aniline, nitrobenzene, toluene, xylene, inorganic compounds (e.g., hydrogen sulfide, hydrogen arsenide)
1922	Steel manufacturing	Carbon monoxide
1940	Viscose rayon	Carbon disulfide, hydrogen sulfide

groups. The report documented her investigation of 68 potteries and factories in nine U.S. states. Information was collected directly from physicians, hospital records, personal inquiries, and examinations. The investigation identified industry practices that increased risk of lead poisoning and compared male and female prevalence of lead poisoning.

As part of the study, Dr. Hamilton identified 18 cases of lead poisoning that occurred in 1910 and 1911 among 314 men employed in Trenton, New Jersey, for a 2-year cumulative incidence of 1 in 31, and compared that with the prevalence in East Liverpool, Ohio, where of 480 men, 31 cases were found in the same time period, or 1 in 15 to 16 employed. (The terminology *2-year cumulative incidence* is used here for clarity. Dr. Hamilton simply reported these as "ratios" or "per cents.")

The report pointed out several potential reasons for the differences between the two cities. Case identification was difficult in general due to absence of employee lists, high turnover in the lead industry, workers' failure to seek treatment, ignorance in the medical community of the symptoms of lead poisoning when medical attention was sought, and in some cases, management cover-up. Hamilton stated that the number of cases in Trenton may have in truth been higher because doctors in Trenton were less aware of whether their patients worked in potteries, whereas in East Liverpool, "everyone knows about the potteries . . . and every doctor has potters and girl helpers among his patients" (Hamilton, 1912, p. 44). Second, because other employment opportunities were available in Trenton, those cases with moderate symptoms may have left the pottery trade for other work, an option that did not exist in East Liverpool. This explanation, whereby only healthy workers are left in an occupation, is now known as the "healthy worker effect." A third explanation offered was that decorating was done more in East Liverpool compared with Trenton, carrying with it greater risks. Finally, the so-called sanitary ware was made in Trenton, but not in East Liverpool. The manufacture of sanitary ware required glazes with lower lead concentrations and practices that created less dust. The report states, "There is no dry rubbing and no dusty gathering and piling together of ware, and no women's skirts stirring up dust, for the employees are all grown men" (Hamilton, 1912, p. 46).

The report also documented higher incidence of lead poisoning in women compared with men in East Liverpool, where the 2-year cumulative incidence was 25 cases per 135 female employees, compared with 31 per 480 males employed. When limiting the comparison to only those employed in the glaze room, the incidence of lead poisoning among women was approximately two times higher than among men (Table 2). She further pointed out that "the contrast between the men and women becomes still greater when one takes into consideration the fact that the average period of employment for the men dippers was 19.5 years and for the women helpers only 2.5 years" (Hamilton, 1912, p. 48).

Table 2 Ratio of Lead-Poisoning Cases Among Dippers and Dippers' Helpers in White Ware Potteries to Number Employed, East Liverpool (Ohio)

Sex	*Employees*	*Cases of Lead Poisoning, and Still at Work*	*Ratio of Cases to Employees of Each Sex*
Male	85	13	1 to 6 or 7
Female	41	14	1 to 3

Source: Hamilton (1912).

—June M. Weintraub and Yvonne L. Michael

See also Cumulative Incidence; Environmental and Occupational Epidemiology; Lead

Further Readings

Hamilton, A. (1912). *Lead poisoning in potteries, tile works, and porcelain enameled sanitary ware factories*. Washington, DC: Government Printing Office.

Hamilton, A. (1943). *Exploring the dangerous trades: The autobiography of Alice Hamilton, M.D.* Boston: Little, Brown.

Sicherman, B. (1984). *Alice Hamilton: A life in letters.* Cambridge, MA: Harvard University Press.

Hardy-Weinberg Law

Genetic epidemiology aims to understand how genetic variation contributes to disease risk. To best characterize risk relationships, one must first understand how the exposure of interest is distributed within and across populations, before attempting to relate

exposure distributions to disease distribution. For genetic exposure, this involves understanding how genetic variation arises in populations and how it is maintained within and across populations over time. The Hardy-Weinberg law describes a state of equilibrium in allele frequencies at a particular genetic locus over generations that are randomly mating. This law also describes a relationship between the allele frequencies and genotype frequencies within a population as a result of random mating.

Variation at a genetic locus can be described by noting the different "spellings," called "alleles," that exist at the same location among chromosomes in a population. For example, the sequence at a particular site may be AT**C**C in some and AT**T**C in others, which would be referred to as two alleles "C" or "T." When such diversity exists, the location is considered to be "polymorphic." Because humans are "diploid," each person carries two copies of the genome, one from their father and one from their mother. So at any polymorphic location, each person carries two alleles, one from each parent. The particular combination of the two alleles carried by a single individual is referred to as a "genotype." Following the example above, there are three possible genotypes for the C/T polymorphism with two possible alleles: CC, CT, TT. The Hardy-Weinberg law characterizes the relationship between alleles and genotypes in a population due to random mating and the equilibrium state of allele frequencies from generation to generation. It is thus often also referred to as Hardy-Weinberg equilibrium (HWE).

Traditionally, the frequency of the first allele is denoted as p and the alternative allele frequency as q. These are the proportions of that particular allele among all chromosomes in the population (e.g., among $2 \times N$, where N is the number of people), and $p + q = 1$. Under an assumption of random mating in a sexually reproducing diploid population with no other population genetic forces such as mutation, natural selection, migration, or drift, it can be shown that the expected genotype frequencies are a specific function of the allele frequencies p^2, $2pq$, and q^2 (see Table 1) and that these values, p and q, will remain constant over generations. Proof of this result was reported by three separate papers in the early 1900s, by Castle (1903), Hardy (1908), and Weinberg (1908). As an example, suppose a population of 10,000 people contained a genetic polymorphism with alleles C and T, where 11,000 of the 20,000 genomes in that population contained a C allele ($p = 11,000/20,000 = 0.55$; $q = 0.45$). The Hardy-Weinberg law, which assumes random mating, would expect the genotype frequencies in the population to be CC genotype $= p^2 = 0.552 = 0.3025$, or 3,025 people with CC; $CT = 2pq = 2 \times 0.55 \times 0.45 = 0.495$, or 4,950 people with CT; and $TT = q^2 = 0.452 = 0.2025$, or 2,025 people with TT. These genotype proportions based on allele frequencies are often referred to as Hardy-Weinberg proportions.

One can measure the amount of departure from Hardy-Weinberg expectations by comparing the observed genotype frequencies in a sample to those expected under HWE based on the allele frequencies for that sample. This value is considered the Hardy-Weinberg disequilibrium coefficient: DHW = observed genotype frequency − expected genotype frequency. For example, if among 1,000 people, 350 were CC, 400 were CT, and 250 were TT, one could calculate the allele frequencies as $p = (2 \times 350 + 400)/2,000 = 0.55$ and $q = 0.45$. The expected genotype frequencies under HWE would be $E(CC) = p^2 = 0.552 = 0.3025$; $E(CT) = 2pq = 2 \times 0.55 \times 0.45 = 0.495$; $E(TT) = q^2 = 0.452 = 0.2025$. The DHW = observed − expected = $0.350 - 0.3025 = 0.0475$. This could also be calculated using the other homozygous genotype: DHW = $0.250 - 0.2025 = 0.0475$. One could test the statistical significance of this by testing the hypothesis that DHW = 0 versus the alternative $D^{HW} \neq 0$ using a z test:

$$z = \frac{\hat{D}^{HW} - \varepsilon(\hat{D}^{HW})}{SE(\hat{D}^{HW})}$$

or through likelihood ratio testing. Because there are often sparse genotype cells for situations with rare alleles, exact tests or permutation approaches are often employed.

Table 1 Hardy-Weinberg Proportions: Assumption of Random Mating

		Males	
		C (p)	T (q)
Females	C (p)	CC p^2	CT pq
	T (q)	TC pq	TT q^2

Note: Expected genotype proportions: $P(CC) = p^2$; $P(CT) = 2pq$; $P(TT) = q^2$

In practice, human populations do undergo population forces such as mutation, selection, and nonrandom mating, yet the genotype frequencies often approach Hardy-Weinberg proportions, making it a very robust property. Departures from Hardy-Weinberg proportions can be detected under severe violations of the assumptions, such as nonrandom mating due to inbreeding or assortative mating; recent strong natural selection; or genetic drift. In addition, violations of Hardy-Weinberg proportions in a sample may be due to sampling error or sampling bias. For example, if a particular genetic locus is related to disease risk, the genotype frequencies among cases may be enriched for particular risk genotypes and may, therefore, not reflect the general population and not demonstrate Hardy-Weinberg proportions. This point has actually been exploited in some statistical genetics methods to detect genetic risk factors by testing for Hardy-Weinberg violations among cases. Finally, violations of Hardy-Weinberg proportions may simply reflect genotyping measurement error. In fact, examination of Hardy-Weinberg proportions is now a standard aspect of quality assessment for large-scale genotyping projects.

—Margaret Daniele Fallin

See also Gene; Genetic Epidemiology; Genetic Markers; Linkage Analysis

Further Readings

Castle, W. E. (1903). The laws of Galton and Mendel and some laws governing race improvement by selection. *Proceedings of the American Academy of Arts and Sciences, 35*, 233–242.

Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science, 28*, 49–50.

Nielsen, D. M., Ehm, M. G., & Weir, B. S. (1998). Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *American Journal of Human Genetics, 63*(5), 1531–1540.

Weinberg, W. (1908). Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, *64*, 368–382.

Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics, 76*(5), 887–893.

Harm Reduction

Noting that no single, widely embraced definition of harm reduction exists, the International Harm Reduction Association (IHRA) proposes that harm reduction be defined as ''policies and programs which attempt primarily to reduce the adverse health, social, and economic consequences of mood-altering substances to individual drug users, their families, and their communities'' (IHRA, n.d., § 14). Examples of mood-altering substances include heroin, cocaine, methamphetamine, alcohol, and tobacco. Because individuals who use mood-altering substances have engaged in collective efforts to reduce the harm of their substance use independently of programs and policies (and, in fact, these efforts have at times predated, and given rise to, programs and policies), IHRA's harm reduction definition should perhaps be extended to explicitly encompass user-initiated actions, undertaken collectively, that are designed to reduce the adverse consequences of mood-altering substance use. This entry discusses the key harm reduction principles of pragmatism, prioritization of goals, and humanism; reviews the history of efforts embodying these principles; and examines particular examples of harm reduction efforts and their effectiveness.

Harm Reduction Principles

Key harm reduction principles include the following:

• *Pragmatism.* The elimination of mood-altering substance use is not an attainable goal in the foreseeable future. Addiction to, or dependence on, a substance may preclude cessation for some individuals, particularly if treatment is inaccessible; individuals may also choose to continue using a particular substance because of the perceived benefits it brings. A pragmatic approach to reducing vulnerability to drug-related harms thus necessarily encompasses efforts to promote safer drug use practices among active substance users, as well as efforts to support individuals who wish to reduce or cease using particular substances.

• *Prioritization of Goals.* Harm reduction programs, policies, and collective user-initiated actions may prioritize their goals so that the most pressing needs of their target populations are addressed first. For example, efforts designed to reduce the spread of bloodborne infections via injection drug use may prioritize goals as follows: (1) reduce the likelihood that individuals will borrow used syringes; (2) minimize the risk that an individual will transition to injection

drug use from another mode of drug administration, and reduce the duration of injecting among current injectors; and (3) facilitate access to appropriate treatment. Importantly, there is no single, optimal method of reducing drug-related harms for all individuals in a population. Rather, multiple possibilities should be available simultaneously, and the same individual may participate in different harm reduction efforts over the course of his or her life.

- *Humanism.* Individuals who use mood-altering substances can and do make rational choices that further their health and well-being, as well as that of their families and communities. The rights and dignity of all individuals who use mood-altering substances merit respect. Drug users are members of broader communities in which they fill multiple social roles, including those of parent, partner, child, and neighbor; furthering users' health and well-being thus furthers community well-being. Users have been central to the development of harm reduction programs and policies, and their ongoing contributions to these efforts should be recognized, promoted, and respected.

These harm reduction principles guide a broad array of programs, policies, and collective user-initiated actions, including laws prohibiting driving while intoxicated and smoking in specified public spaces; syringe-exchange programs; and collective, user-initiated actions to protect promote users' health and that of the broader communities in which they are embedded. Evidence regarding the effectiveness of some of these harm reduction policies, programs, and collective user-initiated actions is presented.

Harm Reduction in Historical Context

The term *harm reduction* was coined in the mid-1980s; the first published use of the term appears to have been in 1987. This term initially emerged to describe efforts seeking to reduce the spread of infection (first hepatitis and later HIV) among people who injected drugs. Early harm reduction efforts sought to contain hepatitis B and C among injectors; in Edinburgh, Scotland, a pharmacist provided sterile syringes without a prescription to local injectors in 1982 and 1984, and in Amsterdam, Holland, substance users organized efforts to provide sterile injection equipment to active injectors in 1983. These efforts intensified and spread geographically with the discovery of the high prevalence of HIV and AIDS among injectors in multiple cities internationally, coupled with the knowledge that (1) HIV could be transmitted via injection equipment and (2) there was neither a cure nor a vaccine for the infection. The term *harm reduction* now also encompasses programs, policies, and collective user-initiated efforts to reduce a broad array of adverse consequences among users of a wide range of substances administered through multiple methods.

The first emergence of the term *harm reduction*, however, should not be confused with the first emergence of harm reduction principles. These principles—of pragmatism, prioritization of goals, and humanism—have been applied to further users' health and well-being in multiple prior contexts. For example, in the late 1800s and early 1900s, physicians in the United States maintained women and men who were addicted to morphine on the drug out of concern that withdrawal might prove fatal or permanently debilitating. Likewise, in the United Kingdom, a government-appointed committee of physicians concluded in 1926 that the practice of maintaining addicted individuals on morphine and heroin should be continued. The principles articulated in this report (the Rolleston Report) continued to influence British drug policy for the next five decades. Likewise, before AIDS had been discovered by science, injectors in New York City in the 1970s noticed the illness's symptoms, attributed it to shared drug paraphernalia, and altered their injection practices accordingly. The current commitment to reducing drug-related harms among active substance users, then, should be viewed as a resurgence or intensification of an existing approach rather than as an entirely novel phenomenon.

Harm Reduction: Assessing the Evidence

Assessing the evidence for the effectiveness of the full range of harm reduction policies, programs, and collective, user-initiated actions lies beyond the scope of this entry. Here, we briefly review findings from four harm reduction efforts.

Per Se Laws: Drinking and Driving

Traffic injuries and fatalities are among the harms produced by alcohol consumption. A dose-response

relationship exists between a driver's blood-alcohol concentration (BAC) and crash risk: Crash risks are 4 times greater at a BAC of 0.08% compared with a BAC of 0.00%, 10 times greater with a BAC of 0.10%, and 25 times greater with a BAC of 0.15%. Per se laws criminalize driving with a BAC that exceeds a particular cut point. Individuals whose BAC exceeds the limit can be prosecuted and/or have their license suspended or revoked. Evidence indicates that alcohol-involved accidents, injuries, and fatalities decline after the introduction of per se laws and after the reduction of the BAC limit in areas with existing per se laws. The magnitude of a per se law's effect may decay over time, particularly if enforcement is lax, and is enhanced if its passage is accompanied by public education campaigns, active enforcement, and the passage of other alcohol-related laws.

Public Clean Air Laws and Tobacco Smoking

Enacted by states and municipal governments, Public Clean Air Laws restrict the locations in which individuals can smoke tobacco. The restrictiveness of these laws can vary, with some laws prohibiting smoking in all work sites and restaurants, and others permitting smoking in designated sections of workplaces and restaurants. While the main impetus of these laws is to reduce exposure to secondhand smoke, their proponents have hypothesized that they may also reduce the opportunity to smoke among active smokers and increase the chances that smokers will quit. Compared with residents of other states, residents of states with highly restrictive clean air laws have (a) 12% lower per capita cigarette consumption rates; (b) 14% lower mean smoking prevalence rates; and (c) 12% higher mean quit rates.

Syringe-Exchange Programs

Syringe-exchange programs (SEPs) are programs in which individuals can acquire sterile syringes and other injection equipment and dispose of used syringes and equipment. Additional ancillary health services may also be available, including referral to drug treatment and instruction on safer injection methods. A large body of research indicates that SEPs are an effective method of reducing the spread of HIV among injection drug users. HIV incidence among SEP participants is substantially lower than that among nonparticipants (hazard ratio: 3.35; CI: 1.29, 8.65), and SEP participation is association with a two- to six-fold reduction in the odds of engaging in drug-related HIV risk behaviors. There is no convincing evidence that SEPs increase drug use frequency or hazardous drug use, either among participants or others in the local community.

Informed Altruism

Collective, user-initiated harm reduction efforts include "informed altruism." In this social process, individuals who plan to inject as a group and who are faced with an inadequate supply of injecting equipment collectively discuss their HIV-serostatus. The order in which people then share the available injection equipment is sequenced so that HIV-positive individuals receive the used equipment last. To date, no data exist on the impact of this collective action on the containment of HIV (and other infections) among injectors.

—Hannah L. F. Cooper

See also Alcohol Use; Drug Abuse and Dependence, Epidemiology of; Health Behavior; HIV/AIDS; Tobacco

Further Readings

Canadian Centre on Substance Abuse. (1996). *Harm reduction: Concepts and practice: A policy discussion paper*. Retrieved August 9, 2007, from http://www.ccsa.ca/pdf/ccsa-006491-1996.pdf.

Des Jarlais, D. C., Friedman, S. R., & Ward, T. P. (1993). Harm reduction: A public health response to the AIDS epidemic among injecting drug users. *Annual Review of Public Health, 14*, 413–450.

International Harm Reduction (Network) Association. (n.d.) *What is harm reduction?* Retrieved August 9, 2007, from http://www.ihra.net/Whatisharmreduction.

Marlatt, G. A., & Witkiewitz, K. (2002). Harm reduction approaches to alcohol use: Health promotion, prevention and treatment. *Addictive Behaviors, 27*, 867–886.

World Health Organization. (2004). *Effectiveness of sterile needle and syringe programming in reducing HIV/AIDS among injecting drug users*. Geneva, Switzerland: Author.

Harvard Six Cities Study

The Harvard Six Cities Study was a large-scale study of the effects on human health of fossil fuel

emissions, in particular, sulfur dioxide and respirable particulate matter (soot). It was inaugurated by Benjamin Ferris and involved many faculty members and researchers at Harvard University, including John Spengler, Douglas Dockery, and Frank Speizer. This study provided evidence that, in concert with similar results from other studies, prompted the Environmental Protection Agency to raise air quality standards.

One impetus for the study was a belief in the United States in the 1970s that, due to the Middle East oil embargo, high-sulfur coal would be used more widely as an energy source in the future. Coal was a well-known source of air pollution (e.g., the combination of pollution due to soft coal combustion and a temperature inversion was associated with a temporary doubling in the death rate in the London smog disaster of 1952), but it was uncertain whether the health effects observed were due to sulfur dioxide, particulate matter, or both. The Six Cities Study was designed in part to investigate this question. Ferris had observed methodological innovations used to study the effects of pollutant exposure on the health of coal miners in Wales and had already incorporated those methods into a study of sulfur emissions from a paper mill in Berlin, New Hampshire; he also applied those methods to the Harvard Six Cities Study.

The Harvard Six Cities Study followed a cohort of 8,111 Caucasian adults in six cities in the northeastern and Midwestern United States for 14 to 16 years, beginning in the mid-1970s. The cities included were Watertown, Massachusetts; Harriman, Tennessee; St. Louis, Missouri; Steubenville, Ohio; Portage, Wisconsin; and Topeka, Kansas. Questionnaires were used to collect data, including smoking history, educational level, age, sex, weight, height, medical history, and occupational exposure to gases, dusts, and fumes, from participants at enrollment and 3, 6, and 12 years afterward. Mortality data were collected through the National Death Index. Ambient air pollution was measured through centrally located monitors in each city.

The principal finding of the study was the positive association of air pollution and mortality. In particular, the study found that comparing residents of the most polluted to those of the least polluted city, higher ambient levels of respirable particulate matter and sulfur dioxide were associated with a 26% increase in mortality from all causes and that increased levels of respirable particles were associated with increased mortality from cardiopulmonary disease. In addition, the study found that the increase in mortality risk was directly proportional to increase in respirable particulate matter concentration. Mortality rate ratios were invariant for smokers and nonsmokers, and for people with and without occupational exposure to dusts, gases, and fumes. Increased air pollution was also associated with increase in a number of illnesses, including asthma and lung cancer.

The relationship between air pollution and mortality found in the Harvard Six Cities Study were confirmed in a much larger study, the American Cancer Society (ACS) study. Results from both studies were influential in the development of higher standards for air quality issued by the Environmental Protection Agency in 1997. These new regulations were questioned by representatives of various industries, who claimed that results from the ACS and Six Cities studies results could be explained by poor research design and data collection techniques, flawed statistical methodology, and factors not considered in the studies, such as temperature differences between cities. Because of confidentiality requirements, data from the ACS and Six Cities studies could not be released for reanalysis by industry representatives, so as a compromise, a third-party reanalysis was conducted by a nonprofit organization, the Health Effects Institute (HEI). This reanalysis, led by Daniel Krewski and Richard Burnett, included an audit of data quality, replication of the original studies, use of different statistical techniques, and inclusion of covariates such as climate, socioeconomic characteristics, and presence of other pollutants. The HEI reanalysis confirmed the original conclusions of both studies, although specific measures of risk were in some cases slightly higher or lower than in the original analyses. One new finding from the reanalysis was the association of lower levels of education with increased mortality.

A follow-up study by Francine Laden, Schwartz, Speizer, and Dockery (2006) based on the Six Cities Study found that reduction in respirable particulate matter was associated with reduced mortality. Results were controlled for the increase in adult life expectancy in the study and follow-up periods and found that a reduction of 1 $\mu g/m^2$ in the average levels of PM2.5 fine particulate matter (matter with a diameter of 2.5 μm or less) was associated with a 3% reduction in mortality, a reduction approximately equal to 75,000 fewer deaths per year.

—Sarah Boslaugh

See also Asthma; Cancer; Cardiovascular Disease; Environmental and Occupational Epidemiology; Pollution

Further Readings

Dockery, D. W., Pope, C. A., III, Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., et al. (1993). An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine, 329*(24), 1753–1759.

Krewski, D., Burnett, R. T., Goldberg, M. S., Hoover, B. K., Siemiatycki, J., Jerrett, M., et al. (2003). Overview of the reanalysis of the Harvard Six Cities Study and American Cancer Society Study of Particulate Air Pollution and Mortality. *Journal of Toxicology and Environmental Health, Part A, 66*(16–19), 1507–1551.

Laden, F., Schwartz, J., Speizer, F. E., & Dockery, D. W. (2006). Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities Study. *American Journal of Respiratory and Critical Care Medicine, 173*(6), 667–672.

Hawthorne Effect

From 1924 to 1927, Elton Mayo of the Harvard Business School, together with Fritz J. Roethlisberger and William J. Dickson, carried out a series of experiments with the level of illumination in a factory called the Hawthorne works of the Western Electric Company in Illinois. After informing the workers that they would be observed to assess their level of productivity, the level of illumination was varied in the factory. In some cases, during the experiment a person associated with the research remained on the factory floor. The researchers expected to find that illumination correlated with increased productivity; however, worker productivity increased throughout the experiment regardless of the level of light. In fact, in one experiment, the level of illumination was decreased steadily; productivity increased until the lights were so low that the workers complained that they were unable to see well enough to work.

The findings of these studies have been widely reported as proof that people change their behavior whenever they know that they are being observed. Furthermore, there is a suggestion that when people feel included in the decision process (as they were in some of the Hawthorne studies), they are empowered and tend to work harder.

The Hawthorne studies have been criticized for poor experimental design. Not all the experiments included a control group, some studies involved very small numbers of workers, and worker turnover may have influenced the results. Portions of the Hawthorne effect have also been attributed to an observer effect, because of the experimenter who remained on the factory floor. While these criticisms are valid, the Hawthorne effect itself continues to be observed in a variety of settings. It has been cited as an explanation for results of studies in widely divergent areas, including patients' perceptions of postsurgical recovery and quality of life, effectiveness of training on reduction of infection rates in day care centers, and the impact of repeated assessments of smoking on the rates of smoking among adolescents.

The Hawthorne effect should certainly be considered when designing an epidemiologic trial. For example, randomization to experimental and control groups will help control the tendency of people to behave differently while in the study. The schema for the control group should also be as similar to the experimental groups as possible; if a placebo is used in this group, it should be formulated to look similar to and be administered in the same way as the experimental drug. These components of the study design would help address the tendency of patients in the trial both to be more compliant in their medication use during the study than they would be in their daily lives and to report improvements in their symptoms—both of which are possible results of a Hawthorne effect and thus would be observable even in the control group. Because of this tendency, the Hawthorne effect is often equated with the widely publicized "placebo effect."

—Felicity Boyd Enders

See also Bias; Control Group; Placebo Effect; Randomization; Study Design

Further Readings

Wickstrom, G., & Bendix, T. (2000). The "Hawthorne Effect": What did the original Hawthorne studies actually show? *Scandinavian Journal of Work, Environment & Health, 26*(4), 363–367.

Hazard Rate

The hazard rate is a measure used to quantify the relative frequency of disease occurrence in a population

and plays an important role in descriptive and etiologic investigations. The definition of hazard rate is given as the instantaneous change in the occurrence of new cases per unit change of time at a particular time point, relative to the size of the disease-free population at that particular point in time. It is a theoretical limit of the rate as the time interval goes to zero and could seldom be obtained in reality. Instead, an *average rate* for a given period is usually estimated and used in epidemiology, analogous to the use of speed as an estimate of average velocity. This average rate is called the *incidence rate*, *force of morbidity or mortality*, or *incidence density*.

Typically, the incidence rate is estimated from studies that involve the follow-up of a population, such as cohort studies. The estimate of incidence rate takes into account both the number of new cases and the size and follow-up time of population:

$$\text{Incidence rate} = \frac{\text{Number of new cases in a given period}}{\text{Total person} - \text{time of observation}}.$$

The denominator is the sum of each individual's follow-up time until the occurrence of disease or until the end of the study. Therefore, the above calculation accounts for the situation that different individuals were observed for different lengths of time. As a hypothetical example, suppose that eight healthy 60-year-old women without coronary heart disease (CHD) were followed up to study the incidence rate of CHD. One of them developed CHD after 1 year of follow-up, two of them developed CHD after 3 years of follow-up, one of them was lost to follow-up without developing CHD after 4 years, and the last four did not develop CHD after 5 years at the end of the study. Then there are three new cases within the 5-year study period and total person-time of observation $= (1 \times 1) + (2 \times 3) + 4 + (4 \times 5) = 31$ person-years and the incidence rate $= 3/31$ person-years $= 0.097$/person-years. For most chronic diseases, especially incurable conditions, such as diabetes and multiple sclerosis, the deathrate, hazard rate, and incidence rate often include only the first occurrence of new cases. For recurring disease such as cancer or heart disease, both first and subsequent occurrence could be of great interest.

The numerical value of hazard rate and incidence rate has a low bound of zero but has no upper bound. Their interpretability depends on the selection of the time unit. The above incidence rate of 0.097/person-years could be expressed as 0.008/person-months, or 97/1,000 person-years. It is thus essential in presenting incidence rates with appropriate time units. For clarity, the numerator is often expressed as a power of 10. Incidence rate should not be confused with prevalence, which is defined as the number of individuals with a certain disease in a population at a specified time divided by the population size at that time.

—Rongwei (Rochelle) Fu

See also Incidence; Mortality Rates; Person-Time Units; Prevalence; Rate; Survival Analysis

Further Readings

Rothman, K. J., & Greenland, S. (1998). Measures of disease frequency. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 29–46). Philadelphia: Lippincott Williams & Wilkins.

Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic research: Principles and quantitative methods.* New York: Wiley.

Health, Definitions of

Epidemiology is often defined as the study of the occurrence and determinants of disease in human populations. Another way to look at epidemiology is to define it as the study of the *health* of human populations, which requires a definition of what is meant by health. The definition of what constitutes health is partly dependent on culture and historical period, so this entry concentrates on two general ways of conceptualizing health that are in common use today in the industrialized world: the medical model and the holistic model.

The medical model defines health as the absence of disease or injury, so a healthy person is one who is not suffering from a disease or injury as defined by current medical practice. This model was common in the industrialized world in the 20th century and is still common in the medical profession. It places an emphasis on treating and curing diseases and injuries after they occur, and it does not focus on prevention. This model also emphasizes diseases that have clearly observable signs and symptoms and that can be treated or cured, and it often places greater weight on physical rather than mental disease and on curing acute diseases rather than enhancing the quality of life of those with chronic diseases.

The holistic model is more commonly used in public health. Those advocating this model often cite the definition of health included in the Preamble to the Constitution of the World Health Organization (WHO), which was adopted in 1946 and entered into force in 1948. The Preamble defined health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" and further stated that "the enjoyment of the highest attainable standard of health" was a fundamental right of every human being (WHO Constitution; unpaginated). Health as a human right was a new concept when this statement was first made, but it has become accepted as a goal if not a reality by many people working in public health.

The WHO definition has been criticized as being both utopian and unmeasurable: Critics point out that every unfortunate aspect of human life, from warfare to religious oppression, could be considered within the scope of "health" by this definition, and further that with so inclusive a definition, the terms loses its meaning because hardly anyone could actually be considered to be healthy. Supporters of the holistic definition counter that many nonmedical aspects of life affect health, and that beginning with a broad rather than narrow model allows consideration of many threats to health, while requiring sound judgment as to which are the highest priority for intervention.

—Sarah Boslaugh

See also Complementary and Alternative Medicine; Epidemiology, History of; Ethics in Public Health; Public Health, History of

Further Readings

Callahan, D. (1973). The WHO definition of health. *Hastings Center Studies, 1*(3), 77–87.

World Health Organization. (1948). *Constitution.* Geneva, Switzerland: Author. Retrieved October 30, 2006, from http://www.who.int/about/en.

Yach, D. (1998). *Health and illness: The definition of the World Health Organization*. Retrieved August 9, 2007, from http://www.medizin-ethik.ch/publik/health_illness.htm.

Health Behavior

The major causes of death in the United States and other developed countries are chronic diseases such as heart disease, cancer, and stroke. Behavioral factors are related to all 12 leading causes of death, and behavioral factors are thought to contribute to almost half of the deaths in the United States. According to Mokdad and others, the most common behavioral contributors to mortality, or death, in 2000 were tobacco use, poor diet and physical inactivity, and alcohol use; other significant causes of death include firearms, sexual behavior, motor vehicle crashes, and illicit use of drugs. These behaviors were responsible for nearly 1 million deaths in just a single year. The resurgence of infectious diseases, including foodborne illness and tuberculosis, and the emergence of new infectious diseases such as antibiotic-resistant infections, HIV/AIDS, hepatitis C, and human papillomavirus (HPV) are also largely affected by human behaviors. The social and economic costs related to these behaviors can all be greatly reduced by changes in individuals' behaviors.

Understanding and improving health behavior is key to improving public health and individual well-being. This entry reviews definitions of health behavior, health trends in the United States, and some of the methods used to identify and assess them. It also examines the determinants of health behavior and ways in which such behavior can be improved. In general, behavioral interventions will be more effective if they are adapted to the audiences or communities they are intended for, if they are theoretically based, and if they are carefully crafted and properly pretested. To gain an understanding of health behaviors and to inform the development of behavioral interventions, researchers over the past decade have most frequently made use of the health belief model, the theory of reasoned action and the theory of planned behavior, social-cognitive theory, social ecological models, and the transtheoretical model (or stages of change model). This entry briefly describes these theories and models.

Health Improvement and Public Policy

As chronic disease prevention has grown in importance, so has the role of governments in identifying concerns and goals for health behavior improvement. It was nearly 30 years ago when landmark government-sponsored reports in the United States and Canada called for widespread health improvement through health behavior change. The Health Objectives for the Nation that were published in 1980 stimulated

a dramatic increase in public, private, and professional interest in preventing disability and death through changes in lifestyle behaviors and participation in screening programs. During the same period, data and surveillance systems were established to begin to better track patterns of health behavior and monitor change over time. More recently, Healthy People 2010 was developed and published by the U.S. Department of Health and Human Services to identify public health priorities and specific, measurable objectives. With the overarching goals of increasing the quality and years of healthy life and eliminating health disparities, Healthy People 2010 strongly emphasizes health behavior as central to improving the nation's health. Seven of the 10 top-priority leading health indicators are behavioral: physical activity, overweight and obesity (with roots in eating and activity behaviors), tobacco use, substance abuse, responsible sexual behavior, immunization, and injury and violence.

Definitions of Health Behavior

Health behavior encompasses a large field of study that cuts across various fields, including psychology, education, sociology, public health, epidemiology, and anthropology, and others as well. Several leaders in the field have created working definitions of health behavior. *Health behavior* is defined as observable, overt actions of people that can be reported and measured. However, the field of "health behavior research" often conceives of health behavior very broadly. Gochman defined health behavior as including overt behavior patterns, actions, and habits that relate to prevention, control, and treatment of health problems, as well as the personal beliefs, motives, and emotional states that influence the actual behaviors. A classic typology of health behavior along the health continuum that was articulated by Kasl and Cobb in the 1960s is consistent with the broad health behavior focus. Interestingly, these definitions emphasize the role of the actions of individuals in health behavior. This is in contrast to public health and epidemiological perspectives, which are more often embraced at present and address individuals as part of a larger population or community. Not quite clear on what this means is the point that groups (such as specific communities) can engage in health behaviors, just as individuals can and governmental policies can establish policies that affect health and thus that their policymaking also can be viewed as a health behavior. Health behavior can be something that one does only to oneself, such as putting on sunscreen, or a behavior that affects others, such as putting up a shade cover so that children at the playground are protected from the sun or establishing a smoke-free workplace policy.

Another useful distinction can be made between episodic behaviors, on the one hand, and lifestyle behaviors or habits, on the other. Health behavior can be something that is done once, or periodically, such as getting immunizations or a flu shot. Other health behaviors are actions that are performed over a long period of time, such as eating a healthy (or unhealthy) diet, getting regular physical activity, and avoiding tobacco use. Behaviors that involve a sustained pattern of action over a period of time are usually considered "lifestyle behaviors" or health habits.

Trends in Health Behavior

Although there is more information about what constitutes healthy behavior and risk factors than ever before, this has not always led to people practicing healthier behaviors. There have been some positive changes: In the late 1980s and 1990s, average daily intake of dietary fat dropped from 36% to 34% of total energy, seat belt use increased from 42% to 67%, and the number of women above the age of 40 who had breast exams and mammograms doubled. Tobacco use has declined substantially among adults and seems to have begun to fall among youth. However, not all the news is favorable. More adults and children are overweight than ever before. Diabetes is increasing in near-epidemic proportions. More adolescents are sexually active. One fifth of children below 3 years of age have not received a basic series of vaccinations for polio, measles, diphtheria, and other diseases. Ethnic minorities and those in poverty experience a disproportionate burden of preventable disease and disability, and for many conditions, the gap between disadvantaged and affluent groups is widening.

Surveillance, Monitoring, and Assessment of Health Behavior

Data systems make it possible to track trends in health behaviors and changes in health-related environmental factors and policies in the United States, and in some cases, to link these changes with changes in disease

incidence and mortality. The Behavioral Risk Factor Surveillance System is a state-based telephone survey that is conducted every year, so it is possible to estimate the rates of various health behaviors in a given state and also to compare patterns of behavior between states, regions, and subgroups of the population such as males and females, whites and blacks, and so on. The Youth Risk Behavior Surveillance System (YRBS) is a parallel survey that is conducted in schools to track the health-risk behaviors of adolescents in Grades 9 through 12 in all states. Monitoring the Future (MTF) is a national survey of representative samples of 8th-, 10th-, and 12th-grade students. Other major national health surveys, such as the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES), also include measures of health behaviors as well as self-reported health problems and even physical measures of health status. Specialized surveys, such as the Continuing Survey of Food Intakes of Individuals (CSFII) sponsored by the U.S. Department of Agriculture, provide more detailed population-based data about specific health behaviors such as food consumption and eating patterns. A new survey, the Health Information National Trends Survey (HINTS), was first conducted in 2002 and surveys a national sample of adults every 2 years to assess progress in meeting health and health communication goals including the public's health behaviors.

Monitoring and surveillance are core epidemiologic functions that emphasize tracking and interpreting patterns of behavior and health among large populations. However, other goals of health behavior assessment include understanding health behavior among individuals and groups and evaluating health behavior change programs. For these purposes, self-report surveys of defined audiences are the most common approach. Other methods include self-monitoring through health behavior diaries, short-term recall interviews, and real-time assessment using time-stamped technologies such as personal digital assistants (PDAs), the Internet, and instrumented diaries. The use of observation and biological assessments to validate self-report is also essential to improving the accuracy of these measures.

Determinants of Health Behavior

Many questions about how health behavior develops, is sustained, and changes have not yet been answered. An understanding of the determinants of health behavior using coherent theories is critical for developing effective interventions that may reinforce or change behavior. Because the determinants of health behavior are complex and multifaceted, one single theory rarely can explain a health behavior. Therefore, some models have been developed that use multiple theories to help understand a specific problem in a particular setting or context. Broadly speaking, these theories and models can be broken down into two categories: (1) theories of behavioral prediction or explanation and (2) theories of behavior change or action. Explanatory or predictive theories help identify factors that may influence a health behavior; if properly specified, explanatory theories should then be able to predict reasonably well who will be more or less likely to perform a given behavior. In contrast, theories and models of behavior change focus on the change process; these theories tend to detail stages through which individuals progress before achieving lasting health behavior change. Although these two types of theory often have different emphases, they are complementary. For example, knowing the reasons why someone smokes is important for the development of effective smoking cessation materials, but equally important is an understanding of how someone who has made several unsuccessful attempts to quit in the past can progress to becoming a nonsmoker.

The major theories can be classified into roughly three categories: (1) individual level, focusing on constructs such as knowledge and attitudes; (2) interpersonal level, emphasizing social factors, such as social norms or social support; and (3) structural or environmental, emphasizing multiple levels of influence, including access to resources, laws, and policies. The most commonly used theories cut across these levels, and the most widely used theories have also most often been subjected to testing in scientific research, which is important to advancing our understanding of health behavior.

The *health belief model* was originally developed to explain why people took or did not take advantage of preventive services such as disease screening and immunizations. The model suggests that if a person believes that a health threat has severe consequences (i.e., it could cause death or serious illness), and that he or she is susceptible to developing or contracting a harmful health problem (that it could happen to him or her), then he or she will be motivated to act. To take action against that threat (i.e., change their

behavior), the individual must perceive that the benefits outweigh the costs of the preventive action. For example, if a flu epidemic is predicted for the winter, a person needs to feel that (a) he or she could get the flu, (b) it would be a serious problem, and (c) an action (such as a flu shot) would help enough to take the time and undergo the possible pain of getting the shot.

The *theory of reasoned action* proposes that the most proximal indicator of actual behavior is behavioral intention. Behavioral intentions are a function of (a) attitudes toward the behavior and (b) subjective norms regarding the behavior. An extension of the theory of reasoned action, the *theory of planned behavior* includes the idea of perceived behavioral control. Perceived behavioral control depends on specific beliefs about the likelihood that certain conditions might affect the ability to control the behavior, and whether or not those conditions would encourage or constrain behavioral performance. For example, if a person thinks she can afford a mammogram for early detection of breast cancer, she would be more likely to make an appointment for it than if she doesn't have health insurance coverage.

Social-cognitive theory (SCT) posits that people and their environments interact continuously. A basic premise of SCT is that people learn not only through their experiences but also by watching the way other people act and the results they achieve. According to SCT, three primary factors affect behavior. First, individuals must have self-efficacy or the confidence in their ability to perform the behavior (akin to perceived behavioral control above). Second, individuals must have personal goals that provide them with meaningful incentives for change. Third, the perceived advantages of practicing the behavior must outweigh the obstacles that hinder behavior change.

Social ecological models of behavior change emphasize the importance of the interplay between individuals and their environments. Social ecological models suggest that health behavior determinants are at multiple levels, so that individual, social, and environmental forces all determine health behavior. Ecological models are receiving increasing attention as many behavioral epidemiologists and public health experts recognize that behavior is not simply the result of individuals' knowledge, attitudes, and beliefs.

The *transtheoretical model* (TTM)—often referred to as the stages of change model—addresses individuals' readiness to change their behaviors from unhealthy to healthy ones. Its basic premise is that individuals are at varying levels of "readiness" to change. This means that people at different points in the process of change can benefit from different programs for change. In developing successful behavior change intervention, the programs work best if matched to the person's stage at that time. For example, a smoker who has never thought about quitting will need different messages (to motivate the smoker to think about quitting) than a smoker who has made repeated unsuccessful attempts and may need messages that will build self-efficacy with respect to quitting.

It is important to bear in mind that the various theories of health-related behavior often overlap. Not surprisingly, these explanations for behavior and models for change share several constructs and common issues. Among these are the idea of behavior change as a process, not an event; the distinction between initial behavior changes and long-term maintenance of healthier practices; and the notion that people often weigh the barriers to action against the expected benefits when making decisions about their health behaviors.

While policies, laws, and regulations can affect health behaviors, there are also many individual factors to consider in public health efforts. Behavior change is incremental, and lasting changes are not achieved easily. Public health programs need to identify and maximize the benefits, or advantages, of positive change; push or pull participants along the continuum of change; and consider changes in educational programs and environmental supports to help people who have made positive health behavior changes to maintain them over the long term.

—Karen Glanz

See also Behavioral Risk Factor Surveillance System; Health Communication; Healthy People 2010; Social-Cognitive Theory; Transtheoretical Model

Further Readings

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Glanz, K., & Rimer, B. K. (2005). *Theory at a glance: A guide for health promotion practice* (2nd ed.) [Monograph]. Bethesda, MD: National Cancer Institute (NIH Publication No. 05–3896). Retrieved August 9,

2007 from http://www.cancer.gov/PDF/481f5d53-63df-41bc-bfaf-5aa48ee1da4d/TAAG3.pdf.
Glanz, K., Rimer, B. K., & Lewis, F. M. (2002). *Health behavior and health education: Theory, research and practice* (3rd ed.). San Francisco: Jossey-Bass.
Gochman, D. S. (1982). Labels, systems, and motives: Some perspectives on future research. *Health Education Quarterly, 9*, 167–174.
Kasl, S. V., & Cobb, S. (1966). Health behavior, illness behavior, and sick-role behavior: I. Health and illness behavior. *Archives of Environmental Health, 12*, 246–266.
Mokdad, A. H., Marks, J. S., Stroup, D. F., & Gerberding, J. L. (2004). Actual causes of death in the United States, 2000. *Journal of the American Medical Association, 291*, 1238–1245. (Correction published in *Journal of the American Medical Association*, 2005, *293*, 298)
Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*, 1102–1114.
Smedley, B. D., & Syme, S. L. (Eds.). (2000). *Promoting health: Intervention strategies from social and behavioral research.* Washington, DC: National Academies Press.

Web Sites

Healthy People: http://www.healthypeople.gov.

Health Belief Model

What does it take for people to act to protect themselves from illness? This is the fundamental question posed by the framers of the Health Belief Model (HBM), and it has continued to be addressed by researchers over the past five decades in the disciplines of public health, health psychology, and health education.

Background

The HBM was originally developed by Godfrey Hochbaum, Irwin Rosenstock, and other research psychologists in the U.S. Public Health Service in the early 1950s as they applied cognitive and learning theory to understanding and predicting health behavior. The original work in this area grew out of an attempt to understand the limited utilization of public health programs for disease prevention and screening (including tuberculosis screening). The HBM is a value-expectancy theory that attempts to describe the valuation of the desire to avoid illness (or treat it effectively) and the types of expectations about health that are essential in influencing preventive (or self-care) behavior. The HBM has evolved over the years from addressing primarily health-screening behavior to applications covering the full range of health behaviors from lifestyle change for primary prevention to management of chronic illnesses and sick-role behavior.

Key Concepts of the HBM

The central variables of the HBM have been redefined over time to incorporate a number of concepts beyond those originally considered (perceived susceptibility to the risk and the perceived benefits of early detection, plus a cue to action) to include the following:

Perceived Threat

Perceived threat is a combination of two concepts:

- *Susceptibility.* This is the subjective perception of the individual's risk of developing an illness. In the context of an existing illness, it includes susceptibility to complications of advanced or recurrent disease, acceptance of the diagnosis, as well as more general susceptibility to health problems.
- *Perceived Severity.* Perceived severity is the sense of how serious an illness is and the consequences of leaving it untreated. This concept includes the perception of the possible physical consequences of an illness (e.g., pain, death) and the broader range of social consequences in a person's life (e.g., disability, stigmatization).

Perceived Benefits

Perceived benefits relate to the anticipated positive effects of taking action. This includes beliefs about the effectiveness of a course of action in reducing the disease threat, as well as other potential benefits not directly related to health (e.g., quitting smoking might be seen as a way to save money or set a good example for one's children).

Perceived Barriers

Perceived barriers are the potential negative consequences or costs associated with taking an action to improve health. The factors that could impede a course of action might include concerns about the expense, possible discomfort or danger associated with the

action (e.g., fears about pain or radiation exposure from a mammogram), and inconvenience or competition with other valued activities (e.g., having to miss work to get to an appointment). The wide range of potential barriers include logistical barriers such as cost or lack of convenient access to services, and emotional barriers such as fears about physical or emotional harm (including fear of getting a cancer diagnosis). In addition, when addressing changes in lifestyle and personal habits that may be rewarding in their own right (eating high-fat foods or smoking cigarettes), the habit strength or the loss of pleasurable activities (if not addiction) may prove to be potent barriers to health behavior change.

Cues to Action

Cues to action (either internal cues such as thoughts, emotions, or sensations, or external events that act as a prompt) were one of the initial concepts in the HBM. Interestingly, this component of the model has not been as systematically studied as several others. Nonetheless, examples clearly exist in effective screening and health maintenance interventions that derive from this concept, such as the success of reminder systems for screening tests. Another example is having a cancer diagnosis of a relative motivate people to obtain a first mammogram or colorectal screening test.

Other Modifying Variables

This category includes an array of demographic and sociopsychological variables that may greatly influence the performance of health behavior directly or may interact with the perceptions of susceptibility or seriousness. One powerful example of this variable is a person's level of education, the addition of which has improved the predictive accuracy of the model. The identification of additional social or psychological variables that may be important independent predictors or modifiers of the other variables is an important area for future research, particularly as the HBM is applied to the maintenance of preventive or self-management behaviors in chronic conditions.

Self-Efficacy

This variable was a relatively late addition to the HBM. The concept of self-efficacy, developed by Albert Bandura in 1977, addresses an additional expectancy that influences the performance of a health behavior. Self-efficacy refers to the level of confidence a person feels regarding his or her ability to perform a behavior. Bandura described a number of processes by which a person's sense of self-efficacy may be influenced, and this issue is particularly important when trying to predict or influence the adoption of new behavior patterns or the changing of lifestyle and habits to improve health outcomes. For example, confidence regarding one's skill at being able to test blood sugar and accurately self-administer insulin is essential to the consistent performance of diabetic self-management.

In summary, the HBM posits that adopting a health behavior change requires several beliefs and situations working in concert. First, people must be aware of the health risk and perceive it to be sufficiently serious and likely to affect them to consider taking action. They also need to believe that a particular behavior will be effective in protecting them from a bad outcome in order to overcome whatever possible costs or downside risks they may be concerned about. Moving them toward action may also require the perception of bodily sensations, reminders, or events in their physical or social environment to prompt them to act sooner rather than later. In addition, they need to feel that the behavior change not only will be effective but is something they are capable of doing.

Empirical Evaluation of the HBM

A vast and wide-ranging body of research has been based on the HBM, and reviewing it even briefly is well beyond the scope of this brief entry. Fortunately, that task has been ably accomplished in detailed periodic reviews by Marshall Becker, Nancy Janz, Victor Strecher, Victoria Champion, and others. The reviews show considerable support for the validity and utility of the HBM variables as predictors of health behavior in retrospective, cross-sectional, and prospective studies. To the extent that there are consistent findings across behavioral domains, perceived barriers appears to be the strongest predictor overall. Perceived susceptibility appears to be the next strongest predictor of preventive behavior, whereas perceived benefits is a better predictor of self-care behavior in chronic illness. More recent multivariate modeling has examined the paths by which the variables act in concert to predict health behavior. For example, perceived

severity, which is frequently among the weaker predictors of behavior by itself, may exercise significant influence on behavior through strengthening the importance of perceived benefits.

For all the literature generated by the HBM, relatively little of it has fed back to the development of the model itself. The model offers very little specificity regarding the measurement of the variables and how the variables combine (e.g., additively or multiplicatively, or all at once or in particular sequences). At the same time, the HBM variables have been used as elements of many of the other health behavior models. A recent example of an attempt to develop the HBM is the work of several European investigators, including Anna-Mari Aalto, in proposing an "Expanded Health Belief Model" in which health locus of control, health value, and social support are added as additional "Modifying Factors" and have shown promise in improving prediction of self-management of chronic conditions such as diabetes and asthma.

Finally, an important goal of the HBM is to inform interventions and practice. HBM variables have been essential parts of tailoring interventions to the needs of individual patients and have proven useful in understanding and accommodating cultural differences in diverse populations around the world. However, much remains to be done. With more precise measurement of the variables, a better understanding of the interactions and paths by which the variables exert their influence, and more complete appreciation of the impact of cultural and environmental contexts, more effective interventions can be developed to address new areas of disease prevention and the management of chronic conditions in diverse populations in public health research.

—Lynn Clemow

See also Health Behavior; Prevention: Primary, Secondary, and Tertiary; Screening; Self-Efficacy

Further Readings

Aalto, A. M., & Uutela, A. (1997). Glycemic control, self-care behaviors, and psychosocial factors among insulin treated diabetics: A test of an extended health belief model. *International Journal of Behavioral Medicine, 4*(3), 191–214.

Hochbaum, G. M. (1958). *Public participation in medical screening program: A sociopsychological study* (Public Health Service Publication No. 572). Washington, DC: U.S. Department of Health & Human Services.

Janz, N. K., & Becker, M. H. (1984). The health belief model: A decade later. *Health Education Quarterly, 11*, 1–47.

Janz, N. K., Champion, V. L., & Strecher, V. J. (2002). The health belief model. In K. Glanz, B. K. Rimer, & F. M. Lewis (Eds.), *Health behavior and health education: Theory, research, and practice* (3rd ed., pp. 45–66). San Francisco: Wiley.

Rosenstock, L. M., Strecher, V. J., & Becker, M. (1988). Social learning theory and the health belief model. *Health Education Quarterly, 15*, 175–183.

Healthcare Cost and Utilization Project

The Healthcare Cost and Utilization Project (HCUP) was developed by the Agency for Healthcare Research and Quality (AHRQ), in partnership with entities that maintain statewide hospital administrative databases, to provide multistate, population-based data on both insured and uninsured patients in a uniform format.

Participating states contribute their statewide hospital administrative databases, which contain information from discharge summaries, and HCUP makes the data available for public use. When a patient is discharged from a hospital or a hospital-affiliated facility, an abstract is created that summarizes the administrative information related to the hospitalization. Within individual states, the discharge summaries are incorporated into databases by an agency of the state government, a hospital association, or another organization designated to collect this information. The individual statewide databases contain similar information; however, data completeness and composition vary somewhat from state to state. The Center for Organization and Delivery Studies within AHRQ edits the state databases, applies a uniform coding system, and incorporates the uniformly coded data into the HCUP databases. The data available through HCUP increase as more states participate and as new databases are developed.

Five HCUP databases have been formed, including both inpatient and outpatient administrative data. The statewide files share common data elements, including primary and secondary diagnoses and procedures, admission and discharge status, patient demographics

(age, gender, median income for zip code; race/ethnicity is available for some states), expected payment source, total charges, and length of stay. Some states also include identifiers that enable linkage to other databases, including the AHA Annual Survey of Hospitals, and Medicare public release data.

The central database is the State Inpatient Database (SID), which is composed of annual, state-specific files, beginning with 1990. The 2005 SID file, with 39 states participating, includes about 90% of all discharges from community hospitals in the United States. Under the definition currently used, records from short-term general hospitals and some specialty hospitals are included in the SID; federal hospitals (Veterans Administration, military, and Indian Health Service hospitals), psychiatric hospitals, alcohol/chemical dependence treatment facilities, and hospitals within prisons are excluded.

Two HCUP databases have been developed based on samples drawn from the SID: (1) *The Nationwide Inpatient Sample* (NIS) is designed to approximate a 20% sample of all U.S. community hospitals. The annual database includes all the discharge data from the sampled hospitals (about 1,000 hospitals in 2003). (2) *The Kids' Inpatient Database* (KID) includes 10% of uncomplicated births and 80% of all other pediatric and adolescent hospitalizations. The NIS and KID samples are both drawn from a sampling frame stratified by number of beds, teaching status, ownership, rural/urban location, and region.

In addition to the inpatient databases (SID, NIS, and KID), HCUP has developed two outpatient databases: (1) *The State Ambulatory Surgery Database* (SASD) contains data from hospital-affiliated ambulatory surgery sites; data from some states include freestanding sites as well. These data can be linked to records in the SID. (2) *The State Emergency Department Database* (SEDD) contains data from hospital-affiliated emergency department visits that do not result in hospital admission.

The HCUP Web site provides access to reports and summary analyses and provides software that allows users to query the databases. The databases can be purchased by researchers. Software tools have been developed by AHRQ that can be used on the HCUP databases and with other administrative databases; they can be downloaded without charge. These tools include (1) AHRQ Quality Indicators, three modules for measuring different aspects of the quality of inpatient care, including avoidable hospitalizations and iatrogenic events; (2) Clinical Classifications Software, which aggregates the codes from the International Classification of Diseases (ICD-9-CM) into a smaller number of clinically meaningful categories; and (3) Co-Morbidity Software, which identifies coexisting conditions, using ICD-9-CM diagnosis codes in hospital discharge records.

—Judith Marie Bezy

See also Health Care Services Utilization; International Classification of Diseases; National Center for Health Statistics

Further Readings

Agency for Healthcare Research and Quality. (n.d.). *Healthcare Cost and Utilization Project (HCUP).* Retrieved September 7, 2006, from http://www.hcup-us.ahrq.gov.

Steiner, C., Elixhauser, A., & Schnaier, J. (2002). The Healthcare Cost and Utilization Project: An overview. *Effective Clinical Practice, 5*, 143–151.

Health Care Delivery

Health care delivery encompasses a complex system in which local, state, national, and international communities provide services that enable people to achieve a level of health at which they are able to live socially and economically productive lives. "Health for All by the Year 2000" was adopted as policy by the World Health Organization (WHO) in 1981; however, variable levels of achievement of this mandate have actually been implemented to date. This entry provides a general description of the present health care delivery system in the United States as well as a brief historical perspective designed to provide a framework for the analysis of future trends, values, and needs related to health.

The United States has a very complex system, often called a "multiplicity of health care systems (or subsystems)," and is currently experiencing significant changes. The health care delivery system of today has undergone tremendous change, even over the relatively short period of the past decade. New and emerging technologies, including drugs, devices, procedures, tests, and imaging machinery, have changed patterns of care and sites where care is provided.

Quite realistically, the present system of health care delivery continues in transition, and the next decade promises even more change with a final product that looks very different than common delivery models used today.

In the United States, there are two basic sources of health care services: the private sector and the public sector. Traditionally, the private sector health care refers to arrangements in which an individual client contracts directly with an independent contractor to provide individual care on a fee-for-service basis. In contrast, health care provided through the public sector is usually funded by public taxes and provides health-related services for the protection of all citizens regardless of ability to pay; typically, the service provider bills the government or voluntary agencies. In both settings, the primary care provider could be a physician, physician's assistant, or advanced practice nurse who is trained, is licensed, and operates within professional ethics. Reimbursement is expanding for other types of specialized health care providers such as dentists, physical and massage therapists, mental health professionals, and numerous others.

Health care may be delivered in numerous settings, from inpatient (hospital or extended-care facilities) to outpatient (ambulatory) settings. Ambulatory settings, defined as any setting where the individual is not a bed patient, include hospital-based ambulatory services such as clinics, walk-in and emergency services, hospital-sponsored group practices, and health-promotion centers; freestanding urgent care, same-day surgery, emergency centers, and retail health clinics; health department clinics; neighborhood and community health centers; nursing centers; organized home care; community mental health centers; school and workplace health services; prison health services; and a private clinician's office.

Although the delivery of health care traditionally has been disease oriented, there is an increasing movement toward primary care delivered in a "medical home" model. The National Center of Medical Home Initiatives for Children With Special Needs, a group within the American Academy of Pediatrics, notes on its Web site that "a medical home is not a building, house, or hospital, but rather an approach to providing comprehensive primary care. A medical home is defined as primary care that is accessible, continuous, comprehensive, family-centered, coordinated, and culturally effective." Clinicians in the medical home model coordinate care between various subspecialists and are able to balance conflicting treatment issues. The care is delivered in settings close to where people live and work. The ultimate goal of this effort is to keep people as healthy as possible at a reasonable cost to the payer and to prevent disease.

Ultimately, the people pay for all U.S. health care costs. Money is transferred from consumer to provider by different mechanisms. The major sources are government, private insurance, independent plans, and out-of-pocket support. Frequently, the patient has little knowledge about the total costs incurred for their medical care, because those who pay the bill for health care are primarily the government and private employers. A report from Reuters released in August 2006 documents that simple adherence to basic medical treatment guidelines would save thousands of lives and $1.35 billion a year in medical costs. This basic medical care must include increased delivery of evidence-based clinical preventive services that focus on screening for early signs of disease and risk-reduction efforts.

There are three levels of health care based on the immediate needs of the client. "Stay well" health care services coined by the emerging retail health clinic, also known as the convenient care clinic movement, announce convenient delivery of health screenings, vaccines, and physical exams for basically healthy people. "Get well" services refer to treatment of routine medical conditions or episodic care currently delivered through emergency or urgent care clinics and overlapping with many primary care visits. "Keep well" services speak to chronic disease management to the maximum level possible at all stages of the health care continuum. Each level of services (stay well, get well, and keep well) is seen as a separate yet dynamic and interactive continuum of health care delivery.

Development of the public sector of health care delivery is rooted in the Puritan ethic, inherent in the historical development of the United States, which places a high value on work and assistance for the poor. It includes official and voluntary public health agencies organized at the local, state, federal, and international levels. State health authority is given by the U.S. Constitution, which provides obligation and duty for the government to protect the health and welfare of its citizens. Clearly, the government's role at all levels swings back and forth according to constantly changing political philosophy.

Official agencies are tax supported and therefore accountable to the citizens and the government

through elected or appointed officials or boards and often uses the structure of a health department. A *local* health department's role and functions usually center on providing direct services to the public and depend on the state mandate and community resources. The usual range of services include vital statistics (record of births, deaths, and marriages), laboratory facilities for testing, communicable disease control, environmental health and safety, personal health services usually for special populations, and public health education and information. A *state* health department coordinates health resources within each state and determines eligibility of resources for needy and medically indigent persons. Under broad federal requirements and guidelines, states administer Medicaid (named differently in various states), which is an assistance program that provides payment for medical costs for categories of individuals who are too poor to pay for the care. On the *federal* level, the U.S. Department of Health and Human Services (DHHS), established in 1979, is the main federal body concerned with the health of the nation. The U.S. Public Health Service within the DHHS consists of eight agencies that provide leadership, protect the public, conduct research, and provide treatment. Medicare is the federal insurance program that provides funds for medical costs to seniors and eligible disabled persons. To direct and coordinate *international* health care issues, the WHO was established as a specialized agency of the United Nations in 1948. It assists governments in strengthening health services, furnishing technical assistance, and encouraging and coordinating international scientific research.

In addition, voluntary (or not-for-profit) agencies are powerful forces in the health field at all levels. Stemming from the goodwill and humanitarian concerns of nongovernmental, free-enterprise agendas, they maintain a tax-free status and often have significant impact on issues of concern for health policy and health research. Examples of these voluntary agencies include the American Cancer Society, American Academy of Nurse Practitioners, Susan G. Komen Breast Cancer Foundation, Young and Healthy, Community Wellness Services, Inc., and numerous other health-related organizations, foundations, and professional associations.

Most of the health care systems described above demonstrate modern health policies and practices in the United States, which are guided by Western scientific principles and values. The practice of Western health care is viewed as professional care based on data from the scientifically proven method of research known as quantitative. Some critics of this model argue that it places too much emphasis on the authority, knowledge, and skills of medical professionals and that it encourages consumer dependence and social distance between the producer (doctor) and the consumer (patient).

Another system, known as a folk system of medicine, embodies the beliefs, values, and treatment approaches of a particular cultural group that are a product of cultural development. Folk health practices are delivered in a variety of settings and practiced by a variety of folk healers. These are often unlicensed (at least by the dominant Western health care system) practitioners such as herbalists, bonesetters, lay midwives, spiritualists, scientologists, and astrologers, to simply name a few. Their treatment uses fewer surgical and pharmacological interventions and aims to restore or prevent imbalance between the person and the physical, social, and spiritual worlds.

An emerging and developing field of health care that aims to deliver the best practices from both Western medicine and folk medicine is known as complementary alternative medicine. Clinicians are usually cross-trained in both paradigms and seek licensure and reimbursement privileges from the public and private sector while often operating on a fee-for-service basis. There is a combined use, in different degrees and at different times, of the services and resources from each system. One or more situational factors, including access, perceived degree of severity of the illness and its symptoms, previous experiences with each system, and ability to pay for the services and treatments, may influence which system is approached. This system is promoted as a holistic approach—incorporating family and support systems, consideration of the individual's viewpoint, and caring.

—Eva A. Meyers

See also Complementary and Alternative Medicine; Governmental Role in Public Health; Health Care Services Utilization; Health Economics

Further Readings

Bernstein, A. B., Hing, E., Moss, A. J., Allen, K. F., Siller, A. B., & Tiggle, R. B. (2003). *Health care in America: Trends in utilization*. Hyattsville, MD: U.S. Department of Health and Human Services, National Center for Health Statistics.

Gavan, C. S., & Edelman, C. L. (1998). Health policy and the delivery system. In C. L. Edelman & C. L. Mandle (Eds.), *Health promotion throughout the life span* (4th ed., pp. 49–71). Toronto, Ontario, Canada: Mosby.

Web Sites

National Center of Medical Home Initiatives for Children With Special Needs: http://www.medicalhomeinfo.org/index.html.

U.S. Department of Health and Human Services, Healthy People 2010: http://www.healthypeople.gov.

Health Care Services Utilization

Health care services utilization refers to how much health care people use, the types of health care they use, and the timing of that care. People use health care services for many reasons: to cure illnesses and health conditions, to mend breaks and tears, to prevent or delay future health care problems, to reduce pain and increase quality of life, and sometimes merely to obtain information about their health status and prognosis. Health care utilization can be appropriate or inappropriate, of high or low quality, expensive or inexpensive. It is an evolving process as the population's need for care has changed over time. Factors that influence health care needs include aging, sociodemographic population shifts, and changes in the prevalence and incidence of different diseases. This entry reviews factors that influence health care utilization and reviews the impact of overall trends found in the United States, as reported by the Centers for Disease Control and Prevention (CDC). The study of trends in health care utilization provides important information on health care delivery phenomena and can spotlight areas that may warrant future in-depth studies because of potential disparities in access to, or quality of, care. Trends in utilization may also be used as the basis for projecting future health care needs, forecasting future health care expenditures, or projecting increased personnel training or supply initiatives.

Multiple forces determine the utilization of health services. Some forces encourage more utilization; others deter it. Factors that may *decrease* health services utilization include decreased supply (e.g., hospital closures, physicians retiring), public health/sanitation advances (e.g., quality standards for food and water distribution), better understanding of the risk factors of diseases and prevention initiatives (e.g., smoking prevention programs), discovery/implementation of treatments that cure or eliminate diseases (e.g., polio vaccine), payer pressure to reduce costs, consensus documents or guidelines that recommend decreases in utilization, changes in practice patterns (e.g., encouraging self-care and healthy lifestyles), and changes in consumer preferences (e.g., home birthing, alternative medicine), to name a few.

Factors that may *increase* health services utilization include increased supply (e.g., ambulatory surgery centers, assisted-living residences), a growing population, aging (e.g., prevalence of more chronic illnesses and functional limitations), new procedures and technologies (e.g., hip replacement, MRI), consensus documents or guidelines that recommend increases in utilization (e.g., annual mammograms), new disease entities (e.g., HIV/AIDS, bioterrorism), increased health insurance coverage, changes in practice patterns (e.g., more aggressive treatment of the elderly), and changes in consumer preferences and demand (e.g., cosmetic surgery, direct marketing of drugs), to name a few.

The relationship between any one correlate of utilization and overall health care utilization is not a direct one. For example, the increased length of the aging process can be a result of the postponement of disease onset or a steady rate of functional loss. The increase in the use of some drugs may reduce the prevalence of some other conditions and their associated utilization. Another example would be the increased use of glucose-lowering and blood-pressure-lowering drugs that may reduce complications of diabetes but may also be associated with increased utilization of physicians' services. Therefore, the independent effect of any one factor on health services utilization is not immediately apparent.

One paradigm of health care utilization identifies *predisposing*, *enabling*, and *need* determinants of care. *Predisposing* factors include the propensity to seek care, such as whether an individual's culture accepts the sick role or encourages stoicism, and what types of care are preferred for specific symptoms. *Enabling* factors include depth and breadth of health insurance coverage, whether one can afford copayments or deductibles, whether services are located so that they can be conveniently reached, and other factors that allow one to receive care. *Need* for care also affects

utilization, but need is not always easily determined without expert input. Many people do not know when they need care and what the optimal time to seek care is, and many conditions are not easily diagnosed or treated. If all people could obtain unlimited health care, need as perceived by both patient and provider might be the only determinant of health care utilization. Unfortunately, there are barriers to needed care such as availability or supply of services, ability to pay, or discrimination, which affect overall utilization.

In the United States, there are at least three major payers for health care: governments (federal, state, and local), employers (through employer-based health insurance), and health care consumers themselves (through out-of-pocket payments). In general, services that are covered by insurance and payment programs are more likely to be used than services that must be paid for directly by consumers. Historically, changes in payment policy have also created incentives to provide services differently; for example, availability of the State Children's Health Insurance Program (SCHIP) share the goal of increasing utilization of services by poor children and their families. An example would be when the Federal Breast and Cervical Cancer Care Programs (BCCCP) were funded and became available, the utilization of these important screening services for women were markedly increased. Thus, the benefit and payment structure of Medicare and Medicaid programs, private insurers, and managed care plans tends to strongly influence utilization patterns.

Utilization of services is also affected by availability of services. Health care providers can accommodate only a finite number of patients. The U.S. Department of Health and Human Services reports that over the past decade, in spite of population changes, the overall supply of some types of health care services and providers has remained relatively constant (e.g., hospital beds, emergency rooms, general surgeons, and radiologists), while the supply of many other types of services and providers has increased substantially (e.g., facilities specializing in new technological procedures or tests such as MRI or laser vision repair, adult day care centers, retail health clinics, and nurse practitioners). Procedures that were once performed only on an inpatient basis are increasingly performed in a variety of outpatient and ambulatory care settings.

Consumer-driven health care is a new paradigm that promises to significantly influence the categories that decrease and increase the use of health care services. Consumerism is on the rise, yet consumer empowerment has not yet arrived. The health care market is still transitioning from one in which purchasers and health plans make most decisions on behalf of consumers. And the tools consumers need to make the best choices, based on both cost and quality, are still evolving. Former Speaker of the House Newt Gingrich reports that there is no other sector of our economy with as *little* information about price and quality as in the $2 trillion health care industry. Transparency in cost and quality is available to American consumers in every other business sector, and 93% of Americans believe they have the right to know cost and quality information about their health care providers and services.

The U.S. Department of Health and Human Services Report (2004) documents trends in American health care and reports that racial minorities receive different, often lower-quality medical care than do white Americans. Although some racial, ethnic, and other disparities in care across different population groups have narrowed over time, other major health care utilization disparities remain that are not easily explained by prevalence, incidence, or risk factors. The sources of these differences in care are complex and not immediately apparent, and they may be rooted in historical patterns of the provision of care, perceptions of both providers and care seekers, financial and cultural barriers to care, as well as numerous other factors. This issue is addressed in *Healthy People 2010*, which states that the two goals of the initiative are to (1) increase the quality and years of healthy life and (2) eliminate health disparities. Much more research in this area is forthcoming.

Health care utilization rates are important indicators of what general types of care specific populations seek, and they also indicate how services may be shifting from one site to another. These data may be used by policymakers, planners, researchers, and others in the health community to profile the use of health care services, the epidemiology of health conditions, demand for and patterns of treatment, disparities in treatment, diffusion of new technologies, and changes in patterns of care and the health care system over time.

—Eva A. Meyers

See also Governmental Role in Public Health; Health Care Delivery; Health Disparities; Health Economics

Further Readings

Gingrich, N., & Frogue, J. (2006, March 8). Sticker shock could help with healthcare costs [Electronic version]. *The Hill Newspaper*. Retrieved August 9, 2007 from http://thehill.com/op-eds/sticker-shock-could-help-with-healthcare-costs-2006-03-08.html.

U.S. Department of Health and Human Services. (2000). *Healthy People 2010*. Retrieved November 1, 2006, from http://www.healthypeople.gov.

U.S. Department of Health and Human Services. (2004). *Health care in America: Trends in utilization*. DHHS Pub No. 2004-1031. Hyattsville, MD: Author.

Utilization Review Accreditation Commission. (2006). *Transforming health care: A consumer-centered model takes hold in the information age* (Issue brief). Retrieved August 9, 2007, from http://www.urac.org/savedfiles/URACConsumerIssueBrief.pdf

Health Communication

The concept of communication involves the processes of encoding, transmitting, receiving, and synthesizing information. While there are numerous definitions of *health communication*, the U.S. Department of Health and Human Services, National Institutes of Health, and Centers for Disease Control and Prevention define the concept as "the study and use of communication strategies to inform and influence individual and community decisions that enhance health." The study of health communication has been influenced by the fields of social and clinical psychology, behavior change theory, medical sociology, cultural anthropology, marketing, and of course, communication theory. Health communication is a broad field for research and program development, as many different levels and channels of communication within varying social contexts are examined. Similarly, the field includes numerous methods and areas, such as media literacy, media advocacy, public relations, advertising, education entertainment, risk communication, individual and group instruction, and partnership development, to name a few.

History of Health Communication

The field of health communication has emerged from a number of social science disciplines. Research efforts in psychology and sociology in the 1960s and 1970s, which focused on the health care system, greatly influenced communication scholars. In particular, the psychological literature on persuasion and social influence during this time period, as well as books such as *The Pragmatics of Human Communication* by Watzlawick, Beavin, and Jackson, helped build a theoretical foundation for the field of health communication. Early work in health communication mainly addressed communication within the health care delivery setting, focusing on diagnosis, cooperation, counseling, and education. However, the early landmark study of the Stanford Heart Disease Prevention Program, initiated in the early 1970s, demonstrated that communication can be a powerful influence in health promotion with a wider reach to a larger audience base.

The field of health communication saw a burgeoning increase in published research in the late 1970s and 1980s. As literature concerning the role of communication in health care and health promotion grew, there appeared to be a greater need for academic legitimization of the field. In 1975, the International Communication Association established a Health Communication Division, while the National Communication Association founded a similar subsection in 1985. Within the field of public health, the Public Health Education and Health Promotion section within the American Public Health Association formally recognized health communication as part of its group in 1997.

In the past two decades, two scholarly journals focusing on the field of health communication were launched. *Health Communication*, primarily devoted to communication in health care, was first published in 1989, while the *Journal of Health Communication*, launched in 1996, takes a more international orientation and focuses more on research and public health practice.

Channels and Levels in Health Communication

Health communication inquiry involves a broad array of communication channels, such as face-to-face communication, personal communication (e.g., telephone, mail, fax), mass media (e.g., radio, television, billboards), and newer, interactive technologies (e.g., the Internet, computer kiosks for tailoring information). Health communication research may also span a diverse range of settings varying from homes and schools to workplaces, hospitals, and public spaces, among other venues.

Additionally, health communication analysis looks at communication at numerous levels, including intrapersonal, interpersonal, group, organizational, and societal communication. Intrapersonal health communication focuses on the internal mental and psychological processes that affect health care and health decisions, while interpersonal communication emphasizes the person-to-person influences in health communication, focusing on the patient-provider relationship, provision of health education and therapeutic interaction, and the exchange of relevant information in health care interviews. Group health communication inquiry examines the role communication plays in the interaction of group members, in contexts such as health care teams, support groups, ethics committees, and families. Organizational health communication focuses on the use of communication to coordinate interdependent groups, mobilize different specialists, and share relevant health information within complex health care delivery or health promotion systems to enable effective provision of information on health care and prevention of relevant health risks. Societal health communication addresses the generation, dissemination, and utilization of relevant health information communicated via diverse media to a broad range of audiences to promote changes in attitudes, beliefs, knowledge, and behaviors both in and out of the health care system. While all these communication levels are important, issues related to interpersonal health communication via the patient-provider interaction and societal communication via the mass media are the most discussed in the health communication literature.

Patient-Provider Interaction

Studying interpersonal communication can help researchers understand the impact of the patient-provider relationship, the role of social support in health, and ways in which interpersonal relationships influence health behaviors and decision making. Since much of the early health communication work focused on communication in the health care delivery setting, there is a large evidence base on how the interaction between patients and their providers can influence patient satisfaction, comprehension of health care information, and compliance with medications. Often when physicians and patients interact, they are coming from different worldviews and definitions of health and illness. Physicians learn and internalize a perspective based on the biomedical model of disease and anchored in the world of biochemistry and technology, whereas the patient's world comprises a complex web of personality, culture, living situations, and relationships that can shape and define the illness experience. Research has shown that patient satisfaction with the medical visit is increased when physicians treat patients in a more partner-like manner, when more positively toned words are spoken, when criticisms are given, when more social conversation occurs, and when the physician treats the patient in a warmer and more immediate nonverbal manner, such as engaging in eye contact.

The way of communicating between patients and providers has changed over the years. For the better part of the 20th century, medical paternalism was the norm, with the health care provider seen as the expert who only provided information to the patient when it was deemed necessary. However, for the past two decades, patients have become more like consumers in their orientation, expecting a more egalitarian relationship and increased participation in the health care dialogue.

Mass Media Communication

Probably the most well-known component of the field of societal health communication is the use of mass media as a channel for large health campaigns. Whether through public service announcements (PSAs) or paid media time, these campaigns aim to disseminate persuasive messages to a large target audience. Campaign planners aim not only to increase the amount of information available on a topic but also to redefine or frame an issue as a public health problem to make it salient, attract the attention of the target audience, and suggest a solution to resolve that problem. Some successful health communication campaigns that have used the mass media in the United States include the American Legacy Foundation's truth® youth smoking prevention campaign, Mothers Against Drunk Driving (MADD) advertisements, and the National Cancer Institute's "5 A Day" program to increase fruit and vegetable consumption. With the increasingly cluttered media environment of health messages, researchers have suggested that it is critical for health communication campaigns to be guided by research, build on the current existing knowledge base of the audience, and develop salient messages. It is also critical for campaigns to acknowledge the political climate and structural changes that

might be needed to provide action opportunities for the target audience. For example, before developing a campaign to encourage breast cancer screenings, public health planners need to consider their target audience's health insurance coverage and access to mammograms.

In addition to more traditional campaigns, health messages are disseminated via the mass media in other ways as well. In the United States and internationally, many public health professionals have worked with media representatives to incorporate medically accurate information and behavior change messages into entertainment programming, a strategy referred to as "entertainment education" or "edutainment." Typically, health-related storylines are incorporated into popular entertainment to raise awareness, increase knowledge, create favorable attitudes, and ultimately motivate people to take socially responsible action in their own lives. A few examples include messages on family planning in Mexican *telenovelas* (soap operas), HIV prevention information in the South African soap opera *Soul City*, and domestic violence in the Brazilian show *Mujeres Apasionadas* (Passionate Women). Similarly, many health messages have been addressed in the United States. As early as the 1970s, entertainment television was recognized as a source for delivering important messages to audiences. For example, an episode of "Happy Days" in which the character Fonzie goes to the library to meet girls and ends up getting a library card reportedly inspired thousands of young people to do the same. According to those involved in the effort, the nationwide demand for library cards increased about 500% after the episode aired. Health agencies have made a concerted effort to work with the media to develop successful entertainment-education strategies. One important organization, the Hollywood, Health, and Society program—a partnership of the Centers for Disease Control and Prevention and the University of Southern California's Annenberg Norman Lear Center—works with the creative community to provide shows with medically accurate information for inclusion in scripts and trains screenwriters and producers on health issues.

Health Communication Planning and Research

Implementing a health communication strategy should emerge from careful planning and, ideally, research at multiple stages. For a health communication program to be effective, it must be based on an understanding of the needs and perceptions of the intended audience. Particularly for larger programs and campaigns, there are four significant stages for developing a health communication strategy. The first stage, which permeates all the others, involves planning and strategic development. In this stage, planners identify the intended audiences, use consumer research to craft a communication strategy and objectives, and draft a communication plan. Stage 2 of the process involves developing salient messages for the audience, drafting materials or activities, and conducting formative research such as focus groups to pretest the messages and materials with the intended audience. In the third stage, the health communication program is implemented, and audience exposure and reaction are tracked. The fourth stage, which should be planned at the beginning of the project, involves outcome evaluation and program refinement. During each phase, it is critical to consider getting feedback from the intended audience.

Interactive Technology and Health Communication

Interactive health communication or interactive technology consists of computer-based media that enable users to access information and services of interest, control how the information is presented, and respond to information and messages in the mediated environment. This technology has created new opportunities for patients to receive support, guidance, and health information tailored for their specific needs. Compared with more traditional media, interactive media may have several advantages for health communication efforts, such as improved access to individualized health information, broader choices for users, potential improved anonymity of users, greater access to health information and support on demand, greater ability to promote interaction and social support among users and between consumers and health professionals, and enhanced ability to provide widespread dissemination and immediate updating of content or functions.

Internet Web sites, electronic message boards, and other technology-based resources offer information on an array of topics from diverse sources. While these sources allow patients to access information on demand and gain more control, they also create some potential challenges. The first is that the emergence of new technologies may potentially widen the gap between those who have access to information and

resources and those who do not. Those most in need of services and prevention information may not have the financial means and skills to seek out this information. Additionally, source credibility is an issue. While the Internet allows a greater diversity of opinion and information to be dispersed, users may not be able to assess the accuracy and quality of this information.

Challenges in Health Communication

While health communication has become an integral part of information dissemination and behavior change campaigns, the field has its limitations and challenges. One critique of the health communication field is that it focuses too much on the individual and his or her own behavioral choices and de-emphasizes the upstream causes of poor health among population groups. Additionally, creating the right message for the specific target audience can be challenging. Different populations will not necessarily find the same messages salient; thus, health professionals need to adapt a targeted approach. Also, research and evaluation of health communication programs can be difficult. There are challenges with collecting reliable data, conducting process and outcome evaluations, and assessing impact in the long term. While health communication can play a key role in the public health arena, proponents have suggested that programs may be most effective and broad reaching if communication methods are complemented with other strategies such as advocacy and community mobilization.

—Lisa S. Wolff

See also Health Literacy; Social Marketing; Targeting and Tailoring; Target Population

Further Readings

Finnegan, J. R., & Viswanath, K. (2002). Communication theory and health behavior change: The media studies framework. In K. Glanz, F. M. Lewis, & B. K. Rimer (Eds.), *Health behavior and health education: Theory, research, and practice* (pp. 313–341). San Francisco: Jossey-Bass.

Freimuth, V. S., & Quinn, S. C. (2004). The contributions of health communication to eliminating health disparities. *American Journal of Public Health*, *94*(12), 2053–2055.

Kreps, G. L., Bonaguro, E. W., & Query, J. L. (1998). The history and development of the field of health communication. In L. D. Jackson & B. K. Duffy (Eds.), *Health communication research: Guide to developments and directions* (pp. 1–15). Westport, CT: Greenwood Press.

Maibach, E., & Parrott, L. R. (Eds). (1995). *Designing health messages: Approaches from communication theory to public health practice.* Thousand Oaks, CA: Sage.

Science Panel on Interactive Communication and Health. (1999). *Wired for health and well-being: The emergence of interactive health communication.* Washington, DC: U.S. Department of Health and Human Services, Government Printing Office.

Siegel, M., & Doner, L. (1998). *Marketing public health: Strategies to promote social change.* Gaithersburg, MD: Aspen.

Singal, A., Cody, M. J., Rogers, E. M., & Sabido, M. (Eds). (2004). *Entertainment-education and social change: History, research, and practice.* Mahwah, NJ: Lawrence Erlbaum.

Thomas, R. K. (2005). *Health communication.* New York: Springer.

Thompson, T. L., Dorsey, A. M., Miller, K. I., & Parrott, R. (Eds). (2003). *Handbook of health communication.* Mahwah, NJ: Lawrence Erlbaum.

U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. (2002). *Making health communication programs work: A planner's guide.* Retrieved June 15, 2006, from http://www.cancer.gov/pinkbook.

Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes.* New York: W. W. Norton.

Health Communication in Developing Countries

The term *health communication* refers to the transmission or exchange of information related to health and is often presented with the intention of influencing health-related behavior. This exchange can occur between pairs of individuals, among and between groups and organizations, and via mass media channels. Health communication is an important vehicle for change as many diseases and conditions can be prevented or managed via behavioral modifications, which can be addressed via effective health communication, in conjunction with the development of the medical and public health sectors. Effective health communication campaigns can be housed within a larger development effort, which provides structural change, while increasing the ability of individuals to take advantage of such changes.

Major health risks in developing countries typically involve some combination or all of the following: maternal and childhood malnutrition, risky sexual behavior, sanitation and hygiene issues, lack of clean water, occupational hazards, alcohol and tobacco use, and heart disease risk factors. Targeting these risk factors can lead to reductions in the spread of/improved management of a range of diseases, from tuberculosis to HIV/AIDS or cancer. As with other efforts targeting underserved populations in developed countries, health communication programs in developing countries often face great barriers due to the target audience's lower levels of education, restricted individual agency, immediate concerns that take precedence over health, and resource constraints at the individual and community levels. Accordingly, these communication efforts often include goals linked to economic development and social mobilization. Clearly, the category of "developing countries" is extremely broad and encompasses great diversity across a number of dimensions, including health communication patterns. Thus, this entry provides a broad overview of health communication in this context and provides a discussion of the range of work in this area, as well as areas for future research. Although health communication involves a wide range of information exchanges, this entry will focus on health communication campaigns, as they are a major focus for developing countries.

Health Communication Components

Health communication has informational, instrumental, social control and communal functions. The *informational* function refers to learning from media and other channels, such as learning how to identify safe water sources from a mass media campaign. The *instrumental* function is provision of information that is useful in enabling practical action, such as prompts to access reproductive services at a particular clinic. The *social control* function comes from defining social norms and defining the limits of what is acceptable and unacceptable in health, such as a campaign to change social norms around intimate partner violence. Lastly, the *communal* function can include building social support and access to social capital. These functions are balanced in a given campaign, depending on the overall goals. To develop high-quality health communication campaigns, practitioners should (1) use social marketing tools to gain a deep understanding of the audience and ways to reach/convince them, (2) ground their communications in theory, and (3) use evaluation tools to understand their successes and failures in detail.

Social Marketing

Social marketing refers to use of techniques from commercial marketing to influence the behavior of a defined group, or audience, for the benefit of that group, individuals within it, or society at large. The key is that the movement refers to both internal influences, such as sociodemographic and cultural variables, and external influences, such as access to goods. Key techniques include consumer research, audience segmentation techniques, and an assessment of the marketing mix, referred to as "The Four P's"—product, price, place, and promotion.

Social marketing tools allow practitioners to understand the audience's information needs, preferences, and the best ways in which to reach them. This assessment typically includes an assessment of culture and its impact on the behavior of interest, as well as factoring in the effects of the target audience's health care system and its constraints. Often, information that cannot be acted on must simply be omitted to avoid inducing unnecessary stress and worry in the population. In addition to culture and structural environment, an important consideration is the ways in which the target audience receives (or wishes to receive) health information. Individuals may receive health information incidentally, from interpersonal contacts or from media channels, but also may seek information purposively. By delineating the sources of health information for a given group, health communication specialists can take advantage of social structures (including family, schools, and religious organizations) to spread information or stop the flow of incorrect information. Thus, the channel, or medium, via which messages are sent to the target audience is an important consideration. Channels include television, radio, books and pamphlets, promotional items, and the Internet. Again, the penetration of each of these media varies greatly within and between developing nations and influences choice of channel. For example, *Soul City*, a campaign aimed at raising HIV/AIDS awareness in South Africa, uses television to reach their urban targets and radio to reach rural groups of interest. The program is discussed in detail below.

Grounding in Theory

Commonly used theories and frameworks for the development and evaluation of health communication programming include the PRECEDE-PROCEED model, the theory of reasoned action/theory of planned behavior, social-cognitive theory, the transtheoretical model, organizational change theory, community organization theory, and diffusion of innovations theory. In the development of health messages, health communication specialists may employ strategies of framing, exemplification, narratives, appeals to sensation seeking, or fear appeals.

Evaluating the Campaign

Formal evaluation of the campaign should occur during and after the campaign takes place and focuses on the knowledge, attitudes, and behavior after consumption of mass media products, such as television or radio programming, which are known as *media effects.* The mechanisms by which media affect health behavior are many and may include an individual's increased awareness, sharing of information among interpersonal contacts, and changing awareness among providers or policymakers. Process analysis allows for an understanding of exposure and response to campaign messages and allows managers to redirect the campaign as needed, particularly if a midcourse evaluation suggests that the goals are not being met. At the end of the campaign, another round of evaluation allows practitioners to understand the results of their efforts and also the mechanisms by which these results were achieved, for further refinement and also for dissemination.

Exemplar Programming: Soul City in South Africa

Much attention has been paid recently to the notion of ''edutainment,'' the use of entertainment media to bring about social change through carefully crafted messaging, but there is great debate among health communication professionals regarding the effectiveness of this approach for actual behavior change. Proponents suggest that by tapping into audience interest and excitement, useful health messages can be transmitted to laypeople. However, critics suggest that although interventionists and even participants may feel that an impact was made, long-term increases in awareness and knowledge levels are often undetectable. An oft-cited example of edutainment as a success comes from *Soul City*, a South African organization that has been using mass media channels, including television, radio, and pamphlets, to educate South Africans about HIV/AIDS, with a particular focus on the underserved. The *Soul City* series includes a 13-part television serial drama, which is broadcast in prime time; a radio drama in 45 parts, which is distributed in nine languages; and color pamphlets distributed via national newspapers. These products are the result of deep research and development efforts, as well as community participatory efforts, to support the efficacy of the content and delivery. A recent evaluation showed significant impact on knowledge, attitudes, and behavior at the individual and community levels. Specific target behaviors included prevention of HIV transmission, providing support for those afflicted with AIDS, discussion of HIV/AIDS with personal contacts, and seeking help when needed.

Changing Information and Communication Technologies

Although many health communication efforts in the developing world focus on campaigns via radio and television or through social networks, there has been a recent boom in the attempt to apply information and communication technologies (ICTs) to health information. ICTs in the form of personal digital assistants (PDAs), cell phones, and general Internet access are being used successfully in developing countries to improve the flow of information to and from rural and underdeveloped communities. Initial implementation of ICTs often focuses on provision of generic information, such as availability of services or health information. As the use of various technologies increases, the level of complexity of information increases, and so does the demand for tailored, individualized information. For example, in many countries, individuals or groups are purchasing cellular phones as a cost-effective substitute for land-based telephones. Since text messages are often free, many users prefer to use text messages, and in South Africa, patient reminders regarding appointments are being sent via this medium to improve patient compliance. As access to the Internet improves, individuals in developing countries will be better able to take advantage of this powerful vehicle for information seeking and delivery.

Exemplar Use of ICTs: Village Knowledge Centers in India

Many ICTs are used to increase information flow to rural and underserved populations. For example, in villages near Puducherry in India, Internet access sites, called Village Knowledge Centers, provide villagers with information about agriculture and economics, as well as information about health facilities and programs offered nearby. Two major health-related goals were (1) to improve access to and appropriate use of primary care facilities and government-sponsored health care programs and (2) to empower women to gain more control over their health, particularly with regard to reproductive issues. The project was successful on both counts and demonstrated improved appropriate use of primary care facilities. Additionally, women report taking advantage of the anonymous nature of Internet browsing to gain information about reproductive issues. The beauty of such programming is that the health communication programs have a broad range of effects in terms of economic development, empowerment of marginalized groups, and education.

Challenges in the Field

An important consideration associated with health communication interventions is that the effort may actually widen, rather than reduce, the "knowledge gap" between groups. It is important to ensure that vulnerable groups are targeted specifically to ensure that they are able to benefit from the information provided. This may include ensuring access to women or minority groups and tailoring messages to reach linguistic minorities. This is of particular concern for new technologies, and care must be taken to ensure equitable spread of such advances.

Other major challenges in the field center on the logistics of planning, managing, and evaluating health communication campaigns. At this time, assessment of program impact, particularly for mass media campaigns, is limited. Yet program evaluation is necessary for refining and revising programs to increase their impact. Similarly, by identifying key factors for success, the potential for dissemination to other settings increases dramatically. Lastly, the issue of sustainability should be a focus throughout the programming and evaluation processes. By developing programs that fit within the resource and cultural constraints of a given environment, they will be primed for success. Future research in this area should focus on greater emphasis on evaluation systems to refine programming, integration of communication with the health care service sector, designing health communication programming as part of greater development efforts, improving local capacity to adapt and sustain programs, and crisis/risk communication.

—Shoba Ramanadhan

See also Community Health; Epidemiology in Developing Countries; Health Communication; Health Literacy; Targeting and Tailoring

Further Readings

Haider, M. (Ed.). (2005). *Global public health communication: Challenges, perspectives and strategies*. Sudbury, MA: Jones & Bartlett.

Maibach, E., & Parrott, E. L. (Eds.). (1995). *Designing health messages: Approaches from communication theory and public health practice*. Thousand Oaks, CA: Sage.

Maxfield, A. (2004). *Information and communication technologies for the developing world*. Baltimore, MD: Johns Hopkins Bloomberg School of Public Health.

National Cancer Institute. (1995). *Theory at a glance: A guide for health promotion practice*. Bethesda, MD: National Cancer Institute.

Randolph, W., & Viswanath, K. (2004). Lessons learned from public health mass media campaigns: Marketing health in a crowded media world. *Annual Review of Public Health, 25*, 419–437.

Health Disparities

Health disparities is a broad term that encompasses two categories—health status disparities and health care disparities. Health status disparities refer to differences in health status (i.e., morbidity, mortality, functional status, or disability) among specific populations. In contrast, health care disparities refer to differences in the access, utilization, quality, or outcomes of health care services among specific populations. The distinction between health status and health care disparities may seem pedantic, but in fact this distinction is crucial because the causes and, consequently, the solutions for elimination of health status and health care disparities are likely different. In addition, health

care disparities may be a contributing factor to health status disparities.

Health Status Disparities

The term *health disparities* is widely used in the United States and is similar in meaning to *health inequalities*, which is the term more commonly used in Europe. Most existing definitions of health status disparities describe them as differences in the incidence, prevalence, morbidity, mortality, and survival rates related to diseases and other adverse health conditions that exist among specific population groupings, such as race, ethnicity, culture, sex, socioeconomic status, education, or geopolitical residence. However, not every difference is necessarily a disparity in this sense. There may be differences in the distribution of morbidity and/or mortality between groups that are not indicative of underlying inequality. For example, higher rates of breast cancer among women compared with men or lower rates of skin carcinoma among African Americans compared with whites are due to lower risks of the disease in one group versus another due to biological factors, not social inequalities. In contrast, higher morbidity rates due to breast cancer in African American versus Caucasian women are often cited as a health status disparity because there are no known biological reason for this difference and it is believed to be due in large part to socioeconomic and health care access factors.

Health Care Disparities

Health care disparities are differences in the access, utilization, quality, or outcomes of health care services among specific population groupings, such as those based on race, ethnicity, culture, sex, socioeconomic status, education, or geopolitical residence. In their seminal report, *Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare*, the Institute of Medicine (IOM) limited their definition to race and ethnicity and to the quality dimension of health care disparities, which they defined as ''racial or ethnic differences in the quality of healthcare that are not due to access-related factors or clinical needs, preferences and appropriateness of interventions'' (Institute of Medicine, 2002a, pp. 3–4). The IOM reported that racial and ethnic minorities tend to receive a lower quality of health care than nonminorities, even when access-related factors, such as patients' insurance status and income, were controlled. They found that the sources of these disparities were rooted in historic and contemporary inequities (i.e., stereotypes, biases, high time pressure, cost containment issue, cognitive complexity, financial and institutional arrangements, and barriers of language, geography, and cultural familiarity) and involved many participants at several levels, including health systems, their administrative and bureaucratic processes, utilization managers, health care professionals, and patients. In conclusion, the IOM offered a multilevel strategy to address these disparities.

In a second 2002 IOM report on health disparities (*Guidance for the National Healthcare Disparities Report*), hyperdisparities is defined as ''greater rates of minority utilization of services that are often less desirable or a suboptimal pattern of patient service utilization that extends to access to care'' (Institute of Medicine, 2002b, p. 76). There is a relatively large research literature on hyperdisparities, but this literature typically views hyperdisparities similarly, to disparities. However, there is a subtle, yet important distinction between disparities and hyperdisparities. For example, disparities might refer to racial/ethnic differences (typically underutilization among minority patients) in health care services that prevent or treat disease or disability. In contrast, hyperdisparities would refer to racial/ethnic differences in procedures or services that are associated with poor quality of care (typically overexposure among minority patients). Another way of describing the distinction between disparities and hyperdisparities is that disparities relate to underuse, by minorities, of ''desirable'' health care services such as annual checkups or mammograms, while hyperdisparities refer to overuse of undesirable health care services (such as Emergency Department visits for routine care). Examples of services studied in the hyperdisparities literature include ambulatory care sensitive hospitalizations, limb amputations (typically a result of poorly controlled diabetes), arteriovenostomy (stunts or cannulae implanted for chronic renal dialysis), excisional debridement (usually related to decubitus ulcers), and bilateral orchiectomy (removal of both testes, typically performed for cancer). In an update of a 1996 study, Gornick found higher rates of utilization of each of these procedures for African American Medicare beneficiaries and also found that these hyperdisparities had increased between 1986 and 1996 (see Gornick, 2000). Many in the field of public health have noted large differences

in how disparities are defined in the United States and Europe. As noted above, the term *disparities* is often used in the United States, while the terms *inequality* or *inequity* are used elsewhere. Interpretation of the terms also vary; the United States concentrates race/ethnicity differences in health care, whereas, for example, the United Kingdom focuses on socioeconomic differences in health care.

Another important difference between U.S. and European terminology is the degree to which these terms imply a moral judgment. The U.S. definitions of health disparities concentrate on the quantitative disproportions found in health-related factors among specific populations. However, there is no indication of whether the existence of these disparities is moral or immoral. In contrast, the terms *inequality* and *inequity* incorporate moral stances into their definitions. The term *inequality* is similar in concept to disparities, but puts more emphasis on the notion that the difference is unwarranted and should be corrected. Unlike *disparity*, the term *inequity* makes strong normative assumptions about the differences in health care.

Equity is defined as social justice or fairness; it is an ethical concept grounded in principles of distributive justice. Equity is inherently normative or value laden, while equality is not. Unlike the empirical concept of equality, equity is concerned with the ethical principle of distributive justice at all levels and in all domains. The concept of health equity focuses attention on the distribution of resources and other processes that drive a particular kind of health inequality. Health equity is the absence of systematic disparities in health (or in the major social determinants of health) between groups with different levels of underlying social advantage/disadvantage—that is, wealth, power, and prestige.

—Thomas A. LaVeist, Lydia Isaac, and Rachel Relosa

See also African American Health Issues; Ethics in Health Care; Health Care Delivery; Health Care Services Utilization

Further Readings

Braveman, P., &. Gruskin, S. (2003). Defining equity in health. *Journal of Epidemiology and Community Health, 57*, 254–258.

Carter-Pokras, O., & Banquet, C. (2002). What is a ''health disparity''? *Public Health Reports, 117*(5), 426–434.

Chang, W. C. (2002). The meaning and goals of equity in health. *Journal of Epidemiology and Community Health, 56*(7), 488–491.

Gornick, M. (2000). *Vulnerable populations and medicare services: Why do disparities exist*? New York: Century Foundation Press.

Institute of Medicine. (2002a). *Unequal treatment: Confronting racial and ethnic disparities in healthcare.* Washington, DC: National Academies Press. Retrieved August 9, 2007, from http://www.nap.edu/catalog.php?record_id = 10260#toc.

Institute of Medicine. (2002b). *Guidance for the National Healthcare Disparities Report.* Washington, DC: National Academies Press.

Smedley, B., Stith, A., & Nelson, A. (Eds.). (2002). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Institute of Medicine, Washington, DC: National Academies Press.

Health Economics

Economics is a social science that examines how resources, particularly scarce resources, are produced, distributed, and consumed. Health economics is a field of study that applies the principles of economics to the study of health care. Economic considerations are integral to any discussion of health care policy or expenditure. In addition, principles derived from economic concepts are applied in many aspects of epidemiology. For instance, ''quality-adjusted life years,'' which are fundamental to an understanding of disability epidemiology, use the economic concepts of utility to express numerically the value of a year of life in differing states of health. This entry reviews the basic principles of the market and the important concepts that come into play when analyzing the medical markets and policies in these markets. It also discusses the use of economic modeling in both individual decision making and public policy.

Economics and Policymaking

Understanding what economics can and cannot do is the first and possibly most important step in using economics as a tool of health policy. Economics cannot be used to solve all problems of medical care access and delivery; however, it can offer a framework to study the implications of individual decision making and help define the alternative mechanisms available to improve resource allocation.

Sound policymaking is based on sound economic principles applied in a sensitive and uniform manner. Economics can illuminate human behavior and the way individuals make decisions, respond to incentives, and interact with each other. Economists do not have the final say about the management of the health care systems, but they can make important contributions to the conversation about health policy and how it relates to the health care system.

The Use of Economics in Health Care

Economics is a way to organize thinking about problems that confront people in their daily lives. Thinking like an economist requires a disciplined approach to a problem. Sound reasoning with systematic frameworks is essential. The value of economics stems from its usefulness in making sense out of complex economic and social issues. While economics is one of the social sciences that attempts to explain human behavior, it is unique among the social sciences in that it establishes a context of scarcity and uncertainty. Specifically, economics explains how scarce resources are allocated among competing alternative uses that attempt to satisfy unlimited human wants.

The goal of economic efficiency stems from the fact that there are never enough resources to provide all the goods and services desired by society. Using resources in one way has the trade-off of not being able to use the same resources in a competing activity or alternative. For example, resources applied to the health economy cannot be simultaneously applied to housing or education.

Adopting the goal of economic efficiency implies that the choices made should maximize the total net benefit from all available resources. In the health economy, this involves the evaluation of health care alternatives by calculating the benefits and costs of each alternative and allocating resources in a way that maximizes the net benefits to the population considered.

What Is Health Economics?

Health economists examine a wide range of issues from the nature of health and health care to the market for health and medical care to the microevaluation of health care interventions. Grossman (1972) developed an economic framework for the study of medical care demand where medical care is simply one of the many factors used to produce health. His model of the production of health looks at the determinants of health, including income, wealth, biology, and public health infrastructure and interventions, and lifestyle choices. Many factors confound the ability of medical care to contribute to the production of good health.

The principal activity of economists outside of the United States is the evaluation of medical interventions. Decision makers with limited resources find it necessary to conduct studies comparing the costs and consequences of diagnostic and treatment options to make informed decisions about efficient allocations of scarce resources.

While U.S. health economists are also involved in medical decision making, the primary focus of U.S. health economists is the market for health care. The demand for health care is seen not only as the desire to feel well (i.e., consumption aspects of health) but also as a way for an individual to invest in human capital, because healthy people are more productive than unhealthy people.

Factors affecting the demand for medical care include socioeconomic factors of the population, patient demographics, access barriers, and the role of providers in determining the services to be provided. The supply of health care encompasses a broad spectrum of economics on such topics as production theory, input markets, and industrial organization. Specific issues to be examined are the cost of production, input substitution (e.g., using a generic drug in place of a brand-name drug, when both treat the same condition), and the nature and role of incentives. An example of incentives would be the many "pay-for-performance" programs in which, for instance, hospitals receive bonus payments for meeting certain quality goals.

Analysis of the overall goals and objectives of the health care system is the subject of macroeconomic evaluations. This is where international comparisons are made. For example, how does the U.S. system compare with other countries in terms of cost, access, and quality? Health systems are constantly changing. Policymakers and planners are always looking for better ways to produce delivery and pay for a growing number of medical care services demanded by the public.

Key Economic Concepts

Santerre and Neun (2004) identified terms that are often used by health economists:

• *Scarcity* addresses the problem of limited resources and the need to make choices. Rationing is unavoidable, since not enough resources are available for everyone's needs.

• *Opportunity costs* recognize the role of alternatives. The cost of any decision or choice made is measured in terms of the value placed on the opportunity foregone.

• *Marginal analysis* recognizes that choices are made at the margin, not on an all-or-nothing scope. In this environment, consideration and decision making are based on incremental benefits and costs of an alternative.

• *Self-interest* is the primary motivator of economic actors. People are motivated to pursue efficiently in the production and consumption decisions made.

• *Markets* are places or mechanisms that bring together demanders and suppliers of goods and services. The market accomplishes its tasks through a system of prices, or the invisible hand. The invisible hand can allocate resources because everyone and everything has a price. Prices increase when more is desired and decrease when less is desired. The price mechanism becomes a way to bring a firm's output decision into balance with consumer desires, which is the role of equilibrium.

• *Supply and demand* serve as the foundation of economic analysis. Pricing and output decisions are based on forces underlying these two economic concepts. Rationing using prices comes about when goods and services are allocated in the market based on the consumers' willingness to pay and the suppliers' willingness to provide at a given price.

• *Competition* forces resource owners to use their resources to promote the highest possible satisfaction of society: consumers, producers, and investors. If the resource owners do this well, they are rewarded. If they are inefficient, they are penalized. Competition takes production out of the hands of the less competitive and places it into the hands of the more efficient—constantly promoting the efficient methods of production.

• *Efficiency* measures how well resources are being used to promote social welfare. Inefficient outcomes waste resources, while efficient use of scarce resources promotes social welfare. Social welfare is promoted through the competitive markets through the relatively independent behaviors on the part of thousands of decision makers. Consumers attempt to make themselves better off by allocating limited budgets. Producers maximize profits by using cost-minimizing methods.

• *Market failure* arises when the free market fails to promote efficient use of resources by not producing the optimal level of output. Sources of market failure include natural monopoly, oligopoly, and externalities of production or consumption and public goods. Other market failures can occur through violations of the competitive market, such as incomplete information and immobile resources.

Economic Modeling

One of the main goals of economics is to understand, explain, and predict the actions of economic actors. To do this, it is necessary to simplify behaviors into their elemental parts. Simplification is accomplished through generalization and the construction of models. A model is a way to organize knowledge and explain a particular issue in general terms. An economic model explains how a part of the economy works.

All scientific models start with assumptions. Economic models start by assuming that decisions are made rationally under conditions of scarcity. That is, people's actions are directed toward achieving an objective given constraints. This assumption makes economics different from other social sciences.

In microeconomics, the assumption of rational behavior establishes a consistent framework for individual decision making. It is assumed that individuals must choose between competing alternatives to satisfy certain objectives. Microeconomic models examine the behavior of individual decision makers—individual households and firms and government agents—or specific markets. For example, we use microeconomic models to study how patients' demand for services vary with income or insurance coverage.

Decision making is dominated by the pursuit of self-interest. Individuals use their resources to advance their own economic well-being. When confronted with alterative actions, they choose the one that makes them better off. Decision makers often practice "rational ignorance," meaning that they make choices based on incomplete information because from their perspective, the cost of gathering the remaining information would outweigh its

perceived worth. Decisions must take into consideration foregone opportunities.

Problem Solving

Most microeconomics can be classified under the frameworks of neoclassical economics. This framework is based on optimizing behavior, which is where an economic actor is seeking to achieve given objectives such as profit maximization, cost minimization, or maximization of satisfaction.

Optimization is nothing more than determining the best action given the decision maker's goals and objectives. Constrained optimization takes into account scarcity of resources. For example, how much medical care should a person consume given that his or her health insurance has changed, and are there other goods and services that the person would need to purchase at a given time period?

Choices in the health economy are made at two levels: (1) Individual actors must decide the best course of treatment or services to consume, and (2) policymakers must decide on the best course of action for the entire community. The health economy must consider the following questions: who to treat, when to begin treatment, where to treat, and how much treatment to offer. Of the many ways to go about finding the best alternatives, economic efficiency will be the main criterion in neoclassical models.

The neoclassical model assumes rational behavior on the part of decision makers. Firms maximize profits given technology and the costs of the resources, while consumers maximize utility or satisfaction from consuming various amounts of goods and services, given limited income and the prices of goods and services considered. The labor force supplies workers in order to maximize utility from consuming goods and services and leisure time available subject to the going wage rate. This more or less independent behavior on the part of economic actors leads to equilibrium.

Within this framework, the optimal consumption of goods and services is where the marginal benefit from consumption equals the marginal cost of consumption. Individuals will continue to purchase goods or services as long as marginal benefits exceed marginal costs. Marginal benefits are declining and marginal costs are rising as more of the goods or services are consumed, and the two converge at some quantity. As soon as marginal benefits equal marginal costs, equilibrium is reached and the consumer will consume no more.

From the perspective of economics, it is wasteful to consume all possible medical benefits. Beyond the point of equilibrium, the marginal benefits are not large enough to compensate for the medical risk. The resources used to provide the excess care are better used elsewhere.

—Diane Mary Dewar

See also Economic Evaluation; Health Care Delivery

Further Readings

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy, 80*(2), 223–255.

Henderson, J. W. (2005). *Health economics and policy* (3rd ed.). Mason, OH: Thomson Southwestern.

Santerre, R. E., & Neun, S. P. (2004). *Health economics: Theories, insights and industry studies* (3rd ed.). Mason, OH: Thomson Southwestern.

Health Literacy

The field of inquiry known as health literacy now represents a robust area of research in public health and medicine and is expanding into oral health and mental health. More than 1,000 published studies, multiple annual national conferences, white papers, and reports from prestigious government agencies and academies have put health literacy on the national agenda. For example, Surgeon General Richard Carmona noted, in 2004, that "health literacy is the currency for everything I do." Health literacy is considered a critical issue for consideration in analyses of health disparities and for all health-related communications. This entry focuses on definitions and measures of health literacy, as well as on links between health, health literacy, and other factors such as socioeconomic status and education.

Reports from the U.S. Department of Health and Human Services (DHHS) (2003) and the Institute of Medicine (IOM) use the following definition of health literacy: "The degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions" (U.S. DHHS, 2003, p. 42).

At the same time, however, both the DHHS report *Communicating Health: Priorities and Strategies for*

Progress and the IOM report *Health Literacy: A Prescription to End Confusion* propose an expanded conceptualization of the term so that both the skills of individuals and the demands of health systems be considered. Thus, the Committee on Health Literacy proposes in the IOM report that health literacy "occurs when the expectations, preferences, and skills of individuals seeking health information and services meet the expectations, preferences, and skills of the people providing health information and services" (IOM, 2004, p. 2).

Unfortunately, well over 600 studies that focused on assessments of health materials and published in medical and public health journals have established a clear mismatch between the reading level of health materials and the reading skills of U.S. adults. National and international studies indicate that health systems are becoming increasingly complex in industrialized nations and require more of health care consumers than ever before.

Literacy and Health Outcomes, an analysis commissioned by the Agency for Healthcare Research and Quality (AHRQ, 2004), indicates that research findings have established a link between reading skills of patients and health outcomes. A number of research studies indicate that those with poor reading skills know less about their disease, medicine, or regimen and are less likely to engage in many healthful behaviors. In addition, studies indicated that those with poor reading skills are more likely to be hospitalized, have increased levels of depression and/or other mental health issues, and have increased rates of cervical cancer than do patients with strong reading skills. Furthermore, studies of diabetic patients found that those with poor reading skills are more likely to have diminished anticoagulation control or lower glycemic control. The AHRQ report concludes that the body of research accumulating over the past three decades offers substantive evidence of an association between reading skills of patients and a variety of health outcomes.

Health and Education

Direct links between socioeconomic status and health status are well established. Evidence from accumulated studies indicates that health, morbidity, and mortality are all related to socioeconomic status as measured by income and educational attainment. At the same time, research findings indicate that both income and education independently predict health outcomes. For example, death rates for chronic disease, communicable diseases, and injuries are all inversely related to education. Adults with lower educational achievement are more likely to die of a chronic disease than are adults with higher educational achievement; those with less than a high school education have higher rates of suicide, homicide, cigarette smoking, and heavy alcohol use than do those with higher education. Previous to the year 2000, however, few health researchers examined education alone or its component parts to elucidate the link between education and health outcomes. This was, in part, because education itself was not a major consideration, but was instead viewed as a marker of socioeconomic status. As findings from the first survey of adult literacy, the National Adult Literacy Survey (NALS, 1992), were disseminated among researchers and practitioners in public health and medicine, interest in education and literacy increased.

Measures and Findings

The findings from the 1992 NALS and the 2003 National Assessment of Adult Literacy (NAAL) established indicators of adult literacy in the United States. Close to half of the U.S. adult population have difficulty using print materials with accuracy and consistency to accomplish everyday tasks related to health and safety, finance, civic engagement, and family life—with little if any improvement during the 12-year span. The United States and international measures all focused on functional literacy skills—adults' use of print materials found in everyday life to accomplish mundane tasks. For example, a label on an over-the-counter medicine is used by the purchaser to determine the appropriateness of the medicine for different age groups, the timing of medicine throughout a day, and safe dosage. The NALS and the NAAL measured literacy skills across three dimensions based on types of materials: those in prose format with full sentences and paragraphs; those in document format such as lists, charts, and graphs; and those with numbers and requiring application of the basic mathematical functions of addition, subtraction, multiplication, and division. The complexity of both the materials and the related tasks were considered in the calibration of literacy skills.

In 2003, researchers identified and coded all health-related materials and tasks to estimate the distribution

of literacy for health-related tasks among U.S. adults, describe the health literacy skills of at-risk or vulnerable population groups, and demonstrate how health-related literacy is connected to health status, wealth, and civic engagement. *Literacy and Health in America* indicates that health literacy is strongly related to educational attainment, nativity, minority and immigrant status, access to resources, health status, reading engagement, and civic engagement. The findings show that social factors have a powerful impact both on literacy and on health outcomes.

The U.S. DHHS worked with the National Center for Education Statistics to add 28 health items to the 2003 NAAL so that they might have a baseline measure of health literacy and thereby track changes over time. Findings published in 2006 indicate that a majority of adults do not have the skills needed to use complex health materials with accuracy and consistency (Kutner, Greenberg, Jin, & Paulsen, 2006). Descriptive data similarly indicate that factors such as educational achievement, poverty, minority status, and nativity all influence health literacy scores. Furthermore, those with lower health literacy scores are more likely to indicate fair or poor health than are those with higher health literacy scores.

Implications

Recommendations from the IOM report on health literacy include attention to the development of causal models explaining the relationships among health literacy, the education system, the health system, and relevant social and cultural systems (Recommendation 2-1) and to the development, testing, and use of new measures of health literacy (Recommendation 2-2). In addition, the IOM recommends that health literacy measures be developed for large ongoing population surveys. In the health field, these include the Medical Expenditure Panel Survey, the Behavioral Risk Factor Surveillance System, and the Medicare Beneficiaries Survey, as well as for accreditation and quality assessment activities, such as those carried out by the Joint Commission on Accreditation of Healthcare Organizations and the National Committee for Quality Assurance (Recommendation 2-2). Recommendation 6-2 states that health literacy assessment should be part of the health care information systems and quality data collection. Many recommended follow-up activities are already under way. Thus, as health literacy enters the health research agenda, it may well prove to be a critical variable for explorations of pathways from education to health outcomes and for analyses of health disparities.

However, researchers have yet to move beyond assessments of health materials such as informed consent documents, patient education materials, and procedural and preparatory texts to examine and measure the literacy-related barriers to participation in health studies and trials. For example, researchers have not yet systematically studied the literacy-related demands placed on participants in research studies; nor have they systematically examined and assessed the complexity and reading level of open-ended documents and, specifically, questionnaires in current use. Health literacy concerns have consequences for health research.

—Rima E. Rudd, Jennie E. Anderson, and Lindsay Rosenfeld

See also Health Behavior; Health Communication; Health Disparities

Further Readings

Berkman, N. D., Dewalt, D. A., Pignone, M. P., Sheridan, S. L., Lohr, K. N., Lux, L., et al. (2004). Literacy and health outcomes. *Evidence report/ technology assessment (Summary), 87*, 1–8.

Institute of Medicine. (2004). *Health literacy: A prescription to end confusion* (pp. 1–18). Washington, DC: National Academies of Science.

Kirsch, I., Jungeblut, A., Jenkins, L., & Kolstad, A. (Eds.). (1993). *Adult literacy in America: A first look at the National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). *The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy* (NCES 2006-483). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

National Center for Education Statistics. (2005). *National Assessment of Adult Literacy.* Retrieved April 4, 2006, from http://www.nces.ed.gov/naal.

National Center for Education Statistics. (2006). *The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy.* Retrieved November 2, 2006, from http://www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid = 2006483.

Rudd, R. E., Anderson, J., Nath, C., & Oppenheimer, S. (2007). Health literacy: An update of medical and public health literature. In J. P. Comings, B. Garner, & C. Smith (Eds.), *Review of adult learning and literacy* (Vol. 7). Mahwah, NJ: Lawrence Erlbaum.

Rudd, R. E., Kirsch, I. S., & Yamamoto, K. (2004). *Literacy and health in America.* Princeton, NJ: Educational Testing Service.

Rudd, R. E., Moeykens, B. A., & Colton, T. C. (2000). Health and literacy: A review of medical and public health literature. In J. P. Comings, B. Garner, & C. Smith (Eds.), *The annual review of adult learning and literacy* (pp. 158–199). San Francisco: Jossey-Bass.

U.S. Department of Health and Human Services. (2003). *Communicating health: Priorities and strategies for progress.* Washington, DC: Government Printing Office.

Web Sites

Harvard University School of Public Health, Health Literacy Studies: http://www.hsph.harvard.edu/healthliteracy.

National Assessment of Adult Literacy: http://www.nces.ed.gov/naal.

Health Plan Employer Data and Information Set

The Health Plan Employer Data and Information Set (HEDIS) is a set of performance measures for managed care health insurance plans produced by the National Committee for Quality Assurance (NCQA), an independent, not-for-profit association founded in 1990. HEDIS measures were developed to facilitate comparison of the performance of managed care plans to each other and to national benchmarks. They have also been used to study trends in managed care over time.

Four types of managed care plans are included in the HEDIS measures. Health maintenance organizations (HMOs) offer a range of benefits for a set monthly fee; managed behavioral health care organizations (MBHOs) are similar to HMOs but provide care for mental health and substance abuse disorders; point of service (POS) and preferred provider organizations (PPO) plans provide free or highly subsidized care within a specified network of health care providers and a lesser subsidy for care provided by doctors outside that network.

NCQA evaluates managed care plans in five dimensions. *Access and service* evaluates the quality of customer service and access to care provided by plans, including availability of sufficient primary care physicians and specialists and consumer-reported difficulties in getting care. *Qualified providers* evaluates the training and licensure of the physicians within the plan, sanctions and lawsuits filed against them, and consumer satisfaction with the plan's physicians. *Staying healthy* evaluates the quality of preventive care provided by the plans, including appropriate use of tests and screening procedures, and plan guidelines to physicians concerning preventive care. *Getting better* reviews managed care plan activities intended to help people recover from illness, including access to the most up-to-date care and provision of health behavior programs such as smoking cessation. *Living with illness* evaluates plan activities related to the management of chronic illnesses such as diabetes and asthma.

NCQA also grants or denies accreditation to managed care plans, with several levels of grading. For HMOs and POS plans, the highest level of accreditation is *excellent*, which is granted to plans that meet or exceed HEDIS requirements for clinical quality and service, and are also in the highest range of regional or national performance. The next level is *commendable*, which is granted to those that meet or exceed HEDIS requirements for clinical quality and service. *Accredited* signifies that a plan met most of the HEDIS basic requirements, and *denied* indicates that the plan did not meet these requirements. For PPOs, the highest level of accreditation is *full*, which is comparable with *excellent* for HMOs and is granted for a 3-year period. PPOs that meet most but not all standards may be granted *1-year* accreditation that is reviewed after a year to see if the plan qualifies for *full* accreditation.

Participation in HEDIS is voluntary, but more than 90% of managed care plans in the United States participated in 2006. Data used in HEDIS evaluations are collected by each participating managed care plan and analyzed by NCQA; plans may elect to have the HEDIS results verified by an independent auditor. HEDIS data and reports are available for purchase through the HEDIS Web site, and brief information about individual plans, including scores on the five dimensions and overall accreditation status, is available through the HEDIS Web site.

—Sarah Boslaugh

See also Health Care Delivery; Health Care Services Utilization; Managed Care; Medicaid; Medicare

Web Sites

Health Plan Employer Data and Information Set: http://www.ncqa.org/programs/hedis.

Healthy People 2010

Healthy People 2010 is the current national health promotion and disease prevention agenda published by the U.S. Department of Health and Human Services. It was developed with leadership from the Secretary of Health and Human Services Council on National Health Promotion and Disease Prevention with the input of the Healthy People Consortium, an alliance of government agencies and public and private health agencies and national membership organizations.

The Healthy People 2010 framework includes 2 overarching goals, 28 focus areas, and 467 specific objectives. This identification of an extensive range of health priorities and quantifiable objectives gives states, businesses, educational institutions, and health care providers baseline data and a structure to create and evaluate public health programs or state-specific health agendas.

History

In 1979, the landmark report *Healthy People: The Surgeon General's Report on Health Promotion and Disease Prevention* provided Americans with the first set of national public health goals that focused on reducing premature deaths and preserving independence for the aging population. Following the 1979 Surgeon General's report, *Promoting Health/Preventing Disease: Objectives for the Nation* was released in 1980. That report identified 226 health objective goals for the nation to achieve during that decade. The Department of Health and Human Services has created a new set of goals and public health priorities every decade since then. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives* was released in 1990, and the current *Healthy People 2010: Healthy People in Healthy Communities* was released in 2000.

To help coordinate public health activates at the national, state, and local levels, Healthy People 2010 provides a comprehensive portrait of the nation's health in 2000, sets national goals for 2010, and releases a midcourse review to monitor progress. By bridging the link between community and individual health, the report solidifies the notion of collective action for both community and individual health improvement. Healthy People 2010 also creates an agenda for program funding, prioritizes research, and provides guidance for new regulatory efforts carried out by the various agencies under the Department of Health and Human Services.

Goals

As our nation's demographic profile shifts to an older and more diverse population, so do our health priorities. The two overarching goals of Healthy People 2010 are poised to address our changing demographics.

The first goal, to increase quality and years of healthy life, attempts to tackle our aging population; the second goal, to eliminate health disparities, aims at closing the health chasm that exists between various ethnic and racial groups.

Focus Areas

Leading Health Indicators

Chosen based on the availability of current data to serve as baselines and their relevance to wide-ranging public health issues, Leading Health Indicators were included for the first time in Healthy People 2010. The indicators serve as a tool for monitoring national progress in the following health areas: physical activity, overweight and obesity, tobacco use, substance abuse, sexual behavior, mental health, injury and violence, environmental quality, immunization, and access to health care.

Objectives

Healthy People 2010 objectives provide the vision and direction for action on a specific health outcome or health status. By separating objectives into two categories, measurable objectives and developmental objectives, a clearer structure emerges. Measurable objectives are formulated to drive specific health promotion actions. They provide baseline data, derived from nationally representative data systems, from which the target is set. As a complement to measurable objectives, developmental objectives provide a vision for a specific health outcome or health status. By identifying areas of increasing importance, the developmental objectives seek to drive new research and data collection systems.

DATA2010

To help health professionals measure progress toward Health People 2010 objectives, the National Center

for Health Care Statistics created an interactive Web-based database system, DATA2010. This system allows users to create tables with baseline data for each objective and for the Leading Health Indicators. A single data point tracks most objectives; progress toward goals is based on the change from the baseline value to the 2010 target. DATA2010 data are gathered from census records, vital statistics, the National Notifiable Disease Surveillance System, hospital discharge databases, and large national health surveys, such as the National Health Interview Survey and Youth Risk Behavior Surveillance System.

—Elizabeth Serraillier

See also Governmental Role in Public Health; Health Disparities; National Center for Health Statistics; Race and Ethnicity, Measurement Issues with; Surgeon General, U.S.

Further Readings

Allison, J. (1999). Can data-driven benchmarks be used to set the goals of Healthy People 2010? *American Journal of Public Health, 89*, 61–65.

Garland, L. (2002). Measuring 2010 national objectives and leading indicators at the state and local level. *Journal of Public Health Management & Practice, 8*(4), 9–13.

National Center for Health Care Statistics. (2007). *DATA2010*. Retrieved August 9, 2007, from http://wonder.cdc.gov/data2010/about.htm.

Sondick, E. J., Lucas, J. W., & Smith, S. S. (2000). Race/ethnicity and the 2000 census: Implications for public health. *American Journal of Public Health, 90*, 1709–1713.

U.S. Department of Health and Human Services. (2000, November). *Healthy people 2010* (2nd ed.). With Understanding and Improving Health and Objectives for Improving Health (Vols. 1 and 2). Washington, DC: Government Printing Office.

Healthy Worker Effect

The term *healthy worker effect* (HWE) refers to the fact that employed persons are generally healthier than persons who are not employed. For this reason, studies based on samples drawn from employed people may produce results that are biased due to the selection effect of employment. This is particularly important in cohort studies involving occupational exposures, when studies comparing working cohorts with the general population may not find effects on morbidity and mortality attributable to hazardous occupational exposures due to the healthy worker effect. Study designs that do not account for the selection of healthier workers at the time of employment as well as the removal of sick workers during employment can result in masked health effects and may cause harmful exposures to appear healthful. Selection of the appropriate comparison group is essential in alleviating or eliminating the HWE in occupational studies.

Definition

HWE most commonly refers to the situation in which workers exposed to occupational hazards exhibit lower mortality rates than a reference population without the exposure of interest. Typically affecting occupational studies, the HWE arises when cohorts of workers are compared with reference groups that include the severely sick, terminally ill, and others incapable of working. The strength of the effect varies by occupation and may partially or completely mask any negative exposure-outcome association. In some circumstances, an exposure with negative health effects may appear to confer protection to the exposed.

History

The HWE was discovered in 1885 by William Ogle, who noticed that workers in more demanding occupations had lower mortality rates than those who were employed in less vigorous work or who were unemployed. More than 100 years later, in 1974, the term *healthy worker effect* was coined by McMichael. A few years later, in 1976, Fox and Collier first quantified the HWE by calculating the standardized mortality ratio using the general population as a reference.

Healthy Worker Effect Modifiers

Diverse factors, including personal characteristics (e.g., race, gender), study design (e.g., completeness and length of follow-up), worker screening (e.g., health/ability to perform at hire, age of hire), and disease traits (e.g., incubation period, clinical presentation), may affect the strength of the HWE. The duration of employment, age of hire, and presence of overt symptoms are examples of factors that may cause the study population to mistakenly appear

healthier than the reference population and that therefore affect the HWE. Workers who have participated in the workforce for a longer duration are less likely to have terminated work prior to standard retirement age due to illness. Those with a higher age of hire are more highly influenced by selective processes and therefore experience a stronger healthy worker effect than workers hired at a younger age; those who are hired at an older age are likely to be in better health than similarly aged individuals in the general population. HWE is also increased in studies involving diseases with overt symptoms that interfere with work or that compel individuals to leave the workforce prior to retirement.

Reducing the Healthy Worker Effect

The choice of comparison group is critical in reducing the HWE, and various suggestions have been proposed. One option is to identify an external comparison group from an industry that lacks the hazard being studied but that applies similar worker screening processes. Another option is to choose an internal reference group from within the occupation. In an internal comparison, rates among those with high exposures are compared with those with low or no exposures. Ensuring complete follow-up of all individuals regardless of subsequent worker status will minimize bias, although complete follow-up cannot completely remove bias introduced by the initial selection of healthy workers into employment.

—Michelle Kirian

See also Bias; Confounding; Environmental and Occupational Epidemiology; Study Design

Further Readings

Baillargeon, J., Wilkinson, G., Rudkin, L., Baillargeon, G., & Ray, L. (1998). Characteristics of the healthy worker effect: A comparison of male and female occupational cohorts. *Journal of Occupational and Environmental Medicine, 40*(4), 368–373.

Li, C. Y., & Sung, F. C. (1999). A review of the healthy worker effect in occupational epidemiology. *Occupational Medicine, 49*(4), 225–229.

Richardson, D., Wing, S., Steenland, K., & McKelvey, W. (2004). Time-related aspects of the healthy worker survival effect. *Annals of Epidemiology, 14*(9), 633–639.

Hepatitis

Hepatitis is a term that describes a number of conditions, syndromes, or diseases leading to inflammation of the liver. The incidence of specific types of hepatitis is endemic in some countries, epidemic in others, and on the rise in still other locales. Hepatitis appears in different guises; it may mimic influenza, it may present only as an acute disease, or it may become chronic in nature and in some cases fatal. There are a variety of causes of hepatitis, including parasites, bacteria, viruses, alcohol ingestion, poisonous mushroom ingestion, overdose of certain medications, and autoimmune response. Students and professionals working in public health should understand the complexities of this condition so that whenever possible it can be diagnosed promptly, necessary treatment initiated, epidemics thwarted, and preventive methods instituted whenever possible.

The major types of hepatitis are known by the name of the viruses that cause each specific type of hepatitis: A, B, C, D, and E. The other currently known types of hepatitis are drug-induced hepatitis, alcoholic hepatitis, and autoimmune or lupoid hepatitis. This entry discusses these diseases in terms of causation, diagnostic tests, risk behaviors, symptoms, and prevention.

Symptoms

Because the initial symptoms of viral hepatitis are similar to influenza, the initial diagnosis may be that of the flu. The following symptoms are most commonly seen in individuals who have one of the conditions known as viral hepatitis:

- Fatigue or malaise
- Loss of appetite
- Nausea
- Vomiting
- Low-grade fever
- Headache
- Diarrhea
- Abdominal pain
- Jaundice (sclera and/or skin)
- Generalized skin itching
- Clay-colored stools
- Weight loss

Diagnosis

The diagnosis of any of the viral hepatitis conditions generally consists of several blood tests and may

include a complete blood count (CBC) and a battery of liver function tests (LFTs) including the following:

- Serum albumin
- Alkaline phosphatase (ALP)
- Alanine transaminase (ALT); also known as serum pyruvate transaminase (SGPT)
- Aspartate aminotransferase (AST); also known as serum glutamic-oxaloacetic transaminase (SGOT)
- Gamma-GT (GGTP or GGT)
- Prothrombin time (PT or pro-time)
- Total bilirubin; direct and indirect bilirubin; also known as conjugated or unconjugated bilirubin
- Urine bilirubin

The three most common virally caused hepatitis conditions are hepatitis A (HepA), hepatitis B (HepB), and hepatitis C (HepC). In diagnosing viral hepatitis, serum will also be collected to determine the presence of antibodies for each of these three types of hepatitis. IgM antibodies generally appear in the serum approximately 3 to 4 weeks after exposure and return to normal levels within approximately 8 weeks. IgG antibodies generally appear in the serum approximately 2 weeks after the IgM antibodies increase and may be perpetually present. A normal test for any of these indicates lack of viral exposure.

Causation and Risk Factors

Each of the three types of hepatitis is caused by a specific virus. In addition, mode of transmission and associated risk factors may be disease specific. HepA is transmitted through contaminated water or food and contact with the stool, blood, or other body secretions of an infected person during the 15- to 45-day incubation period before symptoms are present and during the first 7 days of illness. HepA is considered an acute disease; it does not become chronic and is rarely fatal. The risk factors include communal living, for example, nursing homes, close contact with a recently infected person, intravenous drug use, and travel or recent immigration from endemic areas such as South or Central America or many Asian countries.

HepB is transmitted through blood and other body fluids. It is considered an acute disease but may become chronic depending on the age of the person at the time of infection. The symptoms appear within 1 to 6 months and may include muscle and joint aches. The risk factors are any behaviors that involve exposure to blood or other body fluids, including health care work that may cause an individual to come into contact with blood, unsafe sex with an infected person, intravenous drug use, blood transfusions, and tattooing or acupuncture with contaminated instruments. It is also possible for an infected mother to transmit the virus to her baby during the birth process.

HepC or non-A non-B hepatitis is transmitted through contact with blood, blood products, or solid body organs. Risk factors include sharing intravenous needles, razors, or toothbrushes with infected individuals. It is also possible for HepC to be transmitted through sex with an infected partner or during birth if the mother is infected. Additionally, long-term dialysis recipients or those who have workplace contact with blood may transmit it.

Hepatitis D or Delta agent is caused by a defective viral agent and is seen only in individuals with HepB; therefore, it is found primarily in intravenous drug users, recipients of multiple blood transfusions, and those previously infected with HepB or who are carriers of it.

Hepatitis E is the type of viral hepatitis about which the least is currently known; however, the known outbreaks have been associated with fecally contaminated water rather than person-to-person contact.

Prevention

Prevention of hepatitis is primarily human-behavior driven. Vaccines are available for HepA and HepB and should be used in high-risk populations. Preventive measures for the other viral hepatitis strains include not participating in unsafe sex, not sharing intravenous needles, using contamination-free water and water products such as ice for all human purposes, properly cooking all foods in endemic areas, using appropriate personal hygiene techniques, and not sharing an infected person's personal items such as razors or toothbrushes. Additionally, those who produce or handle food should be educated to refrain from using contaminated water in or around food production.

Other types of hepatitis may be caused by drugs such as acetaminophen, alcohol abuse, or as an autoimmune response. All treatment for hepatitis is generally supportive in nature, ensuring adequate nutrition, hydration, and rest.

—Donna Scemons

See also Bloodborne Diseases; Sexually Transmitted Diseases; Waterborne Diseases

Web Sites

American Liver Foundation: http://www.liverfoundation.org.
Centers for Disease Control and Prevention, Viral Hepatitis: http://www.cdc.gov/hepatitis.
Hepatitis Foundation International: http://www.hepatitisfoundation.org.

Herd Immunity

Herd immunity describes a state in which an infectious disease transmissible through human contact is unlikely to spread because a large proportion of the population is immune to the disease. This immunity can be conferred through natural immunity, previous exposure to the disease, or vaccination. The entire population does not need to be immune to attain herd immunity. Rather, the population density of persons susceptible to infection must be low enough to minimize the likelihood of an infected individual coming in contact with a susceptible individual. Herd immunity can prevent sustained disease spread in populations, thereby protecting susceptible individuals from infection. It is important to emphasize that the concept of herd immunity is applicable only to infectious diseases that can be spread by human contact. For example, herd immunity is unlikely to protect unvaccinated persons from tetanus, due to the ubiquitous presence of *Clostridium tetani* in natural reservoirs of soil and animal droppings.

The percentage of the population that must be immune to produce herd immunity differs for each infectious disease. A disease with a high infectivity rate such as measles will require a higher proportion of immune persons to achieve herd immunity than a disease with lower infectivity such as tuberculosis. In addition, individual- and population-level characteristics influencing disease spread such as susceptibility, demographics, social habits, and clustering affect herd immunity.

Herd immunity is an important consideration for mass vaccination practices. Even if a cheap, safe, and effective vaccine exists, resource, logistical, and societal constraints prevent the vaccination of 100% of the population. A reasonable target level of vaccination may be to achieve the threshold level of herd immunity H, which is calculated as

$$H > 1 - \frac{1}{R_0},$$

where R_0 is the basic reproductive rate, the number of infections an infected individual can be expected to produce on entry into a susceptible population. Mass vaccination can be successful through principles of herd immunity, although disease outbreaks can still occur, although generally to a lesser extent than if herd immunity had not been achieved.

—Brian K. Lee

See also Disease Eradication; Vaccination

Further Readings

Anderson, R. M., & May, R. M. (1985). Vaccination and herd immunity to infectious diseases. *Nature*, *318*(6044), 323–329.
Fox, J. P., Elveback, L., Scott, W., Gatewood, L., & Ackerman, E. (1995). Herd immunity: Basic concept and relevance to public health immunization practices. *American Journal of Epidemiology, 141*(3), 187–197; discussion 185–186.

Heritability

In genetics, heritability is the amount of phenotypic variation in a population that is attributable to individual genetic differences. Heritability, in a broad, general sense, is the ratio of variation due to differences among genotypes to the total phenotypic variation for a character or trait in a population. It is expressed as

$$H = \frac{\text{Genotypic variation}}{\text{Total variation}}.$$

The range of values for heritability estimates is 0 to 1. If $H = 1$, then we are able to say that all variation in a population is due to differences or variation among genotypes (i.e., there is no environmentally caused variation). On the other hand, if $H = 0$, there is no genetic variation. In this case, all variation in the population is from differences in environments during the life experience of individuals. In other words, we can say that all individuals are the same with regard to the effect of their genes on phenotypic variance.

The following example, taken from the volume by Bodmer and Cavalli-Sforza (1976), demonstrates how the concept of heritability can be used to measure the relative contribution of genetic and environmental factors in patterning phenotypic variation.

In the example, we have two genotypes, G1 and G2, and two environments, E1 and E2. Table 1 shows four individuals, two of each genotype, distributed in the two environments, and their phenotype scores for some measurable character or trait.

If we graph these data, they look as shown in Figure 1.

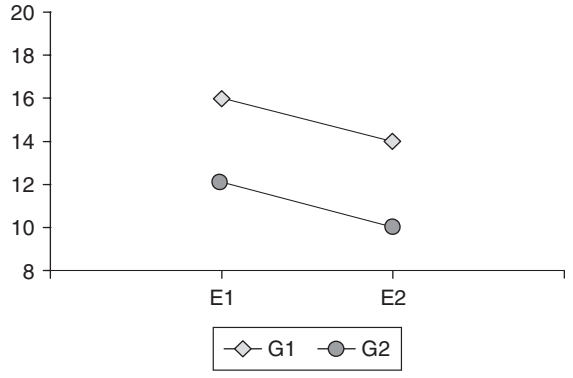


Figure 1 Phenotype Scores Plotted by Genotype and Environment

Source: Adapted from Bodmer and Cavalli-Sforza (1976).

One should note that there is variation, that is, there are differences, among the four individuals on their phenotype scores. Note also that the scores of G1 are higher on average and E1 seems to produce higher phenotype scores, on average, than E2. If we calculate a mean for all four individuals, we get $52/4 = 13 =$ the mean phenotype score. How would we now express the variation among the four individuals?

We calculate a measure of variation called the *variance*. The steps are as follows:

1. First, calculate the deviations of each of the individual scores from the mean.
2. Then square these deviations.
3. Then add up these squared deviations to get the sum of squares or *SS*, which is $SS = (16-13)^2 + (14-13)^2 + (12-13)^2 + (10-13)^2 = 9 + 1 + 1 + 9 = 20$.

This *SS* value is not the variance estimate, but if we divide by $4 - 1$ or 3, we have the actual variance, that is, the value for the total variation among phenotypes.

Once we have calculated total phenotypic variance, we can then estimate how this variance partitions into genetic and environmental components. Only the genetic component will be shown here. We obtain the estimate of the genetic variance component by eliminating the environmental source of variation, that is, placing genotypes in either E1 or in E2. Table 2 shows the four individuals placed in environment E1.

If we calculate a mean for the phenotype scores, the value is 14. The $SS = 16$. These squared deviations are totally the result of variation among genotypes. We can label this second *SS* as SS_g to denote the genetic *SS*. It should be noted that if we had used E2, we would have obtained the same value of 16 for the SS_g. Now, using our *SS* values, we can calculate the actual broad heritability estimate. It would be V_g/V_t or 16/3 divided by $20/3 = 0.80$. In this constructed example, we compute that 80% of the total variance is the result of variance among the genotypes. An important additional point is that this heritability would be valid only under the set of conditions we have shown.

Heritability is, in practice, a statistic and is estimated using equations based on the study design in question. There are essentially two approaches to estimating heritability, one based on correlation and

Table 1 Phenotype Scores for Four Individuals by Genotype and Environment

Individual	*Genotype*	*Environment*	*Phenotype Score*
1	G1	E1	16
2	G1	E2	14
3	G2	E1	12
4	G2	E2	10

Source: Adapted from Bodmer and Cavalli-Sforza (1976).

Table 2 Phenotype Scores for Four Individuals of Two Genotypes in One Environment (E1)

Individual	*Genotype*	*Environment*	*Phenotype Score*
1	G1	E1	16
2	G1	E1	16
3	G2	E1	12
4	G2	E1	12

Source: Adapted from Bodmer and Cavalli-Sforza (1976).

regression methods and the other on analysis of variance methods as we have demonstrated in the example. One study design commonly used in human genetics is twin studies. The methodology is based on the fact that identical twins (monozygotic or one-egg twins) share 100% of their genes in common and nonidentical or fraternal twins (dizygotic or two-egg twins) are like other siblings such as brothers and sisters in that they share 50% of their genes in common. We expect the correlation between identical twins to be equal to 1.0 and that of fraternal twins to be 0.50. The estimate of heritability based on this approach approximates $2[r(\mathrm{MZ}) - r(\mathrm{DZ})]$, where $r(\mathrm{MZ})$ is the correlation coefficient for monozygotic twins and $r(\mathrm{DZ})$ is that for dizygotic twins. This is just one example of the many methods commonly used in studies of heritability of human traits.

In the field of quantitative genetics, the concept of heritability is used to partition observable, phenotypic variation among individuals into genetic and environment components. However, there are problems with this descriptive statistic. First, the concept of heritability has no prescriptive power. In other words, it is not a measurement of how sensitive a character or trait might be to a change in environment. For example, a trait may have complete heritability ($H = 1$), yet can be altered drastically by environmental change. This can be seen in metabolic genetic disorders such as phenylketonuria and Wilson's disease. The broad heritability of the phenotypic outcomes in these diseases should equal 1 without intervention, yet we can effectively treat them through dietary interventions. A second problem with the concept of heritability is that it measures variation only *within* a population. It is not at all helpful in determining the causes of differences *between* populations. In fact, heritability estimates are population and time specific, not trait or character specific.

With respect to humans, it is extremely difficult to separate and measure the relative environmental and genetic contributions to phenotypic variation. Furthermore, although we can estimate the heritability of a trait, these estimates tell us nothing about the specific genes or environmental factors that contribute to observable variation. And, while we can use heritability estimates to quantify the percentage of variation attributable to genes at the *population* level, we cannot use this figure to determine the extent to which an *individual's* phenotype is determined by genes versus environment.

Unfortunately, the heritability concept has also had a history of abuse and misuse when applied to human population differences for traits such as intelligence. Studies have argued that racial differences in measures of intelligence, academic achievement, and crime rates are due to genetic, not environmental, differences. It has been demonstrated that the estimates of heritability for such traits within populations do not inform us about the genetic differences between populations. Although the unfortunate history of its use has contributed to clarifying the limits of the heritability concept, it also cautions us about some of the social repercussions of deterministic science.

—F. John Meaney and Cynthia Taylor

See also Genotype; Phenotype; Twin Studies

Further Readings

Bodmer, W. F., & Cavalli-Sforza, L. L. (1976). *Genetics, evolution, and man.* New York: W. H. Freeman.

Griffiths, A. J. F., Wessler, S. R., Lewontin, R. C., Gelbart, W. B., Suzuki, D. T., & Miller, J. H. (2004). *An introduction to genetic analysis* (8th ed.). New York: W. H. Freeman.

Mange, E. J., & Mange, A. P. (1999). *Basic human genetics* (2nd ed.). Sunderland, MA: Sinauer.

Hill, Austin Bradford (1897–1991)

Austin Bradford Hill has been called "the greatest medical statistician of the [20th] century" (Doll, 1993, p. 795), a distinction achieved without academic degrees in either statistics or medicine. Richard Doll summarized his major contributions as teaching an innumerate medical profession to think quantitatively and persuading them to adopt randomization in studies of therapies and laying the basis for the development of epidemiology by showing how the science could be expanded to discover the causes of noninfectious diseases.

Hill's lectures at the London School of Hygiene and Tropical Medicine (LSHTM) in the 1930s presented "an understandable and logical approach to the collection and interpretation of medical observations" (Doll, 1993, p. 795). The lectures were disseminated

worldwide when published in *The Lancet* and later as a textbook. The introduction of randomization came decades after its acceptance in other sciences because Hill chose to focus first on more basic lessons (particularly the need for valid comparisons and accounting for chance) to avoid alienating the nonscientific medical profession. In the 1940s, Hill convinced colleagues to use randomization to assess the value of streptomycin for treating tuberculosis; the clear results allowed Hill to persuade researchers of the scientific and ethical value of randomization for assessing effects of new treatments.

Hill's greatest contributions to observational research arose from the case-control and cohort designs he pioneered in collaboration with Doll to study the causes of lung cancer. Despite being a champion of randomization, Hill recognized that experiments are not practical for many important epidemiologic questions. Students of epidemiology are most familiar with Hill for his approach to drawing causal conclusions from observational data, particularly his widely misinterpreted considerations for causal inference.

Hill's life and contributions, and many positions and honors, have been chronicled in detail elsewhere. He was born in 1897 into a family described as one of the most intellectual in England. Intending to follow his father into a medical career, Hill first became a pilot in the Royal Navy Air Service. A near-fatal case of tuberculosis ended his service, saving him from the fate of many men of the Great War generation, and set his career path by precluding medical training. He instead studied economics, finding his way to health research through the mentorship of family friend Major Greenwood, and taking statistics classes from Karl Pearson. Hill joined Greenwood's Medical Research Council statistical staff at the National Institute for Medical Research in 1923 and followed Greenwood to LSHTM in 1933 as Reader in Epidemiology and Vital Statistics. From 1945 until his retirement in1961, Hill was Greenwood's successor as both Professor of Medical Statistics and Honorary Director of the Medical Research Council's Statistical Research Unit. Following his death on April 18, 1991, Hill was remembered by colleagues for pragmatism, clarity, persuasiveness, and wit. Doll (1995) wrote, "Those of us who had the benefits of his teaching and guidance in epidemiological research aim only to be able to pass it on to others with the same clarity of logic and expression, tinged, we like to think in our most optimistic periods, with the same sense of humour" (p. 162).

—Carl V. Phillips and Karen J. Goodman

See also Causation and Causal Inference; Hill's Considerations for Causal Inference; Study Design; Tobacco

Further Readings

Doll, R. (1993). Obituary: Sir Austin Bradford Hill, in *Sir Austin Bradford Hill* (multiple authors). *Statistics in Medicine, 12*, 795–808.

Doll, R. (1995). Sir Austin Bradford Hill: A personal view of his contribution to epidemiology. *Journal of the Royal Statistical Society, Series A, 158*, 155–163.

Special issue to mark the 85th birthday of Sir Austin Bradford Hill. (1982). *Statistics in Medicine, 1*(4).

Wilkinson, L. (1997). Sir Austin Bradford Hill: Medical statistics and the quantitative approach to prevention of disease. *Addiction, 92*, 657–666.

Hill's Considerations for Causal Inference

Austin Bradford Hill's widely cited list of considerations, from his 1965 address to the Royal Society of Medicine, presents factors to consider before inferring causation from an observed association. This list is often erroneously referred to as the "Bradford Hill criteria" or "causal criteria," although any list that lacks a basis for determining whether a condition is met, or for compiling such determinations to draw an overall conclusion, does not constitute a set of criteria, and Hill warns in the address that there are no "hard-and-fast rules of evidence" for causation (Hill, 1965, p. 299). This widespread misinterpretation is particularly unfortunate because it distracts from lessons in Hill's address that "offer ways to dramatically increase the contribution of health science"—for example, systematic error is often more important than random error; decisions must always be made using the epidemiologic evidence that exists; weighing costs and benefits matters more than statistical significance (Phillips & Goodman, 2004, p. 1).

Hill sought to challenge the mistake of treating tests of statistical association as sufficient evidence of causation, suggesting that we consider factors such as

(1) strength (of association), (2) consistency (across studied populations), (3) specificity (of cause and effect), (4) temporality (cause before effect), (5) biological gradient (dose-response), (6) plausibility, (7) coherence (with other evidence of various kinds), (8) experiment (intervention changes the outcome), and (9) analogy (with known causes of disease). Such lists of considerations do nothing to relieve the epistemic conundrum—well known since Hume—that causation cannot be directly observed. But pragmatically, conclusions about causation are needed to make decisions, and tools that improve these conclusions are useful.

Unlike many who cite him, Hill recognized that causal considerations (which he called "viewpoints") are sources of insight, but do not provide a basis for objective conclusions. The simple observation that there is no well-defined basis for judging whether a "criterion" has been fulfilled makes clear that the list can serve only as a guide to drawing subjective conclusions. In an address that covered many aspects of making decisions based on evidence, Hill cannot be faulted for not presenting a careful thesis on causal inference, or even the failure to define his viewpoints or provide guidance for those who wish to consider them (e.g., suggesting an epistemic hierarchy among them or methods for assessing whether they have been satisfied). He does, however, share blame for continued simplistic interpretation of an unelaborated list, having included his list in textbooks with no elaboration beyond what appeared in the original paper.

Historical Context

Creating checklists of causal criteria is a proclivity particular to health science and is not observed even in sciences similar to epidemiology, such as economics (in which Hill received his degrees). This may be because the routine of medical diagnosis (perform tests, declare a conclusion, move on) influences health science inquiry, creating a desire among many researchers to find something similarly algorithmic, or it could be because the usefulness of a checklist for infectious disease causation created the desire to find one for other etiologies.

The Henle-Koch postulates (or Koch's postulates) were developed for inferring disease causation by infectious agents. They resemble genuine criteria because infectious agents are fairly isomorphic to their disease manifestations, although the postulates were still too simplistic to account for complexities such as asymptomatic carrier states or susceptibility cofactors. For noninfectious diseases, complexities, including nonspecificity, multifactorial causes, and nondichotomous exposures, render a checklist approach even less informative, although this has been widely ignored, perhaps due to the desire to devise a chronic disease analog to Henle-Koch.

Hill was neither first nor last to present a list of causal considerations, and other versions are arguably more clearly presented and coherent; the close association of the concept with Hill probably owes to his eloquence and fame for other contributions rather than the superiority of his list. The 1964 U.S. Surgeon General's Committee on Smoking explicitly used their own list to determine whether smoking caused the diseases under review, but despite the high profile of that committee's report, it was overshadowed by Hill's similar list. Hill's list also sustained its dominance despite Mervyn Susser's later efforts, dating from the 1970s, to provide more completely fleshed-out causal considerations. Those not inclined to dismiss the notion of causal criteria outright could benefit from Susser's advice: "Epidemiologists have modified their causal concepts as the nature of their tasks has changed.... Indeed, the current set of criteria may well be displaced as the tasks of the discipline change, which they are bound to do" (Susser, 1991, pp. 646–647).

Contribution to Decision Making

Hill's address mixes varied epistemic approaches that reflect the necessity of muddling through a highly incomplete science to inform decisions. On one hand, the considerations are lessons in scientific common sense (a woefully uncommon commodity), showing how hypothetico-deductive reasoning can prevent many patently faulty causal conclusions. For example, we should doubt a causal conclusion if the association appears only when a particular data set is analyzed a particular way, or if a constellation of clearly unrelated health outcomes are associated with a particular "cause." Seen this way, the considerations are predictions based on the causal hypothesis, and a guide to seeking evidence that challenges the hypothesis.

But Hill also asked whether there was another explanation of the set of facts that would be equally likely as or more likely than cause and effect. This suggests a different epistemic approach: ruling out

alternate explanations for an observed association rather than testing predictions of the causal hypothesis suggested by the association. This approach calls for attention to confounding and other errors on the path from reality to data, which is not captured in the list of considerations.

Finally, Hill explicitly referred to different degrees of certainty we might hold about a possible causal relationship, and the costs of making the wrong decision in different situations, arguing that lack of definitive evidence of causation "does not confer upon us a freedom to ignore the knowledge that we already have, or to postpone the action it appears to demand at a given time" (Hill, 1965, p. 300). This view implies Bayesian reasoning and a pragmatic decision-theory approach. It is on this theme that Hill concluded his address and offered his most timeless lessons, a theme quite contrary to the naive, unrealistic dichotomies—proven/unproven theories, supported/unsupported hypotheses, met/unmet criteria—that dominate most discussions of causation in which Hill's name is invoked.

—Carl V. Phillips and Karen J. Goodman

See also Causation and Causal Inference; Hill, Austin Bradford; Koch's Postulates

Further Readings

Evans, A. S. (1978). Causation and disease: A chronological journey. *American Journal of Epidemiology, 108*, 249–258.

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58*, 295–300.

Phillips, C. V., & Goodman, K. J. (2004). The missed lessons of Sir Austin Bradford Hill. *Epidemiologic Perspectives and Innovations, 1*, 3.

Phillips, C. V., & Goodman, K. J. (2006). Causal criteria and counterfactuals; nothing more (or less) than scientific common sense. *Emerging Themes in Epidemiology, 3*, 5.

Susser, M. (1973). *Causal thinking in the health sciences: Concepts and strategies of epidemiology*. New York: Oxford University Press.

Susser, M. (1991). What is a cause and how do we know one? A grammar for pragmatic epidemiology. *American Journal of Epidemiology, 133*, 635–648.

Weed, D. L. (1995). Causal and preventive inference. In P. Greenwald, B. S. Kramer, & D. Weed (Eds.), *Cancer prevention and control* (pp. 285–302). New York: Marcel Dekker.

Histogram

A histogram is a pictorial representation of the information in a frequency distribution or a relative frequency distribution of a data set. Histograms are used to display information about continuous variables and typically require grouping data into classes; this feature distinguishes it from a bar graph, which is primarily used for categorical or discrete data with a limited number of categories that do not require further grouping. A histogram displays the overall pattern and deviation of the data, but due to the necessity of grouping data into classes, it is not an exact representation of the data values.

Histograms are most often created using computer software but may also be created manually. To make a histogram of a data set, proceed as follows:

1. Divide the range of the data into classes of equal width.
2. Count the number of observations in each class. These counts are called frequencies, and a table of frequencies for all classes is a frequency table.
3. Draw the histogram. The width or base of the bar represents the class, and the bar height is the class frequency. The graph is drawn with no horizontal space between the bars (unless a class is empty, so that its bar has zero height).

Example

Table 1 gives the survival times in days of 72 guinea pigs after they were injected with tubercle bacilli in a medical experiment. Figure 1 is a histogram for the same data.

This histogram is skewed to the right; that is, the right side of the histogram contains the larger half of the observations in the data and extends a greater distance than the left side. A histogram is skewed to the left when its left side extends a much larger distance than the right side; a histogram is symmetric if the right and left sides have essentially the same shape. A histogram with one major peak is described as unimodal; when a histogram has two major peaks, it is described as bimodal. If every interval has essentially the same number of observations, the histogram is described as a uniform histogram.

Table 1 Survival Times in Days of 72 Guinea Pigs After They Were Injected With Tubercle Bacilli

45	45	53	56	56	57	58	66	67	73
74	79	80	80	81	81	81	82	83	83
84	88	89	91	91	92	92	97	99	99
100	100	101	102	102	102	103	104	107	108
109	113	114	118	121	123	126	128	137	138
139	144	145	147	156	162	174	178	179	184
191	198	211	214	243	249	329	380	403	511
522	598								

Source: Adapted from Bjerkedal (1960).

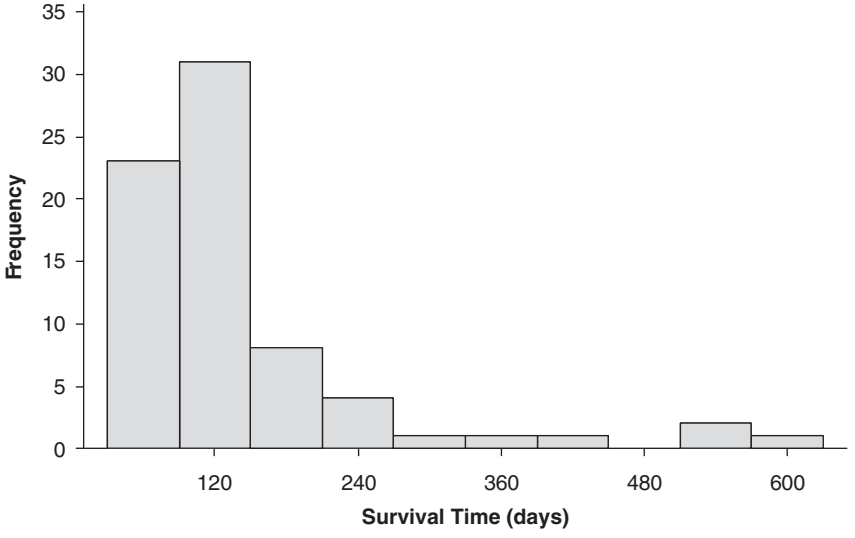


Figure 1 Survival Times in Days of 72 Guinea Pigs After They Were Injected With Tubercle Bacilli

Source: Adapted from Bjerkedal (1960).

Note that the selection of the number of bars to be included, which may also be stated as selection of the width of the bars, is an important decision. The same data displayed in histograms using varying bar widths can appear quite different, a fact that may be explored using Azzalini and Bowman's data and West and Ogden's histogram applet cited below.

—Renjin Tu

See also Bar Chart; Graphical Presentation of Data

Further Readings

Azzalini, A., & Bowman, A. W. (1990). A look at some data on the Old Faithful geyser. *Applied Statistics, 39*, 357–365.

Bjerkedal, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene, 72*, 130–148.

West, R. W., & Ogden, R. T. (1998). *Interactive demonstrations for statistics on the World Wide Web*. Retrieved September 5, 2006, from http://www.amstat.org/publications/jse/v6n3/applets/Histogram.html.

HIV/AIDS

AIDS is an acronym for *acquired immune deficiency syndrome*, a health condition that leads to the deterioration of the immune system and is caused by infection with the human immunodeficiency virus (HIV). AIDS is not a disease per se but rather a health syndrome that results in a weakened immune system, mostly due to the destruction of CD4+ T-cells, and that in turn results in susceptibility to numerous pathogens (viral, bacterial, fungal, and protozoal) that may lead to opportunistic infections and death. Individuals with AIDS are highly susceptible to these life-threatening pathogens and to certain types of cancer.

The number of individuals living with HIV and deaths due to AIDS increases daily, and because not all cases of HIV infection or AIDS are reported, official statistics are usually estimates rather than counts of reported cases and may vary by agency. According to the World Health Organization, in 2005 about 1.2 million Americans were infected with HIV, up from 1.1 million in 2003, and in 2005 about 16,000 died from AIDS (UNAIDS/WHO, 2006, Annex 2). The Centers for Disease Control and Prevention (CDC) estimates are slightly different: just under 1 million cases in the United States in 2005 and just over 17,000 deaths from AIDS (CDC, 2005). Recent data suggest that AIDS disproportionately affects communities of color, including African American and Latinos/Latinas. According to the CDC, in the United States, as of 2005, there were approximately 400,000 AIDS cases among African Americans, 387,000 among whites, and 156,000 among Latinos. The case rate per 100,000 population in 2005 was 75.0 for black non-Hispanics, 26.4 for Hispanics, 10.0 for American Indian/Alaska Natives, 7.5 for white non-Hispanics, and 4.9 for Asian/Pacific Islanders (CDC, 2005).

According to the UN estimates, across the world in 2005, there were 38.6 million individuals living with HIV/AIDS and approximately 2.8 million deaths. The UN report says 38.6 million were living with HIV in 2005, of whom 4.1 million were newly infected. There were 2.8 million deaths. The majority of the cases worldwide (36.3 million) are adults. Additionally, approximately 4.1 million people were newly infected in 2005. The vast majority of these cases (24.5 million, or approximately 63%) are in sub-Saharan Africa, where in recent years there has been an annual death rate from AIDS of 2 million people. In Asia (not including Oceania), there are approximately 8.3 million cases, 1.6 million in Latin America, and 1.5 million in eastern Europe and central Asia. North America and western Europe together have about 2 million cases. In addition to countries in

Table 1 Regional Statistics on HIV/AIDS

	People Living With HIV	*New Infections, 2005*	*AIDS Deaths, 2005*	*Adult Prevalence (%)*
Sub-Saharan Africa	24.5 million	2.7 million	2 million	6.1%
Asia	8.3 million	930,000	600,000	0.4%
Latin America	1.6 million	140,000	59,000	0.5%
North America and Western and Central Europe	2 million	65,000	30,000	0.5%
Eastern Europe and Central Asia	1.5 million	220,000	53,000	0.8%
Middle East and North Africa	440,000	64,000	37,000	0.2%
Caribbean	330,000	37,000	27,000	1.6%
Oceania	78,000	7,200	3,400	0.3%
Total	38.6 million	4.1 million	2.8 million	1%

Source: UN/WHO (2006, figure 2.3). Available from http://data.unaids.org/pub/GlobalReport/2006/200605-FS_globalfactsfigures_en.pdf.

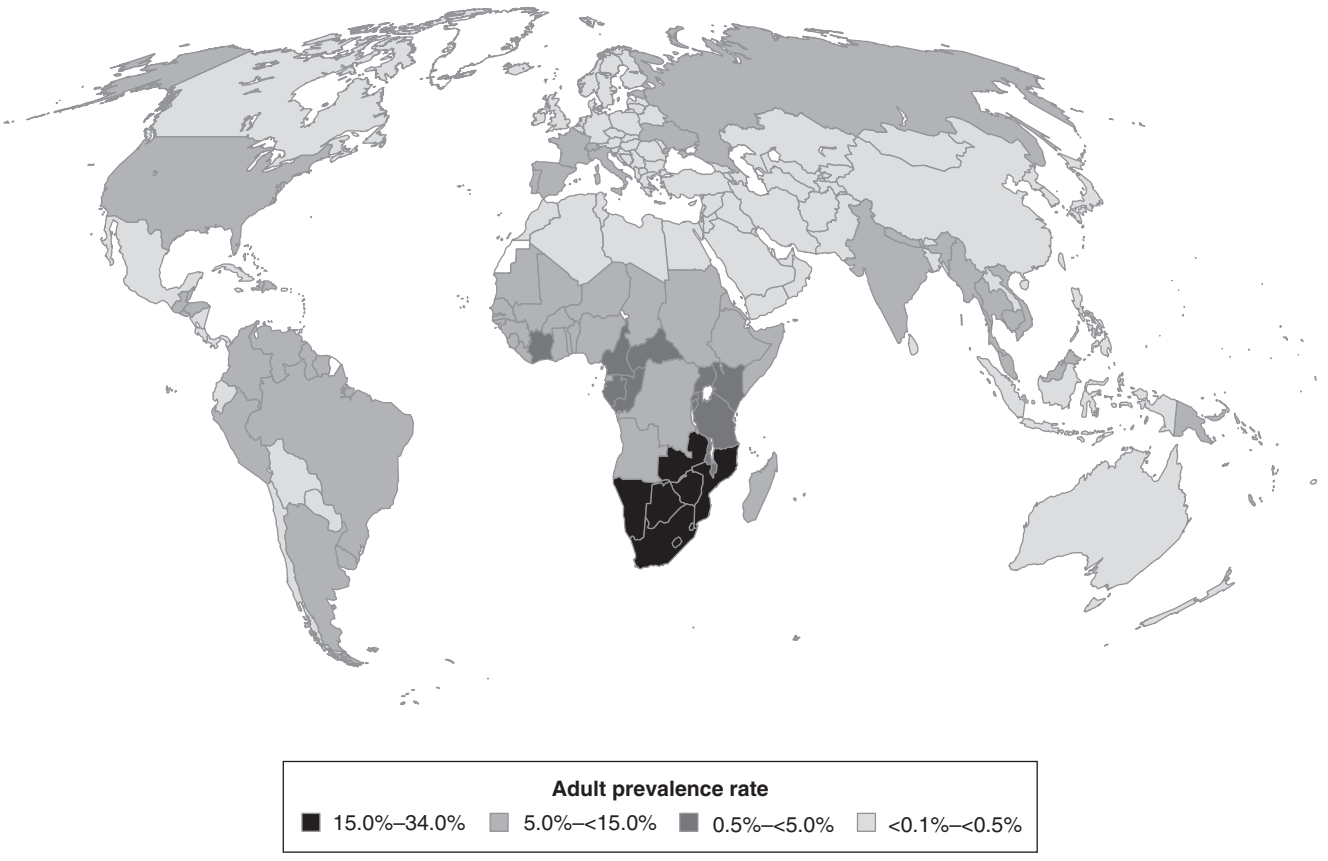


Figure 1 A Global View of HIV: 38.6 Million People Living With HIV/AIDS

Source: UN/WHO (2006, figure 2.4).

sub-Saharan Africa, countries of the former Soviet Union, including Russia, and Asian Pacific nations such as Thailand are experiencing vast increases in new HIV infections.

The first published report of AIDS occurred in the now landmark report of June 5, 1981, by the CDC in its weekly *Morbidity and Mortality Weekly Report* (*MMWR*). This report described five sexually active young gay men in Los Angeles who were diagnosed with *Pneumocystis carinii* pneumonia (PCP). The *MMWR* editor remarked that "the occurrence of pneumocystosis in these five previously healthy individuals with underlying immunodeficiency is unusual." Shortly thereafter, on July 3, 1981, another report appeared in *MMWR* identifying 26 young men in the New York area who had been diagnosed with PCP as well as the skin cancer Kaposi's sarcoma. While both these health conditions had been noted in the United States among elderly, the occurrence of these ailments among these men was of concern because at their age, they would have been expected to have normally functioning immune systems. Initial hypotheses to explain the development of this condition among these men included the use of amyl nitrate as well as the high levels of sexually transmitted infections in this segment of the population. Because the disease appeared to be confined to gay men, it was initially named GRID (gay-related immune deficiency). However, by 1982, it became apparent that the syndrome was not confined to gay men, and AIDS cases were documented among injection drug users. By the middle of 1982, 355 cases had been documented in five different states—California, Florida, New Jersey, New York, and Texas—and the disease was renamed AIDS. By early 1983, AIDS was being reported in 16 different countries around the world,

and in the first 3 years of the epidemic, the rates of diagnosis doubled every 6 months. Shortly thereafter, two research teams working independently in the United States and France each identified a virus and antibodies to the virus that appeared in the systems of individuals with AIDS and that were respectively named HTLV-III (human T-cell lymphotropic virus) and LAV (lymphadenopathy-associated virus). Recognizing that these two names were being used for the identical virus, researchers renamed the pathogen HIV in 1986.

Numerous theories have been postulated about the origins of HIV. The overriding scientific theory is that a similar virus existed in nature and was transmitted from animal to human, probably from a bite or scratch from an African primate, in the late 1940s or early 1950s. The virus then continued to mutate in humans, and these changes, coupled with social and political changes (such as population migration to the cities of Africa and increased trans-Atlantic air travel), created the conditions that allowed rapid transmission of the virus. Given that a 10-year incubation period for HIV is typical, this combination of factors is consistent with the rapid outbreak of AIDS in the 1980s. The virus is related to and may have evolved from SIV (simian immunodeficiency virus), which infects chimpanzees. Although chimpanzees are genetically 98.5% like humans, they do not develop AIDS. There is some evidence based on case studies that HIV may have entered the human population as early as the 1950s, and in 1969, what appears to be an AIDS-related death was documented in a 16-year-old boy in Missouri. Evidence of the HIV virus in the gay population of the United States was detected in stored blood samples of gay men in New York City and San Francisco who were participants in a large-scale hepatitis study in the 1970s.

Individuals who are infected with HIV but who have not yet experienced immune system deterioration are said to be HIV-positive and/or asymptomatic for AIDS. Those individuals who have advanced infection leading to a highly weakened immune system are classified as having AIDS. The period of time from infection to the development of full-blown AIDS varies, but most individuals progress to AIDS within a decade of infection. Of course, this is an ever-changing situation given the development of treatments for HIV infection. Furthermore, a small subset of individuals do not progress to AIDS for extended periods of time and are known as long-term survivors. Current research with these individuals is being undertaken to determine if newer, more effective treatments can be developed on the basis of what is learned about the ability of long-term survivors to avoid progression to AIDS.

The initial stage of HIV infection is referred to as the acute infection stage and is associated with the first few months of infection, up to 6 months, when HIV viral levels are high, as detected in the blood plasma, and HIV antibodies are relatively low. Antibody formation occurs during this period and is usually associated with a decrease in plasma viral levels. Individuals who are first exposed to HIV may experience a response to the infection that mimics flu-like characteristics, including fever, headache, fatigue, and enlarged lymph nodes.

In many cases, the acute infection stage is followed by a period when the individual is asymptomatic and experiences no major disease or syndromes and in which immune markers appear relatively normal. As viral replication continues and immune deterioration advances, individuals may develop symptoms such as high levels of fatigue, night sweats, diarrhea, and lymphadenopathy, which are associated with the beginning stage of AIDS referred to as AIDS-related complex (ARC).

The onset of AIDS or advanced HIV disease is determined by a set of health markers. The CDC has deemed any individual with less than 200 copies of CD4 + T-cells per cubic milliliter of blood or having less than 14% CD4 cells as part of their total T-cell count as having an AIDS diagnosis. CD4 cells are a key immune system marker needed for the body to effectively fight off infections. In addition, 26 other clinical conditions are associated with an AIDS diagnosis. These include PCP; Kaposi's sarcoma (KS), a type of skin cancer; cytomegalovirus (CMV), an infection that affects the eyes; candida, a fungal infection that can cause thrush (a white film in one's mouth, throat, or vagina); toxoplasmosis, a protozoal infection of the brain; mycobacterium avium complex (MAC or MAI), a bacterial infection that can cause recurring fevers; general sick feelings, problems with digestion, and serious weight loss; and tuberculosis, a bacterial infection that attacks the lungs and can cause meningitis. Over the course of the AIDS epidemic, the list of infections has been expanded to better represent infections and diseases that affect women (e.g., cervical cancer) and children (e.g., lymphoid interstitial pneumonia). In addition, individuals

with AIDS usually experience other conditions first evident during ARC, including chronic fatigue, weight loss, frequent fevers and sweats, persistent skin rashes or flaky skin, pelvic inflammatory disease in women that does not respond to treatment, and short-term memory loss. In addition, individuals with AIDS may develop herpes due to infection with the herpes simplex virus (HSV-1 and/or HSV-2), which affects the mouth, genitals, and/or anus, and can often develop into shingles, a painful condition that affects nerve endings. Children who have AIDS develop and grow more slowly than uninfected children.

To date, two main strains of HIV have been identified: HIV-1 and HIV-2. The majority of individuals who are infected with the virus in the United States and industrialized world are infected with HIV-1. HIV-2 is more commonly found in the countries of West Africa, although there is evidence of this strain of the virus in Western countries, including the United States. Both HIV-1 and HIV-2 have the same modes of transmission and are associated with similar opportunistic infections. The main difference between the two strains is that persons infected with HIV-2 seem to develop immunodeficiency at slower rates; in addition, the extent of the immunodeficiency is milder. Furthermore, at the initial stage of infection (known as the acute infection period), individuals with HIV-2 appear to be less infectious than those with HIV-1.

HIV is transmitted through the exchange of blood, semen, or vaginal fluids, and is primarily spread through sexual contact with an HIV-infected individual, through the sharing of needles and/or syringes with an infected individual, or through transfusion with infected blood products. In addition, infection may be passed from infected mother to unborn child (known as vertical transmission) during birth or after birth through the mother's breast milk.

Because HIV is a retrovirus, reproduction is dependent on host cells. Thus, HIV enters the bloodstream through a process known as adsorption and binds to target cells, primarily CD4+ cells, which are agents in the immune system that typically coordinate immune responses to pathogens and foreign bodies. HIV requires two receptors to gain entry into human cells: the CD4 receptors that are found on these cells of the immune system and the chemokine-binding coreceptors required by most strains of HIV. This coreceptor is known as CCR5. In addition, a molecule on the spikes of the HIV virus known as glycoprotein 120 is used for the adsorption process. After HIV has bound to both the CD4 and the chemokine receptors, an area of the virus is exposed that can then fuse with the cell, and that allows the entry of the viral RNA into the cell, during the process of penetration and infection. Once the viral RNA is in the cell, the use of the cell's genetic mechanisms (specifically the use of reverse transcriptase) allows replication and reproduction of viral genetic material to take place, and the protease is used to create protein coats around the newly formed genomes. In a sense, the infected cell becomes the system through which viral reproduction occurs during processes referred to as the lysogeny and activation of HIV. Once the cycle is completed, the newly formed virus buds from the cell, which may ultimately cause the final destruction of the CD4 cell.

In addition to CD4+ cells, HIV can bind to monocyte macrophages, which are immune system agents that typically engulf and destroy infectious pathogens. These cells can act as harbors for HIV by preventing the cell from destroying the foreign agent, HIV, and simultaneously allowing HIV to be disseminated throughout the body. The infection and reduction of both CD4 cells and macrophages cause an immune deterioration and the body's inability to effectively fight pathogens.

Recent research has indicated that variations in the genes that express CCR5 and CCR5 receptors may affect the susceptibility of cells to HIV infection. The most common variation that has been studied in the chemokine-related genes is CCR5Δ32, a mutation more commonly found in northern European whites than in any other racial group. Individuals with homozygous CCR5Δ32 variations may be less susceptible to HIV infection, while those who are heterozygous CCR5Δ32/wild type may progress to AIDS at slower rates.

Most current treatments for HIV infection interrupt the viral replication processes described above. Antiretrovirals typically interfere with the beginning, middle, or end of the replication process, and more recent treatments work by blocking the binding of HIV to receptor cells. Until the mid-1990s, monotherapy (i.e., using a single treatment) was the most commonly available treatment for individuals with AIDS. After the initial formulation of Zidovudine (AZT) in the late 1980s, treatments were slow in their development. However, by the mid-1990s, a multitude of treatments of varying classes had become available for treatment of HIV infection, and combination therapy (i.e., using

multiple treatments simultaneously) became and continues to be the standard of care for HIV infection, although the recommendations for initiating a treatment regimen have changed over the past decade swinging from a "hit-hard-and-hit-early" approach to a "wait-and-see" approach. At the onset of combination therapy, the treatment modality was referred to as HAART (highly active antiretroviral therapy), which has since evolved into ARV or ART (antiretroviral therapy).

ARV generally consists of a combination of treatments across classes and usually although not always consists of a medication from the class of protease inhibitors (PI) in combination with two or more medications from the other two antiviral classes: nucleoside reverse transcriptase inhibitors (nRTIs) and nonnucleoside reverse transcriptase inhibitors (nnRTIs), although the combinations are highly variable. Protease inhibitors (e.g., Invirase, Viracept, Norvir) prevent T-cells that have been infected with HIV from producing new copies of the virus by interrupting replication through binding and blocking the enzyme. HIV protease nRTIs (e.g., Retrovir, Epivir, Combivir) are incorporated into the DNA of the virus and stop the building process, resulting in incomplete DNA that cannot create a new virus. nnRTIs (e.g., Viramune, Rescriptor, Efavirenz) are designed to work at the final stage of the HIV replication process by preventing HIV protease from developing into HIV. In combination, these drugs work to prolong the suppression of HIV replication, resulting in decreases in viral load to lower and in some cases even to undetectable levels in the blood. As a result, immune system cells such as CD4+ may increase in number, thus strengthening the immune system. More recently, two other classes of drugs have been developed—nucleotide reverse transcriptase inhibitors and fusion blockers. Nucleotide reverse transcriptase inhibitors (NtRTIs) (e.g., Vireaid, a branded drug containing tenofovir disoproxil fumarate) are antiretroviral medications, which have a mechanism of action similar to nRTIs whereby they interfere with the DNA building process, resulting in incomplete DNA that prevents the virus from creating a new virus. A fusion blocker or fusion inhibitor (e.g., enfuvirtide) blocks the activity of HIV. When the virus attacks the cell, it sends out a projectile resembling an extremely small harpoon that anchors the virus to a CD4 T-cell. The virus pulls itself in via this anchor until it makes direct contact with the cell. Once full contact is made, the virus inserts its genetic material into the cell. This new class of drugs works at this entry point, preventing the virus from fusing to the cell.

ARV has shown to produce a significant reduction of opportunistic infections. In addition, ARV has also shown to increase survival and provide a sense of hope and optimism in HIV-positive persons. Although ARV has dramatically improved outcomes for HIV-positive individuals, less than perfect adherence to these complex regimens decreases the suppression of viral replication and increases the chances of developing drug-resistant viral mutations. The emergence of drug-resistant or more virulent strains of HIV may override the effectiveness of ARV and may reverse medical advances in HIV antiviral regimens. Further research suggests that suboptimal adherence to HAART has been associated with more rapid disease progression. While ARV has been heralded as a medical breakthrough in the treatment of HIV/AIDS, recent statistics indicate that the rate of decline for AIDS-related deaths has leveled off. In recent years, the development of genotypic and phenotypic tests has allowed researchers and medical practitioners to determine specific ARV drug resistance.

The detection of HIV infection is normally undertaken through the use of antibody screening, and the two most widely used tests are the ELISA (enzyme-linked immunosorbent assay) and the western blot confirmatory test. In a standard protocol, an ELISA is administered and results are confirmed with a western blot. These tests are traditionally undertaken together during one testing episode and the results are available within a week. More recently developed tests include the OraQuick technology, which allows for the detection of antibodies using a plasma level within a 20-min period, and the OraSure technology, which provides rapid results using only an oral swab. While these tests can provide evidence of HIV antibodies, they cannot detect the presence of the virus in the absence of antibodies as is the situation during the acute infection phase. However, HIV polymerase chain reaction (PCR) (i.e., viral load) testing can determine if the virus is circulating in the plasma. Use of the PCR in combination with a "detuned" ELISA can provide an indication of newly acquired HIV infections.

Over the past 25 years, a variety of strategies rooted in the biopsychosocial paradigm have attempted to prevent the further spread of HIV. Initial reactions to the epidemic focused on behavioral change strategies such as educating IV drug users about the use of clean

needles, and promoting condom use during sex for all segments of the population. By the mid- to late 1990s, these approaches seem to have been effective in leveling off the rates of new infections in the United States and western Europe, where there was a drastic decrease in transmission, especially among IV drug users. Efforts to curtail the transmission of HIV through sexual contact have been less effective. Recent statistics indicate an increase in new infections in the United States across all segments of the population, especially among men who have sex with men (MSM), specifically young MSM of color, as well as among African American heterosexual women. Other strategies with regard to sexual intercourse have included abstinence approaches, and most recently efforts have been focused on early detection of HIV so that individuals are aware of their status. This approach is based on the belief that a large proportion of HIV transmission occurs because individuals are unaware that they are infected. These efforts are dovetailed with strategies that encourage HIV-positive individuals to remain adherent to ARV as this is presumed to be associated with decrease in viral levels and in turn creates a lower probability of HIV transmission.

—Perry N. Halkitis

See also Partner Notification; Sexual Minorities, Health Issues of Sexual Risk Behavior; Sexually Transmitted Diseases

Further Readings

Alcamo, E. I. (2003). *AIDS; The biological basis* (3rd ed.). Boston: Jones & Bartlett.

Centers for Disease Control and Prevention. (1992, December 18). 1993 revised classification system for HIV infection and expanded surveillance for AIDS among adolescents and adults. *Morbidity and Mortality Weekly Report 41.* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention. (2005). *HIV/AIDS surveillance report, 2005* (Vol. 17). Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. Retrieved June 18, 2007, from http://www.cdc.gov/hiv/topics/surveillance/resources/reports/2005report/pdf/2005SurveillanceReport.pdf.

Halkitis, P. N., Parsons, J. T., Wolitski, R., & Remien, R. H. (2003). Characteristics of HIV antiretroviral treatments, access and adherence in an ethnically diverse sample of men who have sex with men. *AIDS Care, 15*(1), 89–102.

Karon, J. M., Fleming, P. L., Steketee, R. W., & De Cock, K. M. (2001). HIV in the United States at the turn of the century: An epidemic in transition. *American Journal of Public Health, 91*(7), 1060–1068.

Marmor, M., Jones, K., Thomas, S. M., Halkitis, P. N., Vogler, M., & Tse, D. B. (2006). Resistance to HIV infection. *Journal of Urban Health, 83*(1), 5–17.

Patterson, D. L., Swindells, S., Mohr, J., Brester, M., Vergis, E. N., Squier, C., et al. (2000). Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. *Annals of Internal Medicine, 133*(1), 21–30.

Stine, G. J. (1999). *AIDS update 1999.* Upper Saddle River, NJ: Prentice Hall.

UNAIDS/WHO. (2006). *Report on the global AIDS epidemic.* Geneva, Switzerland: Author. Retrieved June 25, 2007, from http://www.unaids.org/en/HIV_data/2006GlobalReport/default.asp.

U.S. Department of Health and Human Services. (2006, May 4). *Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents.* Retrieved from http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf.

Honolulu Heart Program

The Honolulu Heart Program (HHP) is a prospective cohort of 8,006 Japanese American men living in Oahu, Hawaii, born between 1900 and 1919 (45 to 68 years old at the time of enrollment). The main objective of the HHP was to identify risk factors for cardiovascular and cerebrovascular disease. The HHP started in 1965 and was funded by the National Heart, Lung, and Blood Institute (NHLBI) at the National Institutes of Health. The HHP is part of a larger study called the NI-HON-SAN study, which stands for Nippon, Honolulu, and San Francisco. The NI-HON-SAN study included men of Japanese ancestry born in the same period as the HHP who were living in Japan and San Francisco.

By observing Japanese Americans who had immigrated to Hawaii, the study investigators could examine different environmental, cultural, and lifestyle risk factors associated with the development of cardiovascular and cerebrovascular disease. Previous studies had revealed geographic variation in the prevalence and incidence of cardiovascular disease and stroke. The study participants were the same ethnically, but the Japanese American men had adopted a Western lifestyle. The investigators were able to reduce genetic variability and focus on factors related to immigration

that may be associated with the difference in cardiovascular and cerebrovascular disease incidence and prevalence.

At enrollment, the HHP participants were given a complete physical examination and received repeated follow-up examinations during the study period. The baseline examination (1965–1968) included collection of the following clinical data: blood pressure, resting heart rate, vital capacity, body mass index, serum cholesterol, triglyceride, glucose, hematocrit, uric acid, and urine. The men also provided demographic and lifestyle (smoking, alcohol consumption, diet, and physical activity) information.

The Japanese American men were found to have diets higher in animal protein and saturated fat as well as higher levels of serum cholesterol, triglyceride, uric acid, and glucose than Japanese men in Japan. The Japanese American men had higher rates of coronary heart disease than Japanese men in Japan, but the prevalence of cerebrovascular disease was lower in the HHS cohort than in the Japanese cohort. The investigators identified the following as risk factors for coronary heart disease: hypertension; higher levels of hematocrit, serum cholesterol, triglyceride, uric acid, and glucose; obesity; smoking; lower alcohol consumption; less physical activity; and lower lung function.

Since the study started, the following data have been collected on the participants: hospital discharges, death certificates, and autopsy records for morbidity and mortality due to coronary heart disease, cancer, and stroke. Surveillance of morbidity and mortality in the original cohort is still ongoing. The HHP provides a rich data source for researchers. From 1993 to 1996, the National Institute of Aging studied aging and dementia in the HHP cohort. Some examples of recent findings based on the HHP data include findings that magnesium is protective against heart disease, reduced caloric diet lengthens the life span, and older men can increase longevity by walking at least 2 miles per day.

—Britta Neugaard

See also Cardiovascular Disease; Migrant Studies; Nutritional Epidemiology

Further Readings

Kagan, A. (Ed.). (1996). *Honolulu Heart Program: An epidemiological study of coronary heart disease and stroke.* Amsterdam: Harwood Academic Press.

National Heart, Lung, and Blood Institute. (2006). *Honolulu Heart Program (HHP)*. Retrieved April 2, 2006, from http://www.nhlbi.nih.gov/resources/deca/descriptions/honolulu.htm.

Hormone Replacement Therapy

During the menopausal transition, there is a natural diminution of the sex hormones estrogen and progesterone. Hormone replacement therapy (HRT), or hormone therapy (HT), is a medical treatment for symptoms related to the menopause and the menopausal transition. Clinically, HT has also been administered to protect against disorders such as osteoporosis and atherosclerotic cardiovascular disease (CVD). However, recent studies have cast doubt on the protective effect of HT and have identified risks involved with its use.

Menopause and HT

Menopause is the permanent cessation of menstruation due to the loss of ovarian follicular function. The *perimenopause* refers to the time period immediately preceding menopause when fertility wanes and menstrual cycle irregularity increases. This period continues until 12 months after cessation of menses, at which time the woman is considered menopausal. The mean duration of the perimenopause is 4 years, while signs consistent with the perimenopause may precede the final menses by 2 to 8 years.

There is strong evidence that the transition to menopause is associated with vasomotor symptoms (hot flashes and night sweats). For instance, in one U.S. study, nearly 60% of women reported hot flashes in the 2 years before their final menses. There is also reasonable evidence that this period can cause sleep disturbances in some women. However, there is inconclusive or insufficient evidence that a decrease in ovarian mass is the major cause of mood swings, depression, impaired memory and the ability to concentrate, somatic symptoms, urinary incontinence, or sexual dysfunction. Notably, symptom intensity, duration, frequency, and effects on quality of life are highly variable.

The decision to use postmenopausal HT for the treatment of the symptoms and conditions listed above is complicated. Although many women rely on their health care providers for a definitive answer to the question of whether to use postmenopausal

hormones, balancing the benefits and risks for an individual patient is challenging, especially when the individual risk for HT-associated morbidity cannot be precisely quantitated. Despite this context and until the earlier years after the turn of the 20th century, many were prescribed HT as a means of alleviating vasomotor symptoms, for which its effectiveness has been well demonstrated. However, HT was increasingly promoted as a potential preventive strategy against disorders that accelerate after menopause, such as osteoporosis and atherosclerotic CVD.

These positions were based on results primarily from observational cohort studies. Although previous observational studies suggest that HT prevents cardiovascular and other chronic diseases, some of the apparent benefits may have resulted from differences between women who opt to take postmenopausal hormones and women who do not. Specifically, in these observational studies, those using HT tended to be healthier, have greater access to medical care, were more compliant with prescribed treatments, and maintained a more health-promoting lifestyle. On the other hand, randomized trials, which eliminate these confounding factors, have not consistently confirmed the benefits found in observational studies. For example, by enrolling more than 27,000 women from 50 to 70 years of age (mean age: 63 years), the Women's Health Initiative (WHI) was the largest randomized clinical trial of both estrogen-progestin and estrogen-alone postmenopausal hormone therapies. After a follow-up period of 5 to 7 years, both WHI hormone trials were stopped early because of an overall unfavorable risk-benefit ratio in the estrogen-progestin arm and an excess risk of stroke that was not offset by a reduced risk of coronary heart disease in the estrogen-only arm. Thus, recent clinical trials have raised doubts about the use of HT for prevention of chronic diseases, especially when initiated more than a decade past menopause.

Against this background and based on a synthesis of currently available evidence, the following summary of the risks and benefits of postmenopausal hormone therapy is provided.

Benefits and Risks of Postmenopausal HT

A summary of benefits and risks of postmenopausal HT in primary prevention studies in provided in Table 1.

Definite Benefits

Symptoms of Menopause

There is compelling clinical trial and observational study evidence that estrogen therapy is highly effective for controlling vasomotor and genitourinary symptoms. Although they are less effective than HT, alternative therapies such as antidepressants, gabapentin, clonidine, or vitamin E or the consumption of soy-based products/phytoestrogens may also alleviate vasomotor symptoms. For genitourinary symptoms, the efficacy of vaginal estrogen is similar to that of oral or transdermal estrogen.

Bone Density/Osteoporosis/Fractures

Estrogen slows the aging-related bone loss experienced by most postmenopausal women by reducing bone turnover and resorption rates. More than 50 randomized trials have demonstrated that postmenopausal estrogen therapy, with or without a progestogen, rapidly increases bone mineral density at the spine by 4% to 6% and at the hip by 2% to 3%. These increases are maintained during treatment.

Data from observational studies indicate a 50% to 80% lower risk of incident vertebral fracture and a 25% to 30% lower risk of hip, wrist, and other peripheral fractures among current estrogen users. In the WHI, 5 to 7 years of either combined estrogen-progestin or estrogen-only therapy was associated with a 30% to 40% reduction in hip fracture and 20% to 30% fewer total fractures among a population unselected for osteoporosis. Like estrogen therapy, bisphosphonates and raloxifene, a selective estrogen receptor modulator (SERM), have each been shown in randomized trials to increase bone mass density and decrease fracture rates. Similarly, a recently available option for treatment of osteoporosis is parathyroid hormone. Unlike estrogen therapy that is not combined with a progestin, the bisphosphonates, SERMs, and parathyroid hormone analogs do not appear to have adverse effects on the endometrium and may therefore be considered in women with a uterus. Increased physical activity and adequate calcium (1,000 to 1,500 mg/day through diet or supplements in two to three divided doses) and vitamin D (400 to 800 IU/day) intakes may also reduce the risk of osteoporosis-related fractures and should serve as first-line options for the prevention of declines in bone mineral density and associated fractures.

Table 1 Benefits and Risks of Postmenopausal Hormone Therapy (HT) in Primary Prevention Settings

		Benefit or Risk		
		Relative		*Absolute*
Outcome	*Summary of Effect*	*Observational Studies*	*WHI[a], Except Where Noted*	*WHI[†], Except Where Noted*
Definite Benefits				
Symptoms of menopause	Definite improvement	▼ 70–80% decreased risk	▼ 65–90% decreased risk[b]	
Osteoporosis	Definite increase in bone mineral density and decrease in fracture risk	▼ 20–50% decreased risk for fracture	E + P: ▼ 33% decreased risk for hip fracture E: ▼ 39% decreased risk for hip fracture	E + P: 50 fewer hip fractures (110 vs. 160) per 100,000 woman-years E: 60 fewer hip fractures (110 vs. 170) per 100,000 woman-years
Definite Risks				
Endometrial cancer	Definite increase in risk with estrogen alone; no increase in risk with estrogen-progestin	E + P: No increase in risk E: ▲ > 300% increased risk (1–5 years); > 600% increased risk (≥ 5 years)	E + P: No increase in risk E: Not determined	E + P: No difference in risk E: 46 excess cases per 100,000 woman-years
Venous thromboembolism	Definite increase in risk	▲ 110% increased risk	E + P: ▲ 106% increased risk E: ▲ 32% increased risk	E + P: 180 excess cases (350 vs. 170) per 100,000 woman-years E: 80 excess cases (300 vs. 220) per 100,000 woman-years

(Continued)

(Continued)

		Benefit or Risk		
		Relative		*Absolute*
Outcome	*Summary of Effect*	*Observational Studies*	*WHI[a], Except Where Noted*	*WHI[†], Except Where Noted*
Breast cancer	Increase in risk with long-term use (< 5 years) of estrogen-progestin	E + P: ▲ 63% increased risk (≥ 5 years) E: ▲ 20% increased risk (≥ 5 years)	E + P: ▲ 24% increased risk E: No increase in risk	10–30 excess cases per 10,000 women using 3 for 5 years; 30–90 excess cases per 10,000 women after 10 years use; 50–200 excess cases per 10,000 women after 15 years use (estimate is derived from observational data and WHI E + P findings)
Gallbladder disease	Probable increase in risk	▲ 110% increased risk	E + P: ▲ 67% increased risk E: ▲ 93% increased risk	E + P: 180 excess cases (460 vs. 280) per 100,000 woman-years E: 310 excess cases (650 vs. 340) per 100,000 woman-years
Probable or Uncertain Risks and Benefits				
Coronary heart disease	Probable increase in risk among older women and women many years past menopause; possible decrease in risk or no effect in younger or recently menopausal women	E + P: ▼ 36% decreased risk E: ▼ 45% decreased risk	E + P: ▲ 24% increased risk E: No increase or decrease in risk	E + P: 60 excess cases (390 vs. 330) per 100,000 woman-years E: No difference in risk
Stroke	Probable increase in risk	▼ 12% increased risk	E + P: ▲ 31% increased risk E: ▲ 39% increased risk	E + P: 70 excess cases (310 vs. 240) per 100,000 woman-years E: 120 excess cases (440 vs. 320) per 100,000 woman-years

Colorectal cancer	Probable decrease in risk with estrogen-progestin	▼ 8–34% decreased risk	E + P: ▼ 44% decreased risk E: No increase or decrease in risk	E + P: 70 fewer cases (90 vs. 160) per 100,000 woman-years E: No difference in risk
Diabetes mellitus	Probable decrease in risk	▼ 20% decreased risk	E + P: ▼ 21% decreased risk E: ▼ 12% decreased risk *	E + P: 150 fewer cases (610 vs. 760) per 100,000 woman-years E: 140 fewer cases (1,160 vs. 1,300) per 100,000 woman-years *
Cognitive dysfunction	Unproven decrease in risk (inconsistent data from observational studies and randomized trials)	▼ 34% decreased risk	E + P or E: ▲ 76% increased risk for incident dementia	120–230 excess cases of dementia per 100,000 woman-years

Source: Adapted from Manson and Martin (2001).

Notes: E = estrogen alone; E + P = estrogen plus progestin. Most studies have assessed conjugated equine estrogen alone or in combination with medroxyprogesterone acetate.

[a]WHI = Women's Health Initiative. The estrogen-plus-progestin arm of the WHI assessed 5.6 years of conjugated equine estrogen (0.625 mg/day) plus medroxyprogesterone acetate (2.5 mg/day) versus placebo. The estrogen-alone arm of the WHI assessed 7.1 years of conjugated equine estrogen (0.625 mg/day) versus placebo.

[b]Data are from other randomized trials. The WHI was not designed to assess effect of HT on menopausal symptoms.

* Not statistically significant.

Definite Risks

Endometrial Cancer

Results from a combined analysis of 30 observational studies indicate a tripling of risk for endometrial cancer among short-term (1 to 5 years) users of unopposed estrogen, while those who with use for ≥ 10 years had a nearly 10-fold higher risk. These findings are supported by results from the randomized Postmenopausal Estrogen/Progestin Interventions (PEPI) trial. In PEPI, 24% of women assigned to unopposed estrogen for 3 years developed premalignant endometrial hyperplasia, whereas only 1% of women assigned to placebo were found to have this condition. Notably, the concomitant use of a progestogen eliminates these risks.

Venous Thromboembolism

A meta-analysis of 12 studies of differing design found that current estrogen use was associated with a doubling of risk for venous thromboembolism in postmenopausal women. The relative risks were even greater (2.7 to 5.1) in the three clinical trials included in this analysis. Similarly, in the WHI clinical trials of postmenopausal women, there was a doubling of risk for the combined endpoint of venous and pulmonary thromboembolism among those in the estrogen-progestin arm, while those in the estrogen-only arm had a one-third higher risk for thromboembolism.

Breast Cancer

In contrast to findings for endometrial cancer, combined estrogen-progestin regimens appear to increase breast cancer risk more than estrogen alone. Furthermore, the increased risk of breast cancer among current or recent estrogen users is likely related directly to duration of use. For example, a meta-analysis of 51 case-control and cohort studies revealed that short-term use (< 5 years) of postmenopausal HT did not appreciably elevate breast cancer incidence, whereas long-term use (≥ 5 years) was associated with a 35% increase in risk.

Data from randomized trials also indicate that estrogen-progestin raises breast cancer risk. Results from the WHI indicate that over 5.6 years of follow-up, women assigned to the estrogen-progestin arm were 24% more likely to develop breast cancer than women assigned to placebo. Similarly, in the Heart and Estrogen/Progestin Replacement Study (HERS), 4 years of combination therapy was associated with a 27% increase in breast cancer risk. Although the latter finding was not statistically significant, the totality of evidence strongly implicates estrogen-progestin therapy in breast carcinogenesis. Conversely, over an average of 7.1 years of follow-up, those in the estrogen-only arm of the WHI did not experience an increased risk for breast cancer.

Gallbladder Disease

Large observational studies report a two- to threefold increased risk of gallstones or cholecystectomy among postmenopausal women taking oral estrogen. In the WHI, women randomized to estrogen-progestin or estrogen alone had a 67% and 93% greater risk, respectively, of undergoing cholecystectomy than those assigned to placebo. Increased risks were also observed in HERS.

Probable or Uncertain Risks and Benefits

Coronary Heart Disease/Stroke

On the basis of multiple observational studies demonstrating a benefit of hormone therapy, HT had, until recently, been enthusiastically recommended in the prevention of CVD. The biologic plausibility of such an association is supported by data demonstrating that exogenous estrogen has beneficial effects on both low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol levels. Administration of estrogen also favorably affects lipoprotein(a) levels, LDL oxidation, endothelial vascular function, and fibrinogen and plasminogen activator inhibitor-1. However, estrogen therapy raises both triglyceride and C-reactive protein levels and adversely affects several markers of thrombosis. In addition, estrogen may increase levels of matrix metalloproteinases, which have been implicated in the rupture of atherosclerotic plaques. Hence, the data on risk factors for CVD are inconclusive.

Randomized trials of estrogen or combined estrogen-progestin in women with preexisting CVD have not confirmed the benefits reported in observational studies. In HERS, a secondary prevention trial designed to test the efficacy and safety of estrogen-progestin therapy on clinical cardiovascular outcomes in high-risk women, the 4-year incidence of coronary mortality and nonfatal myocardial infarction was

similar in the active treatment and placebo groups, and a 50% increase in risk of coronary events was noted during the first year of the study among participants assigned to the active treatment group. Moreover, in the Papworth Hormone Replacement Therapy Atherosclerosis Study, the Women's Estrogen for Stroke Trial (WEST), and the Estrogen in the Prevention of Reinfarction Trial (ESPRIT), there were no cardiovascular benefits of the regimens studied. Thus, in clinical trials, HT has not proved effective for the secondary prevention of CVD in postmenopausal women.

Postmenopausal HT trials in women without preexisting CVD (primary prevention) also suggest an early increase in cardiovascular risk and absence of cardioprotection overall. In the WHI, women assigned to 5.6 years of estrogen-progestin therapy were 24% more likely to develop coronary heart disease and 31% more likely to suffer a stroke than those assigned to placebo. In the estrogen-only arm of the WHI, a similar increase in stroke and no effect on CHD were observed.

A closer look at available data suggests that timing of initiation of HT may influence the association between estrogen therapy and CHD. It is hypothesized that estrogen may have differential effects on clinical coronary events depending on stage of the atherosclerotic lesion. That is, estrogen may slow or even reverse progression in early stages of atherosclerosis but have adverse effects on advanced atherosclerotic lesions and vulnerable plaques. Specifically, the prothrombotic and proinflammatory effects of estrogen may manifest among women with subclinical lesions who initiate HT well after the menopausal transition, whereas women with minimal atherosclerotic disease who start HT early in menopause may derive cardiovascular benefit. Nonhuman primate data support this concept. Conjugated estrogens had no effect on the extent of coronary artery plaque in cynomolgus monkeys assigned to estrogen alone or combined with progestin starting 2 years (approximately 6 human years) after oophorectomy and well after the establishment of atherosclerosis. However, administration of exogenous hormones immediately after oophorectomy, during the early stages of atherosclerosis, reduced the extent of plaque by 70%.

Lending further credence to this hypothesis are results of subgroup analyses of observational and clinical trial data. For example, although there was no association between estrogen-only therapy and CHD in the overall WHI cohort, this therapy was associated with a CHD risk reduction of 37% among participants aged 50 to 59 years. In contrast, a risk reduction of only 8% was observed among those aged 60 to 69, and a risk increase of 11% was found among those aged 70 to 79. Due to the relatively small number of cases of myocardial infarction or coronary death, especially in the younger women, these intra- and inter-age-group differences were not statistically significant. However, when the definition of CHD was widened to include coronary bypass surgery or percutaneous coronary interventions, estrogen-only therapy was associated with a significant 45% reduction in CHD among women in the youngest age group. Furthermore, a meta-analysis of 30 trials with more than 26,000 subjects by Salpeter and colleagues found that HT was significantly associated with a 39% reduction in total mortality in those below 60 years of age. There was no association between HT and total mortality in those above the age of 60 years. Similarly, the risk for CVD mortality was 32% lower in those less than 60 years, but 11% higher in those above 60.

Clearly, further research is needed on age, time since menopause, and other clinical characteristics as well as on biomarkers that predict increases or decreases in cardiovascular risk associated with exogenous HT. Also, whether different doses, formulations, or routes of administration of HT will produce different cardiovascular effects remains uncertain.

Colorectal Cancer

Observational studies have suggested that HT reduces risks of colon and rectal cancer, although the estimated magnitudes of the relative benefits ranged from 8% to 34% in various meta-analyses. In the WHI, the sole trial to examine the issue, estrogen-progestin was associated with a significant 44% reduction in colorectal cancer over a 5.6-year period, although no benefit was seen with 7 years of estrogen-only therapy.

Cognitive Decline and Dementia

A meta-analysis of 10 case-control and 2 cohort studies suggested that postmenopausal HT is associated with a 34% decreased risk of dementia. Subsequent randomized trials, however, have failed to demonstrate any benefit of estrogen or estrogen-progestin therapy on the progression of mild to moderate Alzheimer's disease. The WHI, which assessed cognitive function and incidence of dementia among women randomized

to HT at age 65 or older, found no evidence of benefit and a suggestion of increased risk. Whether the discrepancies between observational studies and clinical trials are due to differences in age and time since menopause at HT initiation (as for CHD) remains uncertain.

—Matthew Allison and JoAnn Manson

See also Cancer; Cardiovascular Disease; Clinical Trials; Osteoporosis; Women's Health Issues

Further Readings

American College of Obstetricians and Gynecologists Task Force on Hormone Therapy. (2004). Hormone therapy. *Obstetrics and Gynecology, 104*(Suppl. 4), S1–S129.

Anderson, G. L., Limacher, M., Assaf, A. R., Bassford, T., Beresford, S. A, Balck, H., et al. (2004). Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: The Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association, 291*(14), 1701–1712.

Barrett-Connor, E., Grady, D., & Stefanik, M. L. (2005). The rise and fall of menopausal hormone therapy. *Annual Review of Public Health, 26*, 115–140.

Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. (1995). *Journal of the American Medical Association, 273*(3), 199–208.

Hsia, J., Langer, R. D., Manson, J. E., Kuller, L., Johnson, K. C., Hendrix, S. L., et al. (2006). Conjugated equine estrogens and coronary heart disease. *Archives of Internal Medicine, 166*, 357–365.

Hulley, S., Grady, D., Bush, T., Furberg, C., Herrington, D., Riggs, B., et al. (1998). Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/Progestin Replacement Study (HERS) Research Group. *Journal of the American Medical Association, 280*(7), 605–613.

Kasper, D. L., Branwald, E., Fauci, A. S., Hauser, S. L., Longo, P. L., Jameson, J. L., & Isselbacher, K. J. (Eds.). (2005). *Harrison's principles of internal medicine* (16th ed., chap. 327). New York: McGraw-Hill.

Manson, J. E., Hsia, J., Johnson, K. C., Rossouw, J. E., Assaf, A. R., Lasser, N. L., et al. (2003). Estrogen plus progestin and the risk of coronary heart disease. *New England Journal of Medicine, 349*(6), 523–534.

Manson, J. E., & Martin, K. A. (2001). Clinical practice: Postmenopausal hormone-replacement therapy. *New England Journal of Medicine, 345*(1), 34–40.

National Institutes of Health. (2005). National Institutes of Health State-of-the-Science Conference statement: Management of menopause-related symptoms. *Annals of Internal Medicine, 142*(12, Pt. 1), 1003–1013.

Rossouw, J. E., Anderson, G. L., Prentice, A. R. L., LaCroix, A. Z., Kooperberg, C., et al. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women. Principal results from the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association, 288*(3), 321–333.

Host

See Parasitic Diseases

Human Genome Project

The Human Genome Project (HGP) is one of the largest scientific endeavors in history, and researchers hope that it will lead to breakthroughs in all areas of biology, from basic science to clinical medicine. The HGP is relevant to epidemiologists because it provides data that can be used to identify genes that are associated with predisposition to disease. In addition, the HGP provides insight into the ancestral origin of modern human populations, which may provide clues about the geographic distribution of diseases. This entry reviews the history of the HGP, as well as the goals and outcome of the project.

History and Leadership

The HGP was conceived in 1985 during a scientific meeting held at the University of California, Santa Cruz (UCSC). At this gathering, Robert Sinsheimer, then chancellor of UCSC, proposed that sequencing the human genome was a feasible undertaking. In 1987, Charles DeLisi, director of the Office of Health and Environmental Research (OHER) at the U.S. Department of Energy (DOE), allocated $5.3 million to create the Human Genome Initiative, which funded national laboratories to develop methodologies for genome sequencing. In its own endeavor to sequence the genome, the National Institutes of Health (NIH) established the Office of Human Genome Research (OHGR) in 1988 and appointed James Watson to lead the sequencing project. The DOE and NIH efforts quickly merged into a joint collaboration with common goals. In 1989, the OHGR was expanded and renamed the National Center for Human Genome Research (NCHGR), and it became a grant-awarding

agency. The official beginning of the HGP occurred in 1990, when the DOE and NIH presented a 5-year plan to Congress. Francis Collins was named director of the NCHGR in 1993, following Watson's resignation in 1992 and the interim directorship of Michael Gottesman. In 1997, the NCHGR was elevated to the National Human Genome Research Institute (NHGRI). Led by Aristides Patrinos, the DOE created the Joint Genome Institute in 1997, which created a formal association between several DOE sequencing centers.

In a private effort to sequence the human genome, former NIH scientist J. Craig Venter formed Celera Genomics in 1998, pledging to sequence the human genome within 3 years at a cost of US$300 million. This announcement spurred a subsequent public-private HGP rivalry, which some credit with accelerating the NIH and DOE sequencing efforts.

At a press conference at the White House in June 2000, Collins, Venter, and Patrinos announced jointly that the working draft of the human genome was complete, several years ahead of schedule. The public and private sequencing efforts published the working draft of the human genome in concurrent issues of the journals *Nature* and *Science*, respectively. This working draft was edited and refined as more sequence data became available. In 2003, the finished sequence was announced, with an error rate of fewer than 1 per 10,000 bases (i.e., 99.99% accurate) and with no remaining gaps that can be sequenced with current technology.

Goals of the HGP

The first 5-year HGP plan presented to Congress in 1990 outlined seven project goals: mapping and sequencing the human genome; mapping and sequencing the genomes of model organisms; developing data collection and analysis methods; studying the ethical, legal, and social implications of the project; training scientists; developing technology; and establishing technology transfer. As the project progressed, these goals were revised and expanded in 1993 and 1998. Key examples of activities related to achieving the HGP goals are outlined below.

Mapping and Sequencing the Human Genome

Two techniques were used to generate the majority of sequence data for the HGP. The first, map-based shotgun sequencing, was used by the international public sequencing effort guided by the NIH and DOE. The second, whole-genome shotgun sequencing, was used by Celera Genomics.

Map-based sequencing begins by incorporating (cloning) large fragments of human genomic DNA into pieces of bacterial or, less commonly, yeast DNA. These hybrid DNA molecules are called bacteria artificial chromosomes (BACs) or yeast artificial chromosomes (YACs). Starting with a complete genome, it is possible to generate a library of BAC clones that represent the entire human genome with a high degree of redundancy. These clones are then analyzed for unique DNA elements that can map each clone to a specific location within the genome. This analysis often involves digesting the BACs with enzymes that cut DNA at specific sequence sites. If a BAC contains a unique piece of genomic DNA, the pattern of enzyme cutting will be unique. Two BACs containing overlapping genomic sequences will have overlapping enzyme-cutting patterns. By comparing the patterns of enzyme cutting across a large number of BACs, it is possible to assemble a collection of BACs that represent the entire human genome with minimal overlapping sequences. The second step in map-based sequencing is to sequence the human DNA cloned into each mapped BAC. This is performed by cutting each BAC's genomic DNA into small pieces and subcloning the DNA into vectors suitable for sequencing. This fragmentation, subcloning, and sequencing process is repeated for each BAC, each of which has already been positioned in the genome. Finally, the sequence fragments are assembled using the BAC physical genome map.

Whole-genome shotgun sequencing bypasses the steps of generating and mapping BAC clones, proceeding directly to sequencing genomic DNA fragments. The entire genome is fragmented and subcloned into sequencing vectors, which are sequenced with a high degree of redundancy. Then, the sequence data are assembled by computers to produce the whole-genome sequence.

The HGP revealed several characteristics of the human genome that were previously unknown. First, the completed human genome sequence revealed that there are about 30,000 human genes. This number was surprisingly low, because scientists had previously hypothesized that the human genome could contain 100,000 genes or more. The HGP sequence data also revealed the presence of bacterial DNA

sequences that have been incorporated into the human genome at some point in human history.

Mapping and Sequencing the Genomes of Model Organisms

By the time the finished human sequence was announced in 2003, the HGP sequencing centers had produced complete genome sequences for *Escherichia coli* (bacteria), *Saccharomyces cerevisiae* (yeast), *Candida elegans* (nematode), and *Drosophila melanogaster* (fruit fly), and draft genome sequences for the mouse, rat, and several other model organisms.

Data Collection and Analysis Methods

The vast amount of data generated by the HGP required extensive innovation in the field of bioinformatics. To address the need for powerful bioinformatics tools, the NIH and DOE established the Joint Informatics Task Force (JITF) to oversee all activities related to collection, storage, access, and sharing of HGP data. Most of the data generated by the HGP is stored in databases designed by the National Center for Biotechnology Information (NCBI), which was created by the National Library of Medicine to develop methods to store and analyze data related to human molecular biology.

Ethical, Legal, and Social Implications

When James Watson assumed leadership of the OHGR in 1988, one of his first actions was to set aside a portion of the HGP budget to investigate the ethical, legal, and social implications (ELSI) of sequencing the human genome. The HGP has since allocated 3% to 5% of its budget to ELSI activities, which has included developing policy options and funding grants for ELSI research. Issues examined by the ELSI program include genetic discrimination, privacy and confidentiality of genetic information, reproductive issues, religious and cultural implications of the HGP, and the potential commercialization of genetic material.

Training

The training activities of the HGP were implemented in several ways. Both the NIH and DOE offered pre- and postdoctoral training by offering training grants, research fellowships, and short courses. Emphasis was placed on cross-disciplinary training, including biology, mathematics, computer science, statistics, and engineering.

Technology Development

The HGP would have been impossible if not for the development and improvement of DNA sequencing technology. At the start of the HGP, state-of-the-art laboratories could sequence about 1,000 bases each day, at a cost of about $10 per base. The first automated DNA sequencing machines went on the market in 1987, and sequencing technology has steadily increased in speed, while decreasing in cost. Capillary-based sequencing machines became available in 1997, and the most advanced sequencing machines are now capable of sequencing 1,000 bases per second for a few cents per base.

Technology Transfer

The leaders of the HGP made an early commitment to release sequence data to the public, including industry partners, almost immediately. In 1991, the DOE and NHGRI adopted a policy that sequence data must be publicly released within 6 months of generation. At a meeting sponsored by Great Britain's Wellcome Trust in 1996, the International Human Genome Sequence Consortium strengthened its commitment to data sharing by agreeing to release all sequence data within 24 hr of generation. Focusing on generating free, publicly accessible data facilitated collaborations between the international sequencing centers.

The HGP and Epidemiology

Along with massive amounts of sequence data, the HGP also generated information about variation in the genome. Single nucleotide polymorphisms (SNPs) are fixed sites in the genome that can vary between individuals. On average, SNPs occur once in every 1,000 nucleotides, amounting to several million SNPs in the human genome. SNPs can, in rare cases, cause disease, but most have very little biological effect. However, SNPs can serve as proxies for neighboring genomic regions and can be a useful starting point for genetic association and linkage studies.

Accessing HGP Data

The entire human genome sequence is freely available on the Internet, on Web sites maintained by several

agencies. Initially, Celera Genomics' private sequencing effort planned to sell their sequence data, but the company eventually made their data freely available. The UCSC's Genome Bioinformatics Group hosts an online Genome Browser that enables users to browse the human genome sequence and the sequence of several model organisms. The NCBI maintains dbSNP, a database of SNPs in the human genome, as well as a Map Viewer that allows users to browse genome sequences.

—Megan Dann Fesinmeyer

See also Association, Genetic; Gene; Genomics; Linkage Analysis; National Institutes of Health

Further Readings

Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: Lessons from large-scale biology. *Science, 300*(5617), 286–290.

Green, E. (2001). Strategies for the systematic sequencing of complex genomes. *Nature Reviews Genetics, 2*, 573–583.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860–921.

Little, P. F. R. (2005). Structure and function of the human genome. *Genome Research, 15*, 1759–1766.

Schork, N. J., Fallin, D., & Lanchbury, S. (2000). Single nucleotide polymorphisms and the future of genetic epidemiology. *Clinical Genetics, 58*, 250–264.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science, 291*(5507), 1304–1351.

Hypertension

Hypertension, also known as high blood pressure, is a condition in which the pressure of a person's blood against their arterial walls is consistently too high. Hypertension is often called the "silent killer" because it may present no warning signs or symptoms obvious to the hypertensive individual and therefore may not be detected until another serious medical condition is diagnosed. It is a major risk factor for heart disease, heart failure, and stroke, and can result in other serious medical complications, including blindness and kidney failure.

Blood pressure is measured in millimeters of mercury (mmHg), using a device called a sphygmomanometer. When a person's blood pressure is recorded, it is commonly written as two numbers, for instance, 120/80. The first number is the systolic blood pressure, meaning the pressure when the heart contracts. The second number is the diastolic pressure, meaning the pressure when the heart rests between beats. The Seventh Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressures defines blood pressure levels for adults as follows:

- *Hypertension.* Systolic blood pressure of 140 mmHg or higher or diastolic blood pressure of 90 mmHg or higher
- *Prehypertension.* Systolic blood pressure of 120 to 139 mmHg or diastolic blood pressure of 80 to 89 mmHg
- *Normal or Normotensive.* Systolic blood pressure of less than 120 mmHg and diastolic blood pressure of less than 80 mmHg.

Persons with prehypertension are assumed to be at risk to progress to hypertension. If the systolic and diastolic blood pressures would place the person in different categories, the higher category is used. For instance, someone with a blood pressure reading of 145/85 would be considered hypertensive.

Classifying the blood pressure of children requires consideration of both their own blood pressure and the distribution of blood pressure for other children of their age, sex, and height. Hypertension for children is defined as that at or above the 95th percentile for children of their age, sex, and height, and prehypertension is defined as blood pressure of 120/80 mmHg or high but below the 95th percentile.

Hypertension may be defined as essential or secondary. *Essential hypertension* is the most common type and does not have a specific treatable cause. *Secondary hypertension* is due to some underlying condition, such as a kidney disorder or congenital abnormality; blood pressure generally returns to normal when the underlying problem is corrected. In addition, hypertension sometimes appears as a complication of pregnancy and may take two forms. *Pregnancy-induced* or *gestational hypertension* first appears when the woman becomes pregnant. *Preexisting chronic hypertension* is present in the woman before she becomes pregnant but only becomes apparent during the pregnancy. Hypertension is one of the

symptoms of preeclampsia, a disorder that appears in 5% to 8% of all pregnancies and is a major health risk for both mother and child. Preeclampsia is the second leading cause of maternal death in the United States and is also a leading cause of fetal complications, including low birthweight, premature birth, and stillbirth.

Prevalence

It is difficult to make statements about the prevalence of hypertension in a population because so many cases remain undetected. Study of this topic has generally relied on tabulating the number of individuals currently treated for hypertension, with the understanding that the number thus produced will underestimate the actual prevalence of the disease, or conducting physical examinations of a sample of subjects to determine the number experiencing high blood pressure on the day of their examination. In the latter approach, hypertension is defined as either measured high blood pressure or the use of medications to lower blood pressure (which would result in a normal reading on the day of examination). Using the second definition, data from the National Health and Nutrition Examination Survey (NHANES) for the years 1988 to 2000 suggests that hypertension rates are increasing among U.S. adults (aged 19 and older). In the 1999 to 2000 NHANES, 28.7% of participants were classified as hypertensive, with the highest rates among non-Hispanic blacks (33.5%), the elderly (65.4% among those age 60 or older), and women (30.1%). Overall, 68.9% were aware that they had hypertension, 58.4% were being treated for it, and 31.0% had it under control. Rates of control were significantly lower among women, Mexican Americans, and persons aged 60 years or older.

Hypertension is often thought to be primarily a problem in the industrialized world, but in fact it is also becoming an increasingly common health problem in the developing world. A number of factors have contributed to this, including increasing life spans and the rise in risk behaviors such as obesity, lack of physical activity, and unhealthy diet. Data collection and measurement issues are even more difficult in the developing world than in an industrialized country such as the United States, so estimates of the prevalence and health costs of hypertension must be interpreted with even greater caution. However, the World Health Organization estimates that worldwide, hypertension causes 7.1 premature deaths and the loss of 64 million disability-adjusted life years annually. Management of hypertension in developing countries is often complicated by the lack of recognition of the seriousness of the problem (compared with more traditional concerns such as childhood immunization programs and the provision of clean water and sewage disposal) and the unavailability or high cost of drugs to treat hypertension.

Prevention and Control

The causes of hypertension in an individual may include genetic factors, lifestyle choices, or both. Prevention and treatment of hypertension currently focuses on lifestyle choices and the use of medications. Current lifestyle recommendations to avoid and control hypertension include maintaining a healthy body weight (body mass index below 25); engaging in regular physical activity; eating a healthy diet, including fruits and vegetables high in potassium; reducing sodium (salt) intake; and moderating alcohol use. Types of medications used to treat hypertension include diuretics, beta-blockers, angiotensin-converting enzyme (ACE) inhibitors, angiotensin antagonists, calcium channel blockers, alpha blockers, alpha-beta blockers, nervous system inhibitors, and vasodilators.

The National High Blood Pressure Education Program (NHBPEP), established in 1972 as a cooperative effort among professional, governmental, and voluntary organizations, which is administered and coordinated by the National Heart, Lung, and Blood Institute of the National Institutes of Health, is a notable example of a successful public health campaign. The aim of the NHBPEP is to reduce death and disability related to high blood pressure, and its primary focus is education.. Specific NHBPEP activities include development of community programs to fight hypertension, evaluation and analysis of the results of major hypertension studies, and dissemination of information about hypertension through the mass media and through publications targeted to professionals, patients, and the general public. This campaign has proven highly effective: For instance, when the NHBPEP began, less than a quarter of the American population was aware that hypertension was a risk factor for stroke and heart disease, while today more than three fourths of the population are aware of that relationship. In addition, use of preventive services such as blood pressure screening by the general public, and control of hypertension through medication

and lifestyle changes, has increased substantially since NHBPEP began.

—*Sarah Boslaugh*

See also Chronic Disease Epidemiology; Nutritional Epidemiology; Obesity; Physical Activity and Health

Further Readings

Hajjar, I., & Kotchen, T. A. (2003). Trends in prevalence, awareness, treatment and control of hypertension in the United States, 1988–2000. *Journal of the American Medical Association, 290*(2), 199–206.

National Heart, Lung, and Blood Institute. (2004). *The seventh report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure*. Bethesda, MD: Author. Retrieved December 13, 2006, from http://www.nhlbi.nih.gov/guidelines/hypertension/jnc7full.htm.

Weir, M. R. (Ed.). (2005). *Hypertension.* Philadelphia: American College of Physicians.

World Health Organization, International Society of Hypertension Writing Group. (2003). World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. *Journal of Hypertension, 21*(11), 1983–1992.

Web Sites

Centers for Disease Control and Prevention, High Blood Pressure: http://www.cdc.gov/bloodpressure.

Hypothesis Testing

Statistical inference is the science of making conclusions, decisions, or inferences about a population based on information obtained in a sample. The procedure that leads to rejecting or not rejecting specific statements about a population is called hypothesis testing. If there is only one population under investigation, researchers conduct a one-sample test. If they are comparing two populations, they conduct a two-sample test. Additionally, there are multisample tests (also known as *k*-sample tests) to consider more than two populations.

A statistical hypothesis is an assumption or statement or inference regarding one or more parameters of a population distribution, or the type or nature of a population. Both the following are examples of statistical hypotheses: (1) The specificities of two diagnostic tests are the same and (2) the average score on the ABC test is 70. Statistical hypotheses are always about population parameters, while a decision to accept or reject a hypothesis is based on a test statistic derived from a sample.

Hypothesis testing is a decision-making process for evaluating claims about a population. A hypothesis test determines if an observed value of a statistic (from a sample) differs enough from a hypothesized value of a parameter (from a population) to draw the inference that the hypothesized value is not the true value. "Enough" difference is measured by using subjective judgment and statistical conventions to set a predetermined acceptable probability of making an inference error due to sampling error.

The most commonly used current method for hypothesis testing is formally called "null hypothesis significance testing." Most people today simply refer to this method as significance testing or hypothesis testing and often use the terms interchangeably. However, this method combines the two historical procedures that are known separately as "significance testing" and "hypothesis testing." In this article, unless a distinction is necessary, the term *hypothesis test* will refer to the current method of the "null hypothesis significance test."

History

Fisher's significance test concentrates on a Type I error and the associated *p* value. In the 1920s, R. A. Fisher developed his significance test of a null hypothesis that is a statistical inference based on deductive probabilities yielding a *p* value (which measures the discrepancy between the null hypothesis and the data). This test specifies only one hypothesis (the null hypothesis), but the alternative hypothesis is implicitly defined. The outcome of Fisher's test is a statement of whether significant or nonsignificant results are obtained. The significance or nonsignificance is decided based on the *p* value compared with a predetermined level of allowable Type I error (α).

In the early 1930s, Jerzy Neyman and Egon Pearson developed a method called the Neyman-Pearson hypothesis test that requires that researchers specify two point hypotheses, as well as Type I (α) and Type II (β) error rates in advance of conducting the experiment. These prespecifications are used to create a decision rule for rejecting or accepting the

null hypothesis. Common problems encountered with this approach are that (1) most researchers are unwilling to specify values for the alternative hypothesis and (2) point hypotheses are always untrue if calculations are carried to sufficient decimal places.

Since the early 1990s, there has been a move toward significance testing incorporating power analysis, which is the essence of hypothesis testing. This synthesis of Fisher's and Neyman-Pearson's approaches to testing is formally called the "null hypothesis significance test." In this test, the values of α (commonly 0.05) and $1-\beta$ or power (typically ≥ 0.80) are set before the experiment as in the Neyman-Pearson method. Complementary null and alternative hypotheses are formulated according to Fisher's method. Then the calculated p value is compared with α to make a decision about rejecting the null hypothesis (also from Fisher's method). If the test statistic is sufficiently atypical given the distribution under the null hypothesis, then the null hypothesis is rejected.

An *a priori* power analysis is usually conducted to determine the needed sample size for a given α, $1-\beta$, and effect size. A *post-hoc* power analysis determines the power observed for a particular study given the sample size, α, and the observed effect size. Post-hoc power analysis is most often conducted when results are nonsignificant or when there is a discrepancy between statistical significance and clinical significance (i.e., when a statistically significant result is too small to have clinical significance). For instance, a trial comparing the amount of weight lost by individuals taking a particular drug versus a placebo could return a statistically significant result, but the difference in weight lost might not be large enough to improve their health or reduce their risk of disease, and therefore not be clinically significant. Clinical significance cannot be determined mathematically but is based on clinical judgment.

The calculations for significance and power are based on a closed system formed by α, $1-\beta$, sample size, and effect size. When three of these are established, the fourth is determined. Power increases if the effect size and/or the sample size and/or α increases. At a given α, sample size, and effect size, one-tail tests are more powerful (i.e., $1-\beta$ is larger) than two-tail tests as long as the effect is in the specified direction. Tests on equal sample sizes are more powerful than unequal sample sizes. Using control variables or covariates in an analysis of covariance (ANCOVA) can increase power compared with simply conducting an analysis of variance (ANOVA).

If a test does not yield the desired treatment effect, it should not be immediately assumed that the effect does not exist. Other possible reasons for obtaining statistically nonsignificant results include a poorly executed intervention (leading to a weak effect size) or insufficient sample size (leading to an underpowered test).

Definitions

An alternative hypothesis or research hypothesis, typically denoted as H_A, H_1, or H', is a general statement the researcher wishes to support if using the null hypothesis significance test or Fisher's significance test. This type of hypothesis may be stated as H_A: The means of treatments A and B do not differ or H_A: $\mu A - \mu B \neq 0$. If the researcher is using the Neyman-Pearson's hypothesis test, then this hypothesis is a specific statement containing a value, such as H_A: $\mu A - \mu B = 10$.

The assumptions should be clearly stated when conducting a test of hypothesis. These assumptions typically relate to the population(s) being sampled.

A critical region or rejection region is a set of values of the test statistic that cause the null hypothesis to be rejected. This region of test values indicates that a significant difference exists.

The critical values are values that mark the boundaries of a critical region; for example, ± 1.96 are critical values for the z test with $\alpha = 0.05$. These values separate the critical or rejection region from the noncritical region.

The effect size is a standardized measure of the magnitude of the treatment effect that is independent of sample size. Due to the standardization, it allows comparisons of results across several studies and, therefore, is used in meta-analysis. Jacob Cohen wrote the definitive work on classification of effect sizes.

The hypothesized value is the numerical value stated in the null hypothesis, for example, 5 is the hypothesized value in H_0: $\mu A - \mu B = 5$.

The level of significance, level of the test, or size of the test are all terms for the maximum acceptable Type I error rate, α, or the probability of making a Type I error for a specific hypothesis test. This level is most commonly set at 0.05; it is seldom larger than 0.10, but may be extremely small (0.001 or smaller). The level is selected based on the researcher's

subjective judgment of the consequences of a Type I error in a particular research situation. It may be thought of as limiting the possibility of a false positive. The level of significance is the complement of the confidence coefficient; for example, if a researcher conducts a 95% confidence test, the confidence coefficient is 0.95 and the α level is 0.05.

A null hypothesis or status quo hypothesis, typically denoted as H_0 and called "H-oh," "H-naught," or "H-nought," is the hypothesis to be tested, which is usually stated as the absence of a difference or effect. It is a statement or conjecture about a population parameter that characterizes the distribution of a variable in a population and can be tested using a sample statistic. Typically, the null hypothesis is the opposite or reverse of what the researcher actually believes; it is put forward to allow the data to contradict it.

A noncritical or nonrejection region is a region of values that indicates that the difference found in the sample was probably due to chance and that the null hypothesis should not be rejected.

A one-tailed, one-sided, or directional test indicates that the null hypothesis should be rejected when the test value is in the rejection region on one side of the mean. This region may be in either the left or the right tail of the distribution, depending on the alternative hypothesis (see Table 1). This test investigates whether a change occurred, but only in the direction of interest.

The p value or observed significance level of the test measures the strength of the evidence against the null hypothesis. It is the probability of obtaining a test statistic at least as contradictory to the null hypothesis as the one seen in the sample, given the sample size used and the assumption that the null hypothesis is true. It is also the smallest fixed value of α at which the null hypothesis can be rejected. Smaller values indicate stronger evidence against the null hypothesis, but the p value is not a measure of the effect size. Because the p value is influenced by sample size as well as effect size, a small effect in a large study may have a greater p value than a large effect in a small study.

The power of the test is the probability of rejecting the null hypothesis given the true population value differs from the hypothesized value. Power is defined as $1-\beta$ or $1-$probability of committing a Type II error. The true power of an experiment is not known to the researcher; for example, the probability of rejecting H_0: $\mu A = \mu B$ depends on the true population means and the true means are unknown without a census of the population. However, power for a statistical test can be calculated using hypotheses about how much the population values are expected to differ. Larger values for power may be thought of as limiting the possibility of a false negative.

The sampling experiment is the actual collection of data through a designed experiment, survey, or observational study.

A test statistic is the formula or rule for computing a test value. For example, the general form for most one- and two-parameter hypotheses is

$$\text{test statistic} = \left(\frac{\text{sample statistic} - \text{hypothesized value}}{\text{standard error of statistic}}\right),$$

although this form does not hold for the F or χ^2 (chi-square) tests.

A test value or observed test statistic is the numerical value obtained from a statistical test. This value is compared with the values within the rejection or critical region to determine whether the null hypothesis should be rejected.

A two-tailed, two-sided, or nondirectional test indicates that the null hypothesis should be rejected when the test value is in either of the two rejection regions. It is typically associated with an alternative hypothesis that includes the "not equal" sign (see Table 1). This test investigates whether a change from the hypothesized value(s) in any direction was present in the study.

A two-tailed probability is a probability computed considering differences in both directions or both tails of the sampling distribution. It is the sum of the two tail probabilities.

Method

It is much easier to prove a statement false than to prove it true. For example, researchers may hypothesize

Table 1 Mathematical Operators for Different Types of Hypotheses

Two-Tailed or Two-Sided	*One-Tailed or One-Sided Left*	*One-Tailed or One-Sided Right*
$H_0: =$	$H_0: \geq$	$H_0: \leq$
$H_A: \neq$	$H_A: <$	$H_A: >$

that 30% of the population below age 18 in a particular city have not been properly immunized. If they take a sufficiently large, random sample and conclude that only 1% of the target population has not been properly immunized, then they have strong evidence that their hypothesis was wrong or false. On the other hand, if they discover 28% have not been properly immunized, what should they conclude? It is possible that 30% is the true population percentage; it is also possible that 35% is the true value, or 26% or 29.4%. However, only one value is the *true* population value; they will not know for certain unless they contact everyone in the area.

To deal with this problem, scientists state two hypotheses that are mutually exclusive and exhaustive. These two hypotheses cover the entire range of possibilities (exhaustive) and only one may be true at a time (mutually exclusive). Then, the scientist seeks to reject one of the hypotheses, thereby supporting the other. This idea is based on the concept of proof by contradiction. Science requires a conservative approach to decision making: A conclusion is accepted by the scientific community only if the data supporting it are strong enough to convince a skeptic. In most cases, scientists view finding significance when none exists to be a more serious flaw than failing to find significance when it does exist. For this reason, α is commonly set lower than β.

To state hypotheses correctly, the researcher must translate the conjecture or claim being tested into mathematical statements of the null and alternative hypothesis. Table 2 contains some common phrases used to state hypotheses and their translation into mathematical notation.

Table 2 Phrases Used in Stating Null and Alternative Hypotheses

$=$	$\neq$
Equal to	Not equal to
Same as	Different from
Unchanged	Changed from
Does not vary	Varies
Is congruent to	
$\leq$	$\geq$
Less than or equal to	Greater than or equal to
Does not exceed	Not less than
At most	At least
No more than	
$<$	$>$
Less than	More than
Strictly less than	Strictly more than
Below	Above
Smaller than	Bigger than
Has reduced	Has grown
Fewer than	More than
Is slower	Exceeds
	Greater than
	More often

If researchers wanted to test the prior hypothesis about the proportion of the population below 18 years that have not been properly immunized being 30%, they would state the null and alternative hypotheses as: H_0: $\pi = 0.30$ and H_A: $\pi \neq 0.30$. If they were interested in whether the proportion exceeded 30%, the hypotheses would be stated: H_0: $\pi \leq 0.30$ and H_A: $\pi > 0.30$. Note that, in either case, the two hypotheses cover all the range of possible values.

A test is available for almost any situation in which researchers may wish to draw conclusions from the data. The choice of test depends on the question being asked and the type of data available. However, the quality of the result depends on the assumptions of the test being made. For example, many tests assume that the sample was randomly drawn from the population and may yield biased results if this assumption is violated. There is no formal procedure to check whether the study design meets the requirements to conduct the test; this depends on the researcher's judgment.

All statistical tests, whether significance or hypothesis tests, follow a logical sequence that results in an ultimate decision to reject the null hypothesis or fail to reject the null hypothesis.

The basic steps to follow in hypothesis testing are shown in Table 3. Note that the steps differ slightly depending on whether the researcher plans to use

Table 3 General Steps for Any Hypothesis Test

Rejection Region Test	p *Value Test*
Define the population under study.	
State H_0 and H_A.	
Decide on the appropriate test statistic.	
Set α.	
Determine the rejection region associated with the test statistic and α.	
Collect data.	
Calculate the test value.	
Compare test value to rejection region.	Calculate p value.
	Compare p value to α.
Make decision about H_0.	
Summarize the results.	

a rejection region to make the final decision or to use the p value compared with α approach.

Relationship to Confidence Intervals

If the hypothesized value of a population parameter lies outside the $100(1-\alpha)\%$ confidence interval for that parameter, then the null hypothesis would be rejected at the α level of significance. For H_0: $\mu_1 - \mu_2 = 0$, if a $100(1-\alpha)\%$ confidence interval centered on $\bar{x}_1 - \bar{x}_2$ does not contain 0, then the researcher would reject H_0 at the α level of significance.

Failure to reject a null hypothesis at a specified α level means that a $100(1-\alpha)\%$ confidence interval for the parameter of interest would contain a value indicating no difference from the hypothesized value.

Conclusion

Hypotheses may deal with more than one parameter at the same time. If hypotheses examine three or more parameters, the null hypothesis is typically stated as no difference between any of the parameters, while the alternative states that at least one differs from the others. For example, the typical ANOVA null hypothesis is H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$.

Within the ANOVA example, if the null hypothesis were rejected, then follow-up post-hoc tests may be performed. These tests, called multiple comparison tests, explore which means or combinations of means differ from each other. To this end, there are two special types of hypotheses that are used: pairwise hypotheses and complex hypotheses. Pairwise hypotheses examine all possible pairs of means, and complex hypotheses examine various linear combinations of the means, such as $(\mu_1 + \mu_2)/2 = \mu_3$ or $(\mu_1 + \mu_2)/2 = (\mu_3 + \mu_4 + \mu_5)/3$. Researchers must be aware that when conducting multiple independent tests on the same data, the experiment-wise probability of a Type I error (or the probability of committing a Type I error somewhere within the simultaneous tests) increases. The experiment-wise error rate is controlled by the formula: $1 - (1-\alpha)c$, where c is the number of comparisons being made. For example, if four comparisons are conducted with $\alpha = 0.05$, the overall error rate is approximately 0.185. To keep this error rate low, it is best to use an adjustment factor such as the Bonferonni or Greenhouse-Geiser corrections.

When reporting the results of hypothesis tests, it is best to report both the p value and the effect size(s). Several research journals have begun rejecting manuscripts that contain only p value information. The effect size reported may be one that evaluates the proportion of variance explained in the analysis, such as R^2 (R-squared) or η^2 (eta-squared), or the standardized differences in statistics, such as the standardized differences in means (Cohen's d), or both.

—Stacie Ezelle Taylor

See also Multiple Comparison Procedures; Sample Size Calculations and Statistical Power; Study Design; Type I and Type II Errors

Further Readings

Bluman, A. G. (1992). *Elementary statistics: A step-by-step approach.* Dubuque, IA: Wm. C. Brown.

Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Freund, J. E., & Simon, G. A. (1992). *Modern elementary statistics* (8th ed.). Englewood Cliffs, NJ: Prentice Hall.

Girish, S. (2006). A shift from significance test to hypothesis test through power analysis in medical research. *Journal of Postgraduate Medicine*, *52*(2), 148–150.

Hassard, T. H. (1991). *Understanding biostatistics.* St. Louis, MO: Mosby-Year Book.

Hildebrand, D. K. (1986). *Statistical thinking for behavioral scientists.* Boston: PWS.

Hurlburt, R. T. (2006). *Comprehending behavioral statistics* (4th ed.). Belmont, CA: Thomson/Wadsworth.

Mendenhall, W. (1993). *Beginning statistics: A to Z.* Belmont, CA: Duxbury/Wadsworth.

Ott, R. L., & Mendenhall, W. (1994). *Understanding statistics* (6th ed.). Belmont, CA: Wadsworth.

Ross, S. M. (1996). *Introductory statistics.* New York: McGraw-Hill.

Sirkin, R. M. (1995). *Statistics for the social sciences.* Thousand Oaks, CA: Sage.

I

Icelandic Genetics Database

In 1998, the Althingi (Iceland's parliament) passed the Act on a Health Sector Database (No. 139) authorizing a centralized database of nonpersonally identifiable health data for all Icelanders. The Althingi subsequently granted the firm deCODE genetics, Inc. (henceforth, deCODE) an exclusive license to establish and operate the database. Combined with pedigrees and biological samples, deCODE aimed to use the database to identify alleles of genes that predispose Icelanders to specific diseases. This project was the first large-scale commercial attempt to combine population genomics and epidemiological genetics. The history and implementation of the Act is of great interest as a case study for epidemiological genetics and the issue of genetic discrimination and as what constitutes fair commercial use of publicly contributed and controlled data.

The objective of the bill was to establish a new central database of medical records, with the goal of improving public health in Iceland. The legislation intentionally did not apply to existing collections of medical records or the curating of biological samples. The Act mandated that the Health Sector Database be created and maintained in Iceland by a single licensee, with the licensee paying all associated costs. Standards for maintaining anonymity were described, and, while any Icelander could proactively opt not to be included in the database or be removed at any time by contacting the Icelandic Director General of Public Health, all Icelanders were presumed to grant informed consent. Quite simply, the bill authorized a licensee to enter extant medical records of all Icelanders into a single database. No licensee was specified in the bill.

The history of the Act on a Health Sector Database is inextricably tied to the company deCODE. Founded in 1996 by Kári Stefánsson and Jeffrey Gulcher, deCODE has its headquarters in Reykjavik, Iceland. An Icelander himself, Kári Stefánsson played a key role in initiating the creation of the legislation, and deCODE was subsequently granted the license to establish the Health Sector Database. The stated business goal of the company is to discover, develop, and commercialize drugs to treat common diseases. Research operations were begun in Iceland with the intent to use population genetics and genomics theory to mine the Icelandic population for disease-causing genetic polymorphisms. The significance of the Health Sector Database for deCODE is that the medical records can be meshed with detailed genealogical information and supplemented with genetic data for specific samples of Icelanders.

Opposition within Iceland to the Act and the intentions of deCODE was spearheaded by Mannvernd (Association of Icelanders for Ethics in Science and Medicine). Mannvernd opposed the Act for two primary reasons: (1) Icelanders were presumed to grant consent regarding inclusion in the database and were required to proactively decline to participate, and (2) the exclusive license granted to deCODE would constitute a virtual monopoly on research concerning medical genetics of Icelanders. With regard to the former objection, in 2003, the Icelandic Supreme Court ruled that certain provisions of the Act pertaining to privacy and informed consent were unconstitutional.

With regard to the latter objection, the license to establish the Health Sector Database does not preclude other medical and scientific investigations concerning Icelanders. However, in a small isolated population, the presence of such a large commercial research entity may inhibit the funding and operation of smaller academic or commercial research endeavors. It is too soon to gauge the impact of the project with regard to the issue of genetic discrimination either at the level of an individual or in a population.

Arguably, Iceland is an ideal population for searching for the genetic bases of diseases: There are extensive genealogical records, state-maintained medical records, and historical sources documenting a founder event. Recent founder events have an impact on patterns of linkage disequilibrium such that it is easier to correlate diseases with particular genes or loci on chromosomes. During the initial stages of the development of the company, deCODE investigated genetic variation in European populations to establish that Icelanders constituted a relatively homogeneous population. Early studies estimating the amount of admixture between Norwegian and Irish populations during the founding of Iceland had reported conflicting results. Estimates from these early papers were based solely on blood group frequencies, however, and may reflect differential effects of selection. It is not possible to distinguish the effects of selection versus changes in demography using data from a single locus.

To characterize the impact of Iceland's population history better and to compare evidence for population structure in Iceland with that of other European populations, deCODE published several papers surveying mitochondrial, Y-chromosomal, and autosomal loci. Compared with other European populations, Iceland generally represented a population that showed a stronger signature of genetic drift associated with the initial founding event. These publications by researchers employed by deCODE stirred up an interesting debate in the literature of population genetics concerning appropriate methods for surveying genetic variability and, more important, the interpretation of the results.

deCODE successfully established the Health Sector Database, organized pedigrees of families with particular diseases, and collected samples for sequencing from the same. As of January 2007, scientists of deCODE have published more than 60 papers linking genetic variants found in the Icelandic population to a multitude of diseases, and currently the company has drugs in advanced stages of development for asthma, heart attack, pain, peripheral arterial disease, obesity, type 2 diabetes mellitus, schizophrenia, and vascular disease or stroke. From a theoretical perspective, it will be interesting to observe how rich the vein of heritable diseases present in Icelanders will be: That is, the success of deCODE as a commercial venture of epidemiological genetics is predicated on detecting linkages of disease-causing variation within and between branches of Icelandic pedigrees. As a population, Icelanders may be relatively more homogeneous than other European populations, but many of the Icelandic disease alleles manifest as mutations more recent than the founding event. Therefore, there is a limit to the number of diseases maintained within Icelanders that might be attributed to particular genes. Anecdotal evidence that deCODE is approaching a threshold of discovery within the Icelandic population may be found in the decreased number of publications reporting new associations. Clearly, incorporation of other populations with available pedigree and health records will open avenues for further successful research using similar methods and strategies. Other European populations with state-run health care systems are obvious candidates. The long-term success of deCODE in terms of improving the public health of Icelanders will determine whether legislation in other countries will use as a model the Act on a Health Sector Database.

—Christopher Tillquist and Chandler Gatenbee

See also Family Studies in Genetics; Genetic Disorders; Genomics; Linkage Analysis

Web Sites

deCODE Genetics: http://www.deCODE.com.
Mannvernd: http://www.mannvernd.is/english.

Illicit Drug Use, Acquiring Information on

Determining whether an individual is or is not using a specific drug is a key step in reducing the public health risks associated with illicit drug use; further, this determination can also have legal implications. There

are three common ways of acquiring this information. The first method is to ask the person directly, that is, to elicit a self-report of his or her behavior. The second method is to perform biochemical analysis of bodily fluids or tissues. The third method is to ask a collateral (e.g., spouse or other relative, social worker, or probation officer) who may have knowledge.

Self-Reports

Among the advantages of self-reports are that they can provide estimates of illicit drug use over a very long time window (e.g., "Have you ever, in your life, used ...") that cannot be matched by biological measures, which provide only an assessment of a single recent point in time. Self-reports can also provide qualitative data on how a substance was consumed, motivation for use, and also circumstances of use, none of which can be obtained from biological measures.

Various approaches are used to collect quantifiable self-report data. Among the older methods are paper surveys that the respondent completes and interviews where the interviewer asks the questions and completes the questionnaire. Newer methods include the use of laptop computers, PDAs, and other electronic devices on which the respondent completes either the whole survey or confidential portions of it. Other methods include automated telephone interviews and, recently, Internet-based surveys.

Self-reports of illicit drug use are highly variable in terms of consistency (reliability) and accuracy (validity). Consistency and accuracy are affected by multiple factors, including memory, cognition, and psychological processes, the latter including tendencies toward deception (whether deliberate or unconscious) and social desirability. Memory failures lead to haphazard errors, such as guessing, or to systematic errors such as forward telescoping, where events are incorrectly remembered as having occurred more recently than the actual occurrence. Cognitive factors include the tendency for people to use enumeration or counting to recall infrequent events, while remembering frequently occurring events (e.g., "How many cigarettes do you smoke in a day?") as rates, resulting in less accurate estimation as the means of providing a quantity. When the events are too numerous for each one to be recalled independently, rates are used. Thus, an individual who smokes 30 cigarettes in a day is unlikely to remember each of them. Instead such individuals remember how many per hour, or how often they need to buy a new pack of 20.

Deliberate deception is highly situation dependent and tends to result in underreporting of use. Lower accuracy rates occur in situations where there are negative consequences for use, such as in the criminal justice system, where individuals may potentially fear legal results of exposure. Perceived negative consequences are probably the most important variable implicated in deliberate deception. Self-deception occurs when individuals tell the inquirer what they perceive as the truth, but after having deceived themselves (e.g., "I don't need heroin. I could quit at any time. I just like it" or "I don't need a drink before lunch; I just like it."). Finally, social desirability is the inclination to respond to questions in a socially approved manner. This may result in either over- or underreporting of use, depending on the context. For instance, youth in high school might exaggerate marijuana use to boost their standing with peers. Alternately, pregnant casual smokers may present themselves as nonsmokers due to the social opprobrium associated with any substance use during pregnancy. Some researchers have suggested that there is a personality component that influences the tendency to produce socially desirable responses.

A number of techniques have been employed as means of improving self-report consistency and accuracy. Basic approaches include developing a rapport with the respondent before asking sensitive questions, and the use of simple sentence structure in questions. Other methods include computer-assisted interviewing (CAI); audio computer-assisted self-interviewing (audio-CASI), where only the respondent hears the questions and answers privately via computer; random response options for providing sensitive material, where once the information is entered, the identity of the respondent cannot be recovered; and the timeline follow-back technique and others that provide memory anchors to help respondents with more accurate recall. In the timeline follow-back technique, the respondent is asked to remember a distinctive event such as a holiday, a birthday, or when a significant recent event happened. She or he is then asked to recall consumption using this event as a frame of reference. Another technique, called the bogus pipeline method, compares responses with a fictitious machine that the respondent is told can detect the truth. However, results from research on the effectiveness of this technique is mixed.

Biochemical Analysis

The second method of obtaining information on substance use is by performing a biochemical analysis on a sample specimen of the person's bodily fluids or tissues—that is, by conducting a drug test. Most drug testing focuses on detecting the parent drug or metabolites of the so-called National Institute on Drug Abuse (NIDA) five: cocaine, opiates, PCP (phenylcyclidine, also known as angel dust), amphetamines, and cannabis (marijuana). These are drugs of particular concern for misuse and abuse. However, in certain situations, legal drugs, such as alcohol, may also be the focus of testing.

Commercial drug testing often proceeds in two stages. Stage one uses low-cost screening tests, where false-negative results are minimized. In the second stage, expensive confirmatory assays of the positive screens are conducted, where false-positive results are minimized.

Screening tests show whether a drug is present at levels above a standard (government-set) limit or cutoff. The most commonly used ones rely on an immunological reaction for detection. A biological specimen is exposed to a substrate that contains an antibody of the drug in question. Any drug present above a predetermined amount will bind to the antibody, producing a visible color response.

Confirmatory assays both verify the presence of a given drug and measure its concentration. Gas chromatography/mass spectroscopy is the technique most often used for such analyses. It involves extracting the parent drug or its metabolites from the biological specimen and then breaking it down to the constituent molecules, which are then weighed and measured.

The vast majority of drug tests use urine as the specimen, although more recent ones can analyze hair, sweat, and saliva. Other fluids, such as blood or meconium, may be used in specialized circumstances, such as during postmortem, postaccident, postpartum, or medical examinations.

The ability of a test to detect a drug depends on how the substance is broken down or metabolized by the individual. That, in turn, is contingent on the chemical structure of the drug or metabolites, the mode of ingestion, the level of use, and the individual's specific metabolic processes, among other factors. In addition, drugs and their metabolites are deposited in various biological fluids at different times and at varying levels. For instance, analysis of urine, sweat, or saliva will show recent use (within a few hours), while hair analysis can only detect distant use (few weeks or more, depending on the length of the hair and its rate of growth). Hair analysis, on the other hand, can be used to measure use over long time periods, with each approximately half inch from the crown reflecting 1 month of potential use. On average, cocaine and heroin can be detected in urine for 2 to 4 days, amphetamines for 2 weeks, and marijuana for up to 30 days.

The physical location of drug testing is affected by the underlying rationale for its use. Samples from workplace testing tend be analyzed in laboratories, where confirmation testing of positive screens can be easily performed. Point of collection or on-site tests are generally used for routine monitoring that occurs in a more controlled environment e.g., drug treatment centers). The type of biological fluid used also plays a role. Currently, hair and sweat can only be tested in laboratories, while urine and saliva can be examined in on-site tests.

Use of Drug Testing

Drug testing plays a key role in monitoring drug use among various groups and in many different settings. Workplace testing occurs as part of pre-employment screening, during postaccident investigations, and as part of random monitoring of employee activity. Employers cite various reasons for testing, including the need to reduce both potential liability for employee misconduct and insurance costs related to substance abuse and its consequences. The United States military conducts testing as a precondition of entry. Police engage in drug testing when investigating potential crimes, mostly in cases where drug use is an element of the offense (driving while intoxicated) or as part of postmortem examinations of suspicious deaths. Individuals under the control of the criminal justice system, in particular probationers or parolees, generally submit to mandatory testing as a condition of their sentence. Routine testing also occurs in drug treatment facilities, so that providers can monitor the course and success of treatment. Other settings where drug tests are likely to take place are in high schools, where administrators sometimes make nondrug use a condition of extracurricular participation, and in medical settings, specifically in emergency rooms and as part of pregnancy-related health care.

Large-scale surveys that monitor drug use prevalence do not routinely collect drug testing data. This is likely due to the prohibitively high expenses involved with testing. A notable exception was the Arrestee Drug Abuse Monitoring (ADAM) survey of drug use by arrested individuals. Studies that collect both self-report and drug testing information tend to show differences between what people report as their use and what the drug tests show. Most times, this is attributed to inaccuracies in self-report, although in some instances the discrepancy may be linked to the limitations of the drug test itself.

Collateral Reports

When a spouse or relative, social worker, probation officer, or other reports on illicit drug use, he or she often has only partial knowledge since he or she may only know what the user tells them or what he or she observes directly. This knowledge may be useful in special instances but often underestimates drug use when compared with either self-reports or drug testing.

—Hilary James Liberty and Angela Taylor

See also Drug Abuse and Dependence, Epidemiology of; Interview Techniques; Questionnaire Design; Screening; Sensitivity and Specificity

Further Readings

Harrison, L. (1997). *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIDA Research Monograph No. 167). Rockville, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Institute on Drug Abuse, Division of Epidemiology and Prevention Research.

Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: A review and meta-analysis. *American Journal of Public Health, 84*, 1086–1093.

Siegel, J. (2005). *Handbook of forensic drug analysis.* Burlington, MA: Elsevier Academic Press.

Sobell, L. C., & Sobell, M. B. (1990). Self-report issues in alcohol abuse: State of the art and future directions. *Behavioral Assessment, 12*, 77–90.

Stone, A. A., & Turkkan, J. S. (2000). *The science of self-report: Implications for research and practice.* Mahwah, NJ: Lawrence Erlbaum.

Swift, R. (2003). Direct measurement of alcohol and its metabolites. *Addiction*, *98*, 73–80.

Immigrant and Refugee Health Issues

Over the past 30 years, the United States has experienced one of the largest waves of immigration in its history. Understanding the health status and needs of immigrants and refugees is important because of their growing numbers and their effect on the overall health of the nation. Until recently, most health research did not collect data on nativity and immigration status. When immigration status was collected, many studies grouped all immigrants together, despite marked differences within subgroups with respect to culture, socioeconomic status, patterns of immigration, and health status. This entry provides a general overview of the effects of migration on health and some of the unique health issues that immigrants and refugees face. It also examines reasons why protecting and promoting the health of immigrants and refugees requires improved collection of data on the health of foreign-born populations in the United States; reduction of barriers to accessing health care; and the development of health care systems that can deliver medically, culturally, and linguistically appropriate care.

Demographics

There are four primary categories of immigrants to the United States: legal immigrants, refugees, asylees, and undocumented immigrants. Legal immigrants are individuals who have been granted permission by the U.S. Bureau of Citizenship and Immigration Services to enter the United States either permanently or temporarily. A refugee is a person who is forced to flee his or her country because of persecution or war and who is granted refugee status prior to entering the United States. An asylee is also someone who is fleeing his or her country because of persecution or war, but an asylee enters the United States without legal permission. Once an asylee is in the United States, he or she must apply for refugee status. If denied, he or she will be deported. Undocumented immigrants do not have permission to be in the United States and can be deported when discovered. In 2006, there were an estimated 35.7 million foreign-born persons in the United States, of whom approximately 10 million were undocumented.

Unlike the early 1900s, when the majority of immigrants came from Europe, the majority of immigrants

to the United States since the 1980s were born in Latin America or in Asia. The countries of origin for the majority of immigrants to the United States are listed in Table 1. In 2005, 53,813 persons were admitted to the United States as refugees. The leading countries of origin for refugees were Somalia, Laos, Cuba, and Russia. Immigrant and refugee populations are heavily concentrated in eight states (California, New York, Florida, Texas, New Jersey, Illinois, Minnesota, and Washington), although there has been significant growth in immigrant and refugee populations in nearly all states over the past decade.

Immigrants and refugees are often characterized as poorer and less educated than U.S.-born persons; however, there are significant exceptions to this generalization, including considerable subgroup differences by country of origin. This variation can affect data collection and interpretation, health status, and the potential success of interventions to improve health.

The Healthy Immigrant Effect

The effect of migration on health is controversial. Some studies have shown that first-generation immigrants enjoy superior health and have lower mortality rates compared with U.S.-born persons, despite higher rates of poverty and worse access to health care. This has been dubbed *the healthy immigrant effect.* As immigrants and refugees adopt traditional American health behaviors over time, their health status begins to converge with that of the general population. The literature on how acculturation influences health status and health behaviors is often difficult to interpret because there are few validated and consistent measures of acculturation. In addition, there is evidence that the effects of acculturation vary, depending on the health behavior or outcome being studied, and among men and women. More recent research has found that some immigrant and refugee groups experience much higher rates of disease and poor health than previously suspected. Understanding how migration affects health is challenging because of gaps in national databases, the heterogeneity of immigrant populations, and difficulty in tracking immigrant populations over time. Data on immigrant health status are also often difficult to interpret because of the uncertain impact of selection biases. For example, immigrants and refugees to the United States may represent the most healthy and motivated individuals—those who are able to make the long journey to the United States, have healthier diets and lifestyles, and engage in fewer risk-taking behaviors. They may also return to their native country prior to dying, and therefore not be counted in U.S. death records or vital statistics.

Table 1 Top 10 Countries of Birth of the Foreign-Born Population in 2004

Country of Birth	*Numbers, Civilian Noninstitutionalized Population*
Mexico	8,544,600
China	1,594,600
Philippines	1,413,200
India	1,244,200
Cuba	1,011,200
Vietnam	997,800
El Salvador	899,000
Korea	772,600
Dominican Republic	791,600
Canada	774,800

Source: Data from U.S. Department of Homeland Security (2006).

Infectious Diseases and Immunizations

Many immigrants and refugees migrate from countries with a high incidence of infectious diseases. Before entering the United States, immigrants and refugees are required to undergo screening for significant communicable diseases; however, overseas screening may be incomplete. Depending on the country of origin, immigrants and refugees should be screened for infectious diseases after entering the United States with tuberculin skin tests (TSTs), hepatitis B testing, and evaluation for ova and parasites.

The proportion of tuberculosis (TB) cases in the United States occurring among foreign-born persons increased progressively during the 1990s; in 2003, persons born outside the United States accounted for 53% of reported cases. The majority of immigrants and refugees to the United States come from areas that have a high incidence of TB, and are therefore at

higher risk for tuberculosis than are those born in the United States. Studies have shown that immigrants from TB-endemic areas remain at elevated risk for TB for approximately 10 years after migration. The elevated risk is mostly attributed to reactivation of latent tuberculosis infection (LTBI). Therefore, targeted testing with a TST and treatment of LTBI is central to reducing TB among immigrants and refugees.

Many immigrant and refugee groups have a higher prevalence of infection with the hepatitis B virus, with the highest prevalence among those from sub-Saharan Africa and East and Southeast Asia. Hepatitis B screening identifies susceptible individuals who can be offered vaccine, and infected individuals can be offered treatment and educated about ways of preventing transmission and reducing future liver damage.

Data from several sources indicate that immigrants and refugees are less likely to be up to date on routine immunizations than are those born in the United States, and they may not have appropriate documentation of prior immunizations. The 1999 and 2000 National Immunization Surveys indicate that foreign-born children were almost 45% less likely to be up to date for the recommended immunization coverage compared with U.S.-born children. Indirect evidence of low immunization rates for rubella among foreign-born adults is provided by data from the National Notifiable Diseases Surveillance System, which indicates that the majority of confirmed rubella cases are now known to occur in foreign-born persons. Current guidelines recommend that immigrants and refugees with no records or incomplete immunizations should receive vaccines during routine health visits unless contraindicated.

Mental Health

The literature suggests that immigrants, in general, have lower rates of mental illness compared with those born in the United States and second- or third-generation immigrants. The Epidemiologic Catchment Area Study (ECAS) and the National Comorbidity Survey (NCS) found that first-generation Mexican immigrant adults had a lower prevalence of mental disorders compared with U.S.-born Mexicans and with the rest of the U.S.-born sample. Only one population-based study of mental disorders among Asians in the United States, the majority of whom were immigrants, has been conducted. It found lifetime and 1-year prevalence rates for depression to be roughly equal to the general population rates found in the same urban area.

Among immigrants, refugees are considered to be at high risk for mental disorders because many were exposed to significant trauma and torture prior to immigration. Many studies document high rates of posttraumatic stress disorder, depression, and anxiety among refugees and asylees throughout the world and in the United States, although few of these studies have been longitudinal.

Studies consistently suggest that immigrants and refugees underuse mental health services. The reasons include the stigma associated with mental illness in many cultures, underdiagnosis due to cultural and linguistic barriers, less access to health insurance and a regular source of care, and the use of traditional healers or health providers before seeking Western mental health services.

Chronic Diseases

Studies have shown that as immigrants reside in the United States for longer, they have an increasing prevalence of being overweight and of obesity. As a result, their risk of chronic diseases, such as diabetes, cardiovascular disease, arthritis, and certain types of cancers, increases. Immigrants and refugees from certain racial or ethnic backgrounds may also have a genetic predisposition for diabetes and heart disease, even though they are less likely to be overweight than are those born in the United States. There is a relative paucity of health programs aimed at prevention of chronic disease in immigrants and refugees. Programs aimed at immigrants and refugees should address how acculturation and other transitions associated with immigration may affect nutrition, physical activity, smoking, and other health behaviors linked to chronic diseases.

Access to Health Care

The ability of immigrants and refugees to access health care services varies widely, depending on immigration status, country of origin, and ability to navigate linguistic and cultural barriers. For example, refugees are automatically granted Medicaid or Refugee Medical Assistance for up to 8 months as part of their asylum, but the remainder of the immigrant population is not. Immigrants are less likely to have public or private health insurance than are those born

in the United States and are twice as likely to be uninsured as U.S.-born persons. Uninsurance rates among immigrants vary depending on country of origin, type of employment, and salary. Noncitizen immigrants are also less likely to have Medicaid. In 1996, Congress barred legal immigrants who entered the United States after August 1996 from enrolling in federally funded Medicaid or Medicare for at least 5 years after entry. Undocumented immigrants can only get emergency Medicaid.

Immigrants and refugees face other barriers to accessing and using health care in addition to health insurance. These include linguistic and cultural barriers that prevent effective communication. Up to 30% of immigrant households are linguistically isolated, meaning there is no adult in the household who is proficient in English. Cultural barriers also exist, with some immigrants expressing reluctance to tell health care providers about traditional practices or traditional medication use. Health care systems can address some of these barriers by recruiting bilingual and bicultural personnel, providing professional interpreter services, and increasing the cultural awareness of clinicians.

—Namratha Kandula

See also Acculturation; Asian American/Pacific Islander Health Issues; Hepatitis; Latino Health Issues; Migrant Studies; Tuberculosis; Vaccination

Further Readings

Abraido-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The Latino mortality paradox: A test of the "salmon bias" and healthy migrant hypotheses. *American Journal of Public Health*, *89*, 1543–1548.

Barnett, E. D. (2004). Infectious disease screening for refugees resettled in the United States. *Clinical Infectious Diseases*, *39*, 833–841.

Carrasquillo, O., Carrasquillo, A. I., & Shea, S. (2000). Health insurance coverage of immigrants living in the United States: Differences by citizenship status and country of origin. *American Journal of Public Health*, *90*, 917–923.

Goel, M. S., McCarthy, E. P., Phillips, R. S., & Wee, C. C. (2004). Obesity among US immigrant subgroups by duration of residence. *Journal of the American Medical Association*, *292*, 2860–2867.

Hollifield, M., Warner, T. D., Lian, N., Krakow, B., Jenkins, J. H., Kesler, J., et al. (2002). Measuring trauma and health status in refugees: A critical review. *Journal of the American Medical Association*, *288*, 611–621.

Kandula, N. R., Kersey, M., & Lurie, N. (2004b). Assuring the health of immigrants: What the leading health indicators tell us. *Annual Review of Public Health*, *25*, 357–376.

Kasl, S. V., & Berkman, L. (1983). Health consequences of the experience of migration. *Annual Review of Public Health*, *4*, 69–90.

Lara, M., Gamboa, C., Kahramanian, M. I., Morales, L. S., & Bautista, D. E. (2005). Acculturation and Latino health in the United States: A review of the literature and its sociopolitical context. *Annual Review of Public Health*, *26*, 367–397.

Lucas, J. W., Barr-Anderson, D. J., & Kington, R. S. (2005). Health status of non-Hispanic U.S.-born and foreign-born black and white persons: United States, 1992–95. *Vital & Health Statistics, 10,* 1–20.

Marshall, G. N., Schell, T. L., Elliott, M. N., Berthold, S. M., & Chun, C. A. (2005). Mental health of Cambodian refugees 2 decades after resettlement in the United States. *Journal of the American Medical Association*, *294*, 571–579.

Ngo-Metzger, Q., Massagli, M. P., Clarridge, B. R., Manocchia, M., Davis, R. B., Iezzoni, L. I., et al. (2003). Linguistic and cultural barriers to care: Perspectives of Chinese and Vietnamese immigrants. *Journal of General Internal Medicine, 18*, 44–52.

Singh, G. K., & Siahpush, M. (2001). All-cause and cause-specific mortality of immigrants and native born in the United States. *American Journal of Public Health*, *91*, 392–399.

Taylor, Z., Nolan, C. M., & Blumberg, H. M. (2005). Controlling tuberculosis in the United States. Recommendations from the American Thoracic Society, CDC, and the Infectious Diseases Society of America. *MMWR Recommendations and Reports*, *54*, 1–81.

U.S. Department of Homeland Security. (2006). *2004 Yearbook of immigration statistics*. Washington, DC: Office of Immigration Statistics, U.S. Department of Homeland Security.

Incidence

In epidemiology, *incidence* refers to new cases of disease or new occurrences of medical conditions such as becoming infected with a virus; it is contrasted with *prevalence*, which includes both new and existing cases or occurrences. While historically the term *incidence* was limited to disease or death, it is increasingly being used more broadly to quantify the occurrence of events, not just a new disease. Examples of incident cases or events include when a person develops diabetes, becomes infected with HIV, starts

smoking, or is admitted to the hospital. In each of these situations, individuals go from an occurrence-free state to having the occurrence.

It is important for epidemiologists to make a clear distinction between incidence and prevalence. While incidence refers to a new disease or event, prevalence means an existing disease or events. The following example will help refine the distinction between incidence and prevalence. A person who is newly diagnosed with diabetes has an incident case of diabetes; a person who has had diabetes for 10 years has a prevalent case of diabetes. Prevalence includes incident cases or events as well as existing cases or events. For chronic diseases, a person can have an incident case just once in his or her lifetime. For diseases and occurrences that can be fully resolved, a person can have multiple incident cases of a disease (e.g., common cold).

The study of incident cases informs us about the etiology (or cause) of a disease and its outcome. In research, the study of incident cases allows the epidemiologist to determine the risk factors for a disease or another event. The study of prevalent cases combines the study of new and surviving cases, making it unclear as to if risk factors are causes of a new disease or causes of survival on getting a disease.

To compute incidence, three elements must be defined: (1) the number of new cases of disease or occurrence, (2) a population at risk, and (3) the passage of time. Incidence can be measured as a proportion or a rate. Measured as a proportion, incidence quantifies the *risk* of an occurrence. Measured as a rate, incidence quantifies the speed of a disease or occurrence in a population.

Incidence proportion measures the probability that a person will develop a disease or condition within a given period of time. Accurate measurement of incidence proportion requires that all the individuals "at risk" for the outcome under study be followed during the entire study period (or until getting the disease or event). Because complete follow-up is required to directly compute incidence proportion, it is usually only calculated for studies with a short follow-up period. For incidence proportion, the numerator is the number of new cases of a disease during a given time period. The denominator is the total population at risk during the defined study period.

> *Example:* On a recent 7-day cruise, 84 of 2,318 passengers reported to the ship's infirmary with gastrointestinal illness. The incidence of disease on this ship equals 84 new cases of illness divided by 2,318 total passengers at risk, resulting in an incidence proportion of 4% during a 7-day period.

The incidence rate numerator is likewise the number of new cases. The denominator, however, is the total person-time of observation at risk for the disease or occurrence.

> *Example:* The incidence rate of breast cancer among women of age 40 years or more equals 32 women with breast cancer divided by 3,896 person-years of follow-up at risk or 821 per 100,000 person-years at risk.

An accurate measure of incidence, whether incidence proportion or incidence rate, requires a precise definition of the denominator. Because incidence is a measure of new cases during a given time period, it is important that those in the denominator be *at risk.* They should not have a history of the disease in question if a chronic disease; nor should they otherwise not be able to develop a new case of disease (e.g., women cannot get prostate cancer).

—Allison Krug and Louise-Anne McNutt

See also Prevalence; Proportion; Rate

Further Readings

Hennekens, D. H., & Buring, J. E. (1987). *Epidemiology in medicine.* Boston: Little, Brown.

Last, J. M. (2001). *A dictionary of epidemiology*. New York: Oxford University Press.

Independent Variable

See Dependent and Independent Variables

Indirect Standardization

Rates (incidence, mortality, etc.) calculated from different study groups or populations are often not directly comparable if the groups differ with regard to the distribution of some characteristic associated with

the outcome of interest. Adjustment of rates to facilitate comparison across populations or for the same population across time is called *standardization* and involves adjusting the crude rates "as if" they were calculated based on the same underlying population. The crude rates are "adjusted" for the characteristic on which the groups differ. For example, if the mortality rates of two different communities are to be compared, but the age distributions of the two communities are sufficiently dissimilar, comparison of the crude rates is inappropriate and will likely yield inaccurate conclusions.

There are two methods of standardization, direct (discussed elsewhere) and indirect. Indirect standardization is most often applied when the specific rates for the study population are not calculable, as is required for direct standardization, or when some strata of the study group are so sparsely populated that stable rates are not determinable.

Indirect standardization accomplishes adjustment by calculating the rates for a reference population and then applying weights (as person-years of follow-up) derived from the study population to the rates calculated from the reference population. Thus, the expected number of events in the study group is estimated from the observed number of events in the reference population. The resulting ratio of observed events to expected events is the adjusted rate ratio that compares the study group with the reference group. The standard mortality ratio (*SMR*) is an example of the use of indirect standardization to compare the mortality rate of a study group, an occupational cohort, for example, with that of a reference population.

Ideally, the chosen reference population would be as closely representative of the study population as is feasible. When two different study groups are indirectly standardized to the same reference population, the adjusted rates are directly comparable. When a single study group is compared with a single reference group, the direct and indirect methods of standardization are equivalent.

In general, presenting standardized rates instead of rates specific to a particular study group is advantageous in several ways. First, and perhaps most important, as discussed in the foregoing, the standardizing of rates in effect adjusts for potential confounding by those factors on which the rates are standardized, such as age. A standardized rate is effectively a summary measure that is then easier to compare with other similar summary measures than are unstandardized rates. Statistically speaking, standardized rates have a smaller standard error than do unstandardized rates. Standardized estimates are more stable and less influenced by small cell sizes or sparsely populated strata. Finally, specific rates may be unavailable for certain groups of interest, so indirect standardization is the method for making estimates about such groups.

Standardization is not without its disadvantages, however. In instances where there is effect modification, a standardized rate, because it is in essence an averaging of the specific rates, will tend to obscure any differences between strata. In addition, the magnitude of the standardized rate depends on the chosen standard population and is thus arbitrary. The standardized rate itself doesn't necessarily yield useful information, but the difference or ratio between rates is important.

—Annette L. Adams

See also Confounding; Direct Standardization; Effect Modification and Interaction; Mortality Rates

Further Readings

Hennekens, C. H., & Buring, J. E. (1987). *Epidemiology in medicine*. Boston: Little, Brown.

Koepsell, T. D., & Weiss, N. S. (2003). *Epidemiologic methods: Studying the occurrence of illness*. New York: Oxford University Press.

Szklo, M., & Nieto, F. J. (2000). *Epidemiology: Beyond the basics*. Gaithersburg, MD: Aspen.

Teutsch, S. M., & Churchill, R. E. (2000). *Principles and practice of public health surveillance* (2nd ed.). New York: Oxford University Press.

Inferential and Descriptive Statistics

Statistics is a branch of applied mathematics used to inform scientific decision-making in the absence of complete information about phenomena of interest. The application of statistics is integral to the theory and practice of epidemiology because it allows an investigator to both describe characteristics of exposure and disease in targeted populations and make logical inferences about these populations based on samples of observations. As we will discuss, *descriptive statistics* are estimates used to characterize or

describe the nature of a population in terms of measured variables, while *inferential statistics* are used to answer questions by testing specific hypotheses. This entry provides a general overview of important statistical concepts, distinguishes the categories of descriptive and inferential statistics, and describes how both descriptive and inferential statistics can inform scientific inquiry.

Fundamental Concepts of Statistics

To understand different statistical techniques discussed in this chapter, a brief overview of key concepts is necessary.

Statistic

A statistic is a quantitative estimate of a population characteristic that is based on a sample of observations taken from that population. In many areas of scientific inquiry, it is difficult or impossible, due to time and resource constraints, to observe or survey the entire universe or *target population* of interest. Fortunately, statisticians have shown that with properly conducted *random sampling*, valid and suitable estimates (known as *statistics*) of population values (known as *parameters*) can serve as effective substitutes.

This holds true in large part because under conditions of well-formulated research design and random sampling, mathematical principles of probability can accurately estimate the probable degree of imprecision, or sampling error, around statistical estimates of population parameters. By estimating this degree of imprecision accurately, one can know how well a statistic may capture a characteristic of the population it targets.

Random Sampling

The importance of *random sampling* to the value and utility of statistical analysis cannot be understated. While a complete discussion of *random sampling* and its variants is beyond the scope of this chapter, put simply, random sampling implies that every member of the population of interest has an equal probability of being included in the sample of measurements. To the extent that this assumption is met in the process of data collection, statistical estimates will have desirable statistical properties. To the extent that this assumption is violated, bias is introduced into resulting statistics that may limit or completely invalidate the degree to which the statistics derived from the sample reflect the population parameters under investigation.

Random Variables

Participant characteristics measured in research studies such as patient gender, age, income, presence or absence of a disease, or disease stage are also known as *random variables*. A *random variable* is an observable phenomenon with a definable, exhaustive set of possible values. To understand a random variable, one needs to understand its associated level of measurement. There are essentially two types of random variables:

1. Qualitative random variables, which take on discrete, categorical values and include those that are nominally measured (i.e., exposed vs. not) and those that are ordinal measured (disease stage—Cancer I to IV);
2. Quantitative random variables (i.e., age, income), which take on values that are measured on a continuous and constant incremental scale. Patient age, for example, generally ranges between 0 and 100 years.

Frequency and Probability Distribution

Random variables take on measurable values with an observable frequency relative to the total number of observed elements. This relative frequency constitutes the probability of observing that value in the sample and is an estimate of the probability of that value in the population. Coin flips, for example, have two possible values (heads and tails) that occur with equal relative frequency (i.e., each with a probability of .5). The assortment of the relative frequencies of the possible values of a random variable is known as a probability distribution. Under conditions of random sampling, the probability distribution of a sample of observed values of a random variable has some important properties:

1. It will, on average, reflect the population probability distribution from which they were drawn.
2. Sampling error, or the degree to which sample statistics can be expected to vary due to the size and type of sample, is quantifiable and thus can be accounted for in analyses.

3. As sample sizes get larger and approach the population, the sample estimates converge on (or more closely approximate) population values.

Description and Inference

Description and inference are distinct but related goals that drive epidemiologic research. Description is driven by a more open-ended desire to explore, understand, and characterize observable phenomena, while inference is more narrowly related to answering specific questions or *testing hypotheses*. To illustrate, a descriptive aim in survey research might be to characterize the political preferences of registered voters in New York City, while an inferential aim might be to determine whether African Americans are more likely than Caucasian voters to declare themselves as Democrats. While these two goals can likely be accomplished in the same research study, they require different statistical methods.

Descriptive Statistics

Good statistical practice begins with attempts to gain a basic understanding of the random variables under study. Descriptive statistics achieve this by characterizing or summarizing aspects of random variables and thereby inform the user about (1) the specific nature of his or her sampled observations and (2) likely values of the same characteristics from the population from which they were drawn. Descriptive measures vary in form and use depending on the type of random variable the user wishes to characterize and can be classified as measuring central tendency, dispersion, and extremity. Specific calculations may also vary depending on whether estimation is being made on a sample versus when all or nearly all of the population is measured.

Measures of central tendency are probably the most commonly cited sample descriptive statistics and are designed to estimate the "expected" value of a random variable. The most commonly used *central tendency* measures are the mean (or average) for quantitative variables, median (the 50th percentile) for quantitative or ordinal variables, and the mode (most frequent value—usually used for nominal variables). The proportion, which commonly appears in epidemiologic studies labeled as a risk or prevalence estimate, is a measure of central tendency that is mathematically equivalent to the mean of a dichotomously measured random variable with measured values of 0 and 1.

Measures of dispersion express the variation or "spread" of the values of the random variable among subjects sampled. The most commonly cited dispersion measure is the standard deviation, which is measured as the square root of the average* of the squared deviations from the mean (*1 is subtracted from the denominator used when computing the average when the mean is also an estimate vs. if the mean of the target population is known). Other key measures of variability include the range (difference between the minimum and maximum values), the interquartile range (difference between the 25th and 75th percentiles), and the coefficient of variation. As discussed in the following, these measures play a very important role in inferential statistics.

Measures of extremity attempt to identify values that differ extremely from other observed values (i.e., outliers) and are primarily applicable to quantitative and discrete ordinal variables with many values. The most common outlier measures include (1) values more than 3 standard deviations from the mean and (2) the values beyond the fifth and 95th percentiles of the ranked distribution. Because outliers can heavily influence central tendency and dispersion measures, measures of extremity provide valuable information to put these measures in proper context and possibly identify key population subgroups. Extremity measures should always be examined prior to employing any inferential statistics as well.

Inferential Statistics

Progress in science is driven in part by *hypotheses*, or educated guesses, often derived from *theories*, about the nature of real-world phenomena and relationships between multiple variables. Inferential statistics are used to test these hypotheses about relationships and differences between key measures of random variables taken on research study samples.

Ideally, research studies with inferential or analytical aims are designed so that observed variation can be isolated to differences on one or more factors of interest. In fact, the quality of a study is often judged by the degree to which it achieves this goal. To the degree that a study design attends to these matters, inferential statistics examine whether any differences found in the study are likely to reflect "true" differences in the factors studied or sampling error or random chance.

The process of *statistical inference* follows certain guidelines. Data are collected and compared using an

appropriately selected *inferential test statistic* that estimates the probability that an observed *difference*, *ratio*, or *association* between two or more descriptive statistics could occur by chance alone (i.e., could be due to sampling error) if the samples were drawn from the same population. Using an a priori decision rule about what constitutes an implausible probability that this assumption is true, typically a probability of .05, the inferential test statistic is used to determine whether to either accept the hypothesis that samples are drawn from the same probability distribution (i.e., accept the null hypothesis) or to reject that hypothesis in lieu of the alternative that there is a difference or association (i.e., reject the null hypothesis). In short, when the correct experimental conditions are met, inferential statistics provide a probabilistic basis for judging the verity of a research hypothesis.

The most commonly used inferential statistics include the independent samples *t* test, the chi-square test of independence, Pearson's correlation, and one-way analysis of variance. Each of these tests is designed to yield a valid test statistic and null-hypothesis probability value under specific conditions. The independent samples *t* test, for instance, tests whether the observed means for two independent samples were drawn from the same probability distribution. It is important to understand that this test assumes that the outcome being compared is quantitative and that the observed data in each group are randomly and independently sampled, are normally distributed in the population, and share a common standard deviation. If these conditions are not met, the test statistic and probability value may not be valid, depending on the degree and type of departure from these assumptions.

Over the years, statisticians have developed hundreds of inferential statistics, each with specific types of hypotheses and "data situations" in mind. Rather than try to enumerate the different types of tests, the point we convey here is that while inferential statistics may vary in their computation, assumptions, and applicability, their primary purpose is for hypothesis testing and inference.

A Caution

It is critically important to understand that even the most robust descriptive and inferential statistics cannot overcome the limitations of poor research and analytic design. Probability values from test statistics are limited first and foremost by the degree to which a selected research design rules out alternative explanations and yields unbiased measurements from the population(s) of interest. In addition, correct application of inferential statistics assumes that a test is chosen that correctly fits the measurement level and characteristics of the data at hand as well as that all assumptions for the test are met. Thorough descriptive analysis helps the analyst check these assumptions and thus forms the foundation of good statistical practice. However, with this caution in mind, applying descriptive and inferential statistics correctly provides researchers and epidemiologists the valuable opportunity to study and learn about important research problems in the absence of complete information.

—Brian M. Waterman

See also Bias; Descriptive and Analytic Epidemiology; Hypothesis Testing; Measures of Association; Measures of Central Tendency; Measures of Variability; Sampling Distribution; Sampling Techniques; Study Design; Target Population; Validity

Further Readings

Hays, W. L. (1994). *Statistics* (5th ed.). Orlando, FL: Harcourt Brace.

Kahn, H. A., & Sempos, C. T. (1989). *Statistical methods in epidemiology*. New York: Oxford University Press.

Mohr, L. B. (1990). *Understanding significance testing*. Quantitative Applications in the Social Sciences, No. 73. Newbury Park, CA: Sage.

Weisberg, H. F. (1992). *Central tendency and variability*. Quantitative Applications in the Social Sciences, No. 83. Newbury Park, CA: Sage.

Influenza

Influenza, also known as the "flu," is a contagious disease caused by infection with the influenza virus. It is a common disease, with annual attack rates of 10% to 30% worldwide each year, with most cases in the northern hemisphere occurring during the "flu season" of December through March. Flu is characterized by fever, by respiratory symptoms, including rhinorrhea, cough, and sore throat, and sometimes by myalgia and headache. In children and infants, gastrointestinal symptoms such as vomiting and diarrhea occur in 50% of cases, but such symptoms are rare in adult cases. The course of influenza is usually self-limiting (i.e., resolves without medical intervention) and lasts

3 to 5 days, but serious complications, including pneumonia, may develop that prolong the illness and may prove fatal.

Influenza is spread through the respiratory secretions of infected persons, primarily through airborne secretions spread by coughing, sneezing, and talking or by direct (e.g., kissing) or secondary contact (e.g., touching a surface touched by an infected person and then touching one's nose). Influenza has an incubation period of 1 to 4 days, and infected persons can transmit the virus from 1 day before the onset of illness through the fourth or fifth day of infection. The most common strains of influenza are Types A and B, which cause the annual epidemics and against which the flu vaccine offers protection; Type C causes mild illness, is not implicated in the annual epidemics, and is not included in the flu vaccine. Strains of Type A influenza virus are classified by two proteins found on the surface of the virus, hemagglutinin (H), and neuraminidase (N); Type B virus is not divided into subtypes. New strains of flu are constantly evolving, and the need to identify each new strain led to development of the standard five-part nomenclature, which identifies the virus type, the site of first identification, the strain number, the year of isolation, and the subtype (for Type A virus). For instance, "A California/7/200(H3N2)" refers to a Type A virus first isolated in California in 2004 as Laboratory Strain 7, with subtype H3N2. Popular names of flu viruses generally refer to the geographical region where the outbreak began or was first reported, such as the "Spanish Flu" or the "Hong Kong Flu," and strains are often referred to by their subtype as well, for instance the H5N1 strain of avian flu.

Although there have been no flu pandemics since the 1960s, flu is still a serious health concern that annually causes many cases of disease and death. The Centers for Disease Control and Prevention (CDC) reports that 5% to 20% of the United States population contracts the flu in an average year and that more than 200,000 are hospitalized and more than 35,000 will die from complications of the flu. Assessing the burden of the flu worldwide is more difficult, but the World Health Organization (WHO) estimates that 3 to 5 million cases of severe illness are caused by the flu each year and between 250,000 and 500,000 deaths.

History

A disease resembling human influenza was described by Hippocrates in the fifth century BCE, and pandemics of febrile respiratory diseases have been recorded regularly since that time. The development of a typical influenza pandemic is the same today as it was in ancient times: it begins in a specific geographic area and is spread along common trade and transportation routes, with high attack rates in all age groups and substantial numbers of hospitalizations and deaths.

The most severe flu pandemic was the Spanish Flu pandemic, which occurred in 1918–1919 and was caused by an unusually severe strain of Type A virus. This virus is estimated to have had an attack rate of 20% to 30% and a case fatality rate of 15% to 50% in adults, and an attack rate of 30% to 45% in children. It is estimated that at least 20 to 50 million persons died in the first 12 months of the pandemic, more deaths globally than caused by any disease since the bubonic plague of the 14th century, and this is almost certainly an underestimation due to the underreporting from Africa and Asia. The Spanish Flu pandemic was also unusual in that many of the dead were young adults, in distinction to the usual pattern in which elderly people are at greater risk of complications and death from the flu. It was called Spanish Flu because Spain, a neutral country during World War I, was the first to report the disease; the United States and other European countries involved in the war did not report the epidemic due to wartime censorship.

The Spanish Flu pandemic began with illness reported among soldiers in the United States, which spread to Europe as soldiers were dispatched to serve in World War I. The number of U.S. troop deaths due to influenza and pneumonia, 43,000, eventually rivaled the 54,000 deaths due to battle. By May 1918, Spanish Flu was reported in Africa, and in India and China by August 1918. The flu's effect was particularly severe among people living on isolated islands: For instance, it is estimated that 20% of the population of Western Samoa died of the Spanish Flu within 2 months after introduction of the virus. A second wave of Spanish Flu swept the United States beginning in August 1918, causing severe illness and high death rates; some authorities believe the virus mutated while in Europe, explaining why this second wave of infection in the United States was much more deadly than the first. There were no effective measures against the flu at this time, although many preventive measures were instituted, including the use of face masks and camphor necklaces and prohibitions against public spitting, none of which are currently regarded as effective.

The last major flu pandemic was caused by the Hong Kong Flu, in 1968–1969, which may have caused as many as 1 million deaths worldwide. The greater availability of supportive health care and use of antibiotics to control secondary infections, plus the advent of the flu vaccine, have probably helped reduce mortality from the flu.

Prevention and Control

The primary means of preventing spread of the flu is by ordinary health habits such as covering your mouth and nose when coughing and washing your hands frequently, and through widespread use of the flu vaccine. The current recommendation by the CDC is that most people can benefit from getting an annual flu vaccine. Persons who should not receive the vaccine include children below 6 months of age, persons allergic to chicken eggs (because the virus used for vaccination is grown on embryonated eggs), persons who have had a severe reaction to vaccination in the past, and persons who have developed Guillain-Barré Syndrome, an inflammatory disorder of the peripheral nerves, within 6 weeks of a previous vaccination. Persons considered to be at high risk of complications from the flu, and who are therefore particularly recommended to get an annual vaccine, include children aged 6 months to 5 years, pregnant women, people above the age of 50 or with certain chronic medical conditions, people who live in nursing homes or other long-term care facilities, health care workers, and people who live in households with, or care for, people in the high-risk groups.

Because the influenza virus mutates regularly and exists in many strains, a new flu vaccine must be developed annually. This process requires scientists to forecast which strains will be circulating in the upcoming flu season. The quality of the match between the vaccine and the strains of flu circulating is a major factor in the effectiveness of the vaccine: the CDC reports that the vaccine is 70% to 90% effective in preventing influenza in healthy people below age 65 when the match between vaccine and circulating strains is close. The flu vaccine may be administered as a "flu shot" containing inactivated flu vaccine or as a nasal-spray vaccine containing live attenuated flu virus. Each vaccine contains two Type A and one Type B influenza viruses. The flu shot is recommended for anyone above 6 months of age, while the nasal-spray vaccine is only approved for use in persons aged 5 to 49 years who are not pregnant.

Four antiviral drugs have been approved for use in the United States against the flu: amantadine, rimantadine, zanamavir, and oseltamivir (TamiFlu). All have been approved for both prophylactic use and treatment of existing cases, although in the latter case they must be taken within 2 days of illness onset. All have been effective in the past against Type A virus, but only oseltamivir and zanamivir have proven effective against Type B viruses. However, due to evidence that many influenza A viruses circulating in the United States have become resistant to amantadine and rimantadine, the CDC issued a recommendation in July 2006 that neither drug be used for treatment or prevention of Type A influenza in the United States during the 2006 to 2007 flu season.

The possibility that a new, deadly strain of the flu will appear that can be transmitted among humans is a serious public health concern. This scenario has been proposed with the H5N1 strain of avian flu, which was identified in Asia in 1997 and can infect both birds and humans. Although 225 confirmed human cases were reported to the WHO between 2003 and 2006, mostly from Asian countries, the disease is not yet capable of causing a pandemic because it is not efficiently transmitted between humans: Most cases are due to transmission from bird to human. However, if the N5N1 strain should mutate to a form that could be efficiently transmitted between humans, it could cause a worldwide pandemic because most humans would not have resistance to the new strain, and existing vaccines would probably not be effective against it. In addition, current stocks of antivirals are not sufficient to treat the number of cases anticipated from a novel and severe flu strain: The CDC estimates that the United States might have as many as 200 million cases, with as many as 800,000 hospitalizations and 300,000 deaths, in the first 3- to 4-month period of the disease. Efforts made to prepare for such a pandemic include stockpiling of antiviral drugs, development of new drug distribution systems, improved surveillance and monitoring for the emergence of new viruses, and creation of priority lists specifying who should receive vaccines, antivirals, and hospital care if supplies are insufficient to treat everyone.

—Sarah Boslaugh

See also Avian Flu; Epidemic; Public Health Surveillance; Vaccination

Further Readings

Barry, J. M. (2004). *The great influenza: The epic story of the deadliest plague in history*. New York: Viking.

Centers for Disease Control and Prevention. (2006). *Key facts about influenza and the influenza vaccine*. Retrieved January 10, 2007, from http://www.cdc.gov/flu/keyfacts.htm.

Institute of Medicine. (2005). *The threat of pandemic influenza: Are we ready?* Washington, DC: National Academies Press.

Kolata, G. (1999). *Flu: The story of the great influenza pandemic of 1918 and the search for the virus that caused it*. New York: Farrar, Straus & Giroux.

Nelson, K. E., & Williams, C. M. (2007). *Infectious disease epidemiology: Theory and practice* (2nd ed.). Boston: Jones & Bartlett.

Smith, N. M., Bresee, J. S., Shay, D. K., Uyeki, T. M., Cox, N. J., & Strikas, R. A. (2006). Prevention and control of influenza: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *Morbidity and Mortality Weekly Report*, *55*(RR10), 1–42. Retrieved January 10, 2007, from http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5510a1.htm.

World Health Organization. (2003). *Influenza: Fact sheet*. Retrieved January 10, 2007, from http://www.who.int/mediacentre/factsheets/fs211/en.

Informed Consent

Informed consent is at the heart of the ethical conduct of epidemiologic research. In 1974, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was created as a part of the National Research Act. This commission was developed in response to atrocities in the Tuskegee Syphilis Study and other research projects. Specific concerns focused on including individuals in research without their knowledge and, in the case of the Tuskegee study, provision of substandard medical care for the perceived benefit of the research project. The Belmont Report, developed by the Commission in 1978, laid the framework for the ethical conduct of research in the United States. Similar documents have been developed elsewhere, most notably the Declaration of Helsinki. All such documents require potential study participants to provide informed and voluntary consent to participate in research.

Informed consent is the *process* of providing information to potential study participants about their rights and the study's goals, procedures, and risks. A written informed consent form is provided to potential study participants in most epidemiologic research and is a key component of the informed consent process; however, informed consent goes beyond this form and is woven throughout all interactions with the participant, from recruitment to study completion and beyond.

Informed Consent Form

In the United States, the informed consent forms used in epidemiologic research include, at a minimum, the following eight elements:

1. information about the purpose of the research and the study procedures (highlighting any components that are experimental),
2. potential risks of the research for participants,
3. potential benefits of the research for participants and for others,
4. information about alternative treatment available (if it exists),
5. information about the level of confidentiality that will or will not occur,
6. information about compensation for injury if more than minimal risk exists,
7. information about who to contact to learn more about the research or to contact if a participant thinks his or her rights have been violated or he or she has been harmed, and
8. a clear statement that participation in the research is voluntary, that the choice of participation will not affect care otherwise provided at the study site, and that the participant can quit at any time.

For studies that involve special circumstances, such as criteria that must be met for continued participation, additional information in the consent process is required.

The language level and format must optimize readability and understanding for potential study participants. Typical recommended reading levels range from fifth grade to eighth grade, depending on the educational level of the population. Innovative formats, such as question and answer, are currently in favor because they enhance reader comprehension.

In addition to providing detailed information, researchers must ensure that the potential participants understand the study procedures. This is particularly true for studies involving risk to the participants. Simply signing the form is not sufficient for meeting this

standard. True consent is only granted by individuals who have autonomy in their decision.

In most research, individuals must sign the informed consent form to participate in the study. The purpose of the signed informed consent form is to ensure that participants have seen it, read it, and understood the content. However, it is important that the researchers assess understanding. A signed informed consent form is not sufficient for meeting the standard. For individuals to provide consent, they must be informed and have autonomy in their decision. Depending on the risks involved in the research, different levels of assessing understanding of procedures and rights are used. In high-risk studies, often multiple methods are used to provide information to participants, including video information about the study, conversation with research assistants well trained in the informed consent process, and written materials. For research involving significant risk, such as randomized trials of experimental medications, the standard for understanding is such that potential participants are quizzed and only after passing a test on the information contained in the informed consent may the individual choose to participate and sign the form.

In special circumstances, researchers can ask for a waiver of the signed informed consent form. This is only permitted in research involving minimal risk when a signature requirement might preclude the study (e.g., random digit dial telephone survey of health behaviors), or the signature is the only identifier collected (e.g., anonymous survey). While informed consent is still required, a signed form may be waived.

Informed Consent Process

Researchers are responsible for ensuring that participants are fully informed and provide voluntary, autonomous consent. Several strategies may be employed to ensure that the informed consent process is conducive to autonomous participation: Create an environment where the participant feels in control of decision-making (e.g., separate the potential participant from the pressure of friends or family), and provide reassurance that it is acceptable to decline participation if hesitation is evident. Additionally, researchers must inform participants about procedures and their rights throughout the study and provide an atmosphere in which participation continues to be voluntary.

—Louise-Anne McNutt

See also Cultural Sensitivity; Ethics in Health Care; Ethics in Human Subjects Research; Tuskegee Study

Further Readings

Office for Human Research Protections, Department of Health and Human Services. (1998, September 30). *Informed consent checklist*. Retrieved August 8, 2006, from http://www.hhs.gov/ohrp/humansubjects/assurance/consentckls.htm.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979, April 18). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Retrieved August 8, 2007, from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm.

Injury Epidemiology

Injury epidemiology involves characterization of occurrences of injuries, identification of risk factors, and the evaluation of prevention programs for injuries. Although the term *injury* could refer to almost any adverse health event, this term is generally used to refer to damage to human tissue resulting from exposure to energy delivered in excess of the threshold that human biological systems can tolerate. This excess transfer can occur during exposure to various forms of energy, including mechanical, chemical, electrical, and ionizing and nonionizing radiation. Injuries may also occur as a result of errant medical interventions and, in the case of strangulation or drowning, could result from the lack of an element vital to the body (i.e., oxygen). Injuries are often classified as either intentional (homicide, suicide, violence) or nonintentional (falls, motor vehicle crashes, burns, lacerations, strains, and cumulative trauma). Injury epidemiology is applied in various diverse environments such as the workplace, home, transportation, and sports and recreational settings.

The application of systematic epidemiologic methods in the analysis of injury events is a relatively recent phenomenon, with its origins in the early 1960s. Despite being a rather new field, injury epidemiology research has led to the development of numerous interventions (e.g., seat belts, bicycle and motorcycle helmets, ergonomic design improvements, workplace safety programs, automobile safety design, and transportation regulations) that have achieved

significant public health improvements. Because one of the largest public health impacts of natural and man-made disasters (hurricanes, tornados, floods, earthquakes, volcanoes, drought, famine, war, terrorism) is injury, the field of disaster epidemiology involves the application of injury epidemiology methods as well as evaluation of infectious disease and mental health conditions.

Injuries have historically been referred to by terms such as *accidents*, *mishaps*, *casualties*, *acts of God*, or other terms that imply a random nature or unavoidability. In fact, most injuries show clear nonrandom patterns that can be characterized epidemiologically and have identifiable risk factors. In addition to prevention, research efforts have identified interventions that were subsequently implemented to reduce the consequences of injuries once they have occurred (e.g., seat belts, airbags).

The field of injury epidemiology observes a distinction, similar to that in medicine, between "acute" injury and chronic injury, repetitive injury, or cumulative trauma. There is currently a debate in the field as to whether terms such as *acute injury* and *chronic injury* should refer to the time period for the delivery of exposure or the duration of the injury. Also, in some circumstances, cumulative trauma has been classified as a disease rather than an injury. However, when injury is defined as the result of excess energy transfer beyond a tolerance threshold (whether this is delivered suddenly or over a long period of time), logically cumulative trauma should be considered as a type of injury. Until a standard terminology is adopted for injury labels such as *acute* or *chronic*, epidemiologists should clearly define the meaning of their injury terminology to avoid confusion and to assure better comparability of results across different studies.

Public Health Burden of Injuries

Epidemiologic research on injuries receives much less public attention than research into cancer or cardiovascular diseases. However, the public health impact of injuries is similar to that of cancer and heart disease, depending on how they are measured and which sectors of the population are considered. Injuries are the fifth leading cause of death in the United States after heart disease, cancer, and cerebrovascular and pulmonary diseases. Worldwide, road traffic injuries rank as the ninth leading cause of death following these four diseases and HIV/AIDS, diarrheal diseases, and other infectious diseases. This worldwide ranking for motor vehicle injury is expected to rise to number three by the year 2020. However, when ranked by years of life lost, intentional and nonintentional injuries rank as the leading cause of premature mortality, higher than cancer and heart disease (Table 1). Young people are especially affected by injuries. For example, in the United States, injuries are the leading cause of death for those aged 1 to 34 years, homicide is the second leading cause of death for persons aged 15 to 24, and suicide is the third leading cause of death for persons aged 15 to 34. Additional injury statistics are presented in Table 1.

Data Sources, Study Design, and Theoretical Framework

Much of the public health research concerning injuries has used descriptive epidemiology to characterize the magnitude of injury impacts and to assess ecologic associations. This research has relied on a wide diversity of injury surveillance databases, including data from law enforcement and justice organizations, the Centers for Disease Control, Departments of Transportation, the Consumer Product Safety Commission, the Bureau of Labor Statistics, state and county health agencies, academic institutions, and other agencies and registries (Table 2). Analyses of these sources have provided numerous insights into injury occurrence and prevention. One metric often used in evaluating the impacts of injuries is a measure of the number of years of life lost due to injury or disease. This is often calculated to quantify the years of life lost before a selected age (e.g., 65 or 75 years). Because of the higher rates of injuries among younger age groups compared with those of chronic diseases, this measure will provide a different (and many argue more accurate assessment) of the public health impacts of injuries.

A key analytical challenge in descriptive injury epidemiology is obtaining accurate estimates of the population at risk. In traffic safety research, for example, accurate estimates of the size of the driving population are required to estimate automotive crash risks. A more precise measure for calculating motor vehicle crash risk is an estimate of the number of miles driven by the populations at risk for motor vehicle injury. For example, when using this type of statistic to compare crash risk among older drivers, studies have

Table 1 Selected Injury Statistics

Approximate Years of Life Lost (< 75 years old)[a] by Selected Cause of Death, United States 1995	
All injuries	2,120
Unintentional injuries	1,250
Suicide/homicide	870
Malignant neoplasms	1,800
Heart disease	1,675
HIV	540
Cerebrovascular	230
Total number of fatal injuries, United States (1997)	146,000
Fatal injuries due to motor vehicle accidents	29.1%
Fatal injuries due to firearms	22.2%
Fatal injuries due to poisoning	12.1%
Fatal injuries due to falls	8.6%
Fatal injuries due to suffocation	7.3%
Fatal injuries due to other	20.7%
Firearm-Related Death Rates in Selected Countries[b]	
United States	55
Norway	11
Canada	10
Australia	8
Sweden	3.5
Injury-Related Morbidity Statistics	
Estimated proportion of individuals sustaining an injury in a given year	25%
Hospital discharges related to injury	8%
Emergency room visits related to injury	37%
Estimated annual costs (United States) due to injury ($ billions)	260

Source: Based on National Center for Health Statistics data, cited in MacKenzie (2000).

Notes

a. Years of life lost before age 75, per 100,000.

b. Mortality rates among males aged 15 to 24 (1992–1995) per 100,000 population.

shown older drivers actually have a higher crash risk than younger drivers. Crash rates based on population estimates as the denominator have not necessarily shown this difference. Estimates of occupational driving risks or occupational injury risks are also problematic because reliance on job titles alone can provide misleading results if the amount of driving or specific "at-risk" tasks vary across and within occupational groups.

A variety of documents are used to record injury events. These include death certificates, medical records, coroner's reports, police reports, crash reports, occupational injury reports, local and national surveys, and registry forms. Injury severity can range from minor first aid events to serious hospitalization and death. Only a small percentage of accidents result in injuries. Various coding systems exist to classify the type, severity, and anatomical location of injuries. International Classification of Disease (ICD) codes use two types of codes to further characterize injuries: N codes, which classify the nature of injury, and E codes, which classify the cause of injury. For some specific settings, such as occupational injury, these ICD codes may not be sufficient for prevention programs, and more tailored, specific classification systems are used. Various scoring systems have been developed to characterize the extent of tissue damage, long-term impairment and functionality, and the body regions affected. Examples of these scoring systems include (1) Glasgow Coma Scale (GCS), (2) Revised Trauma Score (RTS), (3) Abbreviated Injury Scale (AIS), (4) Injury Severity Score (ISS), and (5) New Injury Severity Score (NISS).

Study Design

Analytical injury epidemiologic studies focus on identifying injury risk factors or evaluating the impacts of injury prevention programs or safety regulations. The study designs typically used in injury epidemiology include cohort studies, case-control studies, cross-sectional surveys, and ecologic studies. One variation of the case-control study that has special application for injury epidemiology is the case-crossover design. In this design, the case also serves as its own control. Different periods of time are sampled, and exposure is measured during the sampled time periods and compared with exposure at the time of the injury. For example, in a study of mobile phone use while driving and the risk of a crash, mobile phone use was

Table 2 Injury Data Sources

Database Description	*Web Site*
U.S. National Center for Health Statistics	http://www.cdc.gov/nchs
U.S. National Center for Health Statistics, National Death Index	http://www.cdc.gov/nchs/r&d/ndi
U.S. National Center for Health Statistics, Web-Based Injury Statistics	http://www.cdc.gov/ncipc/wisqars/default.html
National Highway Traffic Safety Administration (NHTSA)	http://www.nhtsa.dot.gov
Fatal Analysis Reporting System (maintained by NHTSA)	
General Estimates System—sample of all types of motor vehicle crashes (maintained by NHTSA)	
Crashworthiness Data System (CDS)—detailed traffic accident investigations	http://www-nrd.nhtsa.dot.gov/departments/nrd-30/ncsa/NASS.html
Bureau of Labor Statistics	http://www.bls.gov/iif
Census of Fatal Occupational Injuries (OSHA)	
Consumer Product Safety Commission—National Electronic Injury Surveillance System (NEISS)	
Bureau of Justice Statistics—National Crime Victimization Survey	
Federal Bureau of Investigation—Uniform Crime Reports	

assessed from billing records for the time period immediately before the crash and compared with mobile phone use the previous day, the same day for the previous week, and at other similar times. The case-crossover design can be used when (1) the induction period between exposure and disease is short, (2) exposure varies over time (as opposed to being constant), (3) exposure has a "transient" effect; that is, distant exposures do not have an impact on the current injury, and (4) confounders remain constant. Injury occurrence often satisfies these four conditions, and because the case-crossover design is cost efficient and efficient in controlling for confounders, it is well suited for injury research. Case-crossover studies are typically analyzed using matched case-control design methods.

Theoretical Framework for Injury Research and Prevention

Due to his work in developing a conceptual model that unified injury research and prevention, William Haddon is frequently considered the father of modern injury epidemiology. His model or matrix, referred to as the Haddon Matrix, has application not only in injury epidemiology but also in other areas of public health research and prevention. The Haddon Matrix is based on the fundamental epidemiologic principles that characterize the interaction of host, agent, and environment to describe and address public health problems. The Haddon Matrix classifies three time periods in the injury process as the rows in the matrix: *Pre-event*, *Event*, and *Postevent.* The columns of the matrix consider the factors related to (1) the host or person injured, (2) the agent or vehicle of energy transfer, (3) the physical environment, and (4) the social environment. Characterizing the cells of this matrix by identifying injury risk factors or factors that reduce the consequences of an injury provide the framework for examining causation and prevention on multiple levels.

Special Topics

Transportation safety, violence (including domestic violence), firearms involvement in injuries (homicides,

suicides, and unintentional injuries), disaster epidemiology, and occupational safety and health are areas where there are numerous applications of injury epidemiology.

Many developments in injury epidemiology have resulted from its application to the area of traffic safety. In fact, the Haddon Matrix was first developed and applied to traffic safety research. Numerous surveillance databases (e.g., FARS, NASS GES, NASS CDS, and state motor vehicle crash databases) have been developed to monitor trends in motor vehicle safety. The combined research and policy efforts of injury epidemiology, biomechanics, engineering, emergency or acute care delivery, and regulations and law enforcement have led to dramatic improvements in traffic safety. Although the 20th century started as the so called Century of Road Death with the introduction of the automobile into our daily lives, the rates of motor vehicle related fatality have decreased significantly from the mid-1930s to the 1990s. This decrease in motor vehicle-related fatality has resulted from improvements in highway design, motor vehicle design, seat belts and children's safety seats, air bags, speed limits, and licensing restrictions. Epidemiologic evaluations of human factors (age, sex, experience, alcohol consumption, fatigue), vehicle factors (mechanical failure, design), and environmental factors (road conditions, traffic, weather) have helped lead to traffic safety improvements; however, considerable work remains to be done in this field. For example, the rates of mortality due to motorcycle crashes are tenfold higher than the rates due to car and truck crashes. Another potential area for research is the evaluation of crash risk associated with various driver distractions.

Violence is a subject that garners substantial media and public attention, especially on a local level, yet receives much less public attention on an epidemiologic scale. As defined by the National Research Council, violence is "behavior by individuals that intentionally threatens, attempts, or inflects physical harm on others." Violence includes homicide, robbery, rape, assault, firearm use, suicide, and domestic violence against a partner, child, or elderly person. Violence is more prevalent in certain age and race groups and in certain geographic areas. Domestic violence takes the form of physical and sexual abuse as well as neglect. Underreporting of domestic violence affects our full understanding of both victims and characteristics or risk factors of the perpetrators of domestic violence. Rates of all types of violence also vary internationally. For example, the United States has an approximately fourfold higher rate of homicide compared with other developed countries in Western Europe and Australia/New Zealand.

In the United States, firearms-related injury deaths are the second most common type of injury death (22%) after motor vehicle crashes (29%); however, firearms make up only a small proportion (less than 1%) of all nonfatal injuries. The rates of firearms-related mortality are at least five times higher in the United States compared with Western European countries (Table 1). Firearms are involved in 68% of homicides and 57% of all suicides among men. Guns are present in approximately 40% of U.S. households, and research has shown that people who live in homes with guns are more likely to die from homicide or suicide and that in cases of domestic assault, those involving firearms are much more likely to result in death than domestic assaults in homes without guns. Although the importance of agents such as firearms and alcohol on injury risk is clear, further understanding of the interactions of the complex relationships between these factors and behavioral, environmental, and social factors is needed.

Investigation of injury occurrence associated with natural (hurricanes, earthquakes, tsunamis, heat waves, drought, tornados, and other weather events) and man-made (war, terrorism, major industrial accidents) disasters offers unique challenges to injury epidemiology. The difficult conditions under which data must be collected, the accuracy of the data, and the rapid time frame in which the information is demanded represent just a few of these challenges. The recent example of estimating Iraqi civilian deaths from the ongoing war reflects these challenges. University researchers, using a national cluster survey design, estimated nearly 600,000 violent deaths compared with 150,000 deaths estimated from the Iraqi Health Ministry based on daily estimated body counts from hospitals and morgues and approximately 63,000 based on tabulations of news media reports. This wide range of civilian fatality estimates highlights the uncertainties involved with these types of public health inquiries. Wide international disparities on the impacts of natural disasters provide an indication of the importance of socioeconomic conditions and emphasize the need for evaluating preparedness and prevention efforts aimed at reducing the impact of such events. Recent projections from climate change experts also suggest an increased frequency or intensity of climate-related

events that will likely call for ongoing epidemiologic surveillance. The role of injury epidemiology in these events is to identify the major health problems, determine the extent of injury, provide information on injury causation, help prioritize health interventions, and finally monitor health trends and program impacts.

Despite continued improvements in the workplace safety environment, occupational injuries remain an important worker health and safety issue that has resulted in significant health impacts among workers, with estimated costs of more than $150 billion annually. Sprains and strains are the most prevalent type of nonfatal occupational injury in the United States involving days away from work and contributed to more than 40% of the 1.3 million injuries and illnesses (in private industry) that required days off. Cuts and lacerations, contusions, and fracture injuries each contribute roughly 9% to 13% of workplace injuries that result in lost work time. The back is the region most commonly affected, followed by upper extremity (including hands and fingers) and lower extremity regions. Most injuries directly result from overexertion, being struck by or struck against an object, and falls. However, many other factors or conditions immediately preceding the injury-causing event are likely to influence injury occurrence and offer opportunities for intervention. Overall average injury rates range from 3.8 per 100 workers per year for lost time injuries to 8.3 per 100 workers for non-lost time injuries.

Unfortunately, no complete or comprehensive data systems exist to monitor occupational injuries, although several surveillance systems are in place such as the Bureau of Labor Statistics (BLS) system, the National Center for Health Statistics (NCHS) National Health Interview Study, and States Workers Compensation data are used to monitor workplace injury trends. Large individual companies also have developed health and safety surveillance systems that can be used for injury epidemiology. The industry sectors of manufacturing, mining, construction and agriculture/forestry/fishing typically have higher work-related injury rates. Within these sectors, the significant variation in injury risk by occupational groups, job tasks, and work location have not yet been fully characterized from an epidemiologic perspective.

In discussing the public health burden of injury and the need for further emphasis on this public health issue, injury researchers Thacker and MacKenzie (2003, p. 1) note that "epidemiologists have a critical role in describing these problems, conducting studies to determine what prevention interventions work, and helping the media, policy makers, and ultimately the public appreciate the impact of injuries." To have an important role in injury control and prevention, injury epidemiologists will need to move from predominantly descriptive epidemiology in injury research to more analytical studies. Descriptive epidemiology will remain important. To improve this area of research, however, there is a continual need to develop, improve, and enhance injury surveillance systems especially in the areas of nonfatal injury, occupational injury, and firearms-related injury. Injury reduction among children and injury reduction among the elderly remain top priorities and will continue to be important given the increasingly larger proportion of older aged individuals in many populations around the world. Advances in injury control and prevention will depend on the cooperative efforts of injury epidemiologists and scientists in engineering, biomechanics, sociology, criminology, and other fields.

—Michael A. Kelsh

See also Child and Adolescent Health; Disaster Epidemiology; Environmental and Occupational Epidemiology; Firearms; Vehicle-Related Injuries; Violence as a Public Health Issue; War

Further Readings

Baker, S. P., O'Neil, B., Ginsburg, M. J., & Li, G. (1992). *The injury fact book* (2nd ed.). New York: Oxford University Press.

Haddon, W. (1968). The changing approach to the epidemiology, prevention and amelioration of trauma: The transition to approaches etiologically rather than descriptively based. *American Journal of Public Health*, *58*, 1431–1438.

Hagberg, M., Christiani, D., Courtney, T. K., Halperin, W., Leamon, T. B., & Smith, T. J. (1997). Conceptual and definitional issues in occupational injury epidemiology. *American Journal of Industrial Medicine*, *32*, 105–115.

MacKenzie, E. J. (2000). Epidemiology of injuries: Current trends and future challenges. *Epidemiologic Reviews*, *22*, 112–119.

Robertson, L. S. (1998). *Injury epidemiology: Research and control strategies* (2nd ed.). New York: Oxford University Press.

Thacker, S. B., & MacKenzie, E. J. (2003). The role of the epidemiologist in injury prevention and control: An

unmet challenge (Preface). *Epidemiologic Reviews*, *25*, 1–2.
Weeks, J. L., Levy, B. S., & Wagner, G. R. (1991). *Preventing occupational disease and injury.* Washington, DC: American Public Health Association.

Insect-Borne Disease

The war between humans and insects predates recorded history. Not only have insects played an adversarial role in human development by destroying crops and killing livestock, they are also apt vectors for viruses, bacteria, parasites, and fungi, which infect and cause illness in humans. Insect-borne diseases affect population rates, trade, travel, and productivity. While the true impact on the global economy is incalculable, the cost is most likely billions and billions of dollars annually. Insects are on every continent, and consequently so are the diseases they carry. The incidence rates of many insect-borne diseases have decreased due to improved public health initiatives undertaken by both governmental and nongovernmental organizations. Unfortunately, such efforts must confront the amazing adaptability of both insects and pathogens.

Transmission Cycle

The majority of insect-borne diseases are transmitted directly through a bite, which is the means by which the pathogen is transmitted to the host. However, this mode of transmission is not common to all insect-borne diseases. An example of a unique mode of transmission occurs with Chagas' disease (American trypanosomiasis), which is transmitted by the reduviid commonly called the kissing bug. The kissing bug, which likes to bite its victims while they sleep, takes its blood meal and simultaneously defecate its infected excrement. It is only after the meal is complete, and the kissing bug has left its sleeping victim, that the person feels the need to scratch the site. This scratching smears the infected insect excrement into the fresh bite wound, thus infecting the victim.

There are four necessities for successful transmission of insect-borne diseases: first, a susceptible host (i.e., no acquired immunity either by natural infection or by vaccination); second, a suitable vector (the specific species of insect); third, the proper environmental conditions (not too hot, cold, dry, or moist); and fourth, presence of the pathogen. When all four of these criteria are present, the possibility of successful disease transmission exists. Alternatively, each one of these four requirements also represents a weakness, a place where the disease transmission cycle can be broken and where public health intervention can occur to prevent disease spread.

Plague

History's most infamous insect-borne disease is plague (caused by the bacterium *Yersinia pestis).* The bacterium is transmitted from animal to animal and then makes the leap from animal to humans via the bite of an infected rodent flea. Plague still evokes fear and panic in people living today, even though only 1,000 to 3,000 cases occur annually worldwide, with the United States claming 10 to 15 of these cases. Without proper antibiotics, plague has a mortality of 50% to 90%, but with proper diagnosis and treatment, the mortality drops to 15%.

Plague earned its notorious reputation in the mid-1300s when it swept through Central Europe and caused the deaths of one third of the population over the course of a few short years. The plague spread rapidly, aided at the time by the unsanitary conditions, increasing trade and travel routes, urbanization, and a lack of knowledge about how the disease was transmitted.

Mosquito-Borne Diseases

The most notorious insect for disease spreading is the mosquito. The mosquito is a known carrier of malaria, dengue fever, yellow fever, lymphatic filariasis, Japanese encephalitis, West Nile encephalitis, St. Louis encephalitis, Venezuelan encephalitis, and many more diseases. The mosquito is a worthy opponent, with its diverse range of breeding grounds (saltwater marshes to abandoned car tires), its ability to develop resistance to insecticides, and its flexibility of hosts from which to take a blood meal. Additionally, increasing global temperatures have aided the mosquito in acquiring an even broader habitat.

Combined annually, malaria, dengue, and yellow fever kill millions of people and cause illness in the hundreds of millions. Malaria, caused by a blood-borne parasite, is endemic in 91 countries and exposes 40% of the world's population to illness. With up to 500 million cases of malaria occurring annually, the

economic impact is tremendous, especially for Africa, where 90% of the cases occur.

Dengue, including dengue hemorrhagic fever and dengue shock syndrome, is the most important mosquito-borne virus. It is found in 100 tropical and subtropical countries and causes 20 million cases annually. The dengue virus has four serogroups (closely related viral strains). Exposure to one serogroup provides lifelong immunity. However, exposure to a different serogroup, for example in a second infection, can cause dengue hemorrhagic fever or dengue shock syndrome. This more serious infection is believed to occur when a patient's immune system is unable to recognize the slight difference between the two viral serogroups and is unable to mount the necessary response.

Other Insect-Borne Diseases

The impressive list of insect borne diseases is not limited to mosquito-borne illnesses. Endemic to Africa and Latin America, onchoneriasis, or river blindness, is transmitted by the black fly. This fly thrives in fast-flowing water with ample vegetation. Its bite transmits subcutaneous filariasis, which then progresses to itching, nodules, and, in severe infections, blindness. It is estimated that 17.6 million people are affected by onchoneriasis in Africa alone.

The tsetse fly, found in sub-Saharan Africa, is able to transmit sleeping sickness. With a prevalence of 300,000 cases, sleeping sickness causes inflamed nodules, headache, and fever, and one strain possibly causes inflammation of the heart and, in some cases, death.

Leishmaniasis, caused by a single-celled protozoan, is one of the commonest parasitic infections in the world. Leishmaniasis occurs on every continent except Australia and Antarctica and is transmitted by the bite of a sandfly. There are three different disease manifestations, visceral, cutaneous, and mucocutaneous, based on species type, cells infected, and the victim's immunity. The spread of this disease has recently accelerated due to increasing rates of exposure to the sandfly during infrastructure development such as dam construction, road development, and mining.

Public Health

Public health workers face many obstacles in their battles with insect-borne diseases. Factors that influence the emergence or re-emergence of insect-borne diseases include changes in public health policy, resistance to drugs and insecticides, shifts in emphasis from prevention to emergency response, demographic and societal changes, genetic changes in pathogens, urbanization, deforestation, and agricultural practices. The threat of insect-borne diseases is not new to humans, and even with our recent technological insights into disease transmission, diagnosis, and treatment, many insect-borne diseases continue to emerge or re-emerge. Nevertheless, continued outreach and research are essential for disease transmission to be controlled.

—Jerne Shapiro

See also Epidemiology in Developing Countries; Malaria; Plague

Further Readings

Heymann, D. (2004). *Control of communicable diseases manual* (18th ed.). Baltimore: United Book Press.

Kenrad, N., Masters Williams, C., & Graham, N. (2005). *Infectious disease epidemiology.* Sudbury, MA: Jones & Bartlett.

Travel Medicine. (2003). *Insect borne diseases*. Retrieved February 23, 2007, from http://www.travmed.com/thg/thg_pdf_2001/08-Insect-Borne%20D-01.pdf.

Web Sites

Centers for Disease Control and Prevention, Division of Vector Borne Infectious Diseases: http://www.cdc.gov/ncidod/dvbid/index.htm.

Institutional Review Board

An institutional review board (IRB), or ethics committee, reviews proposed and ongoing research conducted with human subjects. The purpose of the IRB review is to protect the rights and safety of human participants. The IRB review process was initiated in response to the Tuskegee Syphilis Study, in which subjects received substandard medical care without their consent. Initially focused on biomedical research, IRBs now review social science research and often liberal arts research as well (e.g., living history interviews). Institutions seeking federal funding must have an IRB, and the IRB must review and approve federally funded research studies. Most institutions require IRB approval for all research that

involves human subjects, not just that funded by the federal government.

The IRB process begins before participants are recruited for a study. The study protocol must satisfy the three basic principles of the Belmont Report: that the study design provide sufficient *beneficence* (maximizes the benefits compared with the risks of participation), *justice*, and *respect* for persons. Once the study is approved, the IRB is charged with overseeing the research from an ethics perspective. This oversight is usually exercised through two mechanisms. First, it provides all participants a means of contacting the IRB directly if they have any concerns, and, second, the IRB conducts periodic reviews of the study to monitor the research progress and address any ethics concerns. The review process usually occurs annually. Although rarely used, the IRB can require additional reviews and actively conduct surprise inspections of research records.

IRBs for institutions receiving U.S. federal funds must have at least five members, and most have more. The IRB must include members who represent diverse bodies of knowledge relevant to the conduct of ethical research. In addition, members should be demographically and culturally diverse. Several specific membership criteria must be satisfied:

- At least one member must be *from the scientific community* and knowledgeable about scientific research.
- At least one member must be *outside the scientific community*; this person should advocate for the nonscientific issues relevant to ethical conduct of research, such as legal issues and standards for professional conduct.
- At least one person must be from *outside the institution*; this person is usually a community member and represents the community standard for assessing the ethics of a study.

When the research proposed is outside the expertise of the IRB members, the IRB can invite experts in the research area to provide additional information in the review; however, these consultants are not allowed to vote. Most IRBs use a consensus approach (i.e., votes must be unanimous) to reach a decision, although some IRBs allow a majority vote. When a majority vote is used, the community member typically still has substantial power because most IRBs will not override the perspective of the community member. The administration of an institution (e.g., president of a university, director of a hospital) must allow the IRB to function independently, without undue influence related to funding pressures or other administration priorities.

The Office of Human Research Protections (OHRP) is responsible for the registration of IRBs and their oversight. The OHRP Web site provides substantial information about IRBs and the review process.

—Louise-Anne McNutt

See also Ethics in Human Subjects Research; Informed Consent; Tuskegee Study

Web Sites

FDA/Office of Science Coordination and Communication, Guidance for Institutional Review Boards and Clinical Investigators: http://www.fda.gov/oc/ohrt/irbs/default.htm.

U.S. Department of Health and Human Services, Office of Human Research Protections, Assurances: http://www.hhs.gov/ohrp/assurances/assurances_index.html. [An institution that engages in human subjects research must have an "assurance" indicating that the OHRP has cleared the researcher's IRB to authorize research projects. Individual investigators within the institution then apply individually to their institution's IRB for each project.]

Intent-to-Treat Analysis

Intent-to-treat analysis (ITT) requires that all study subjects be included in outcome analyses in the study condition in which they were assigned, or were intended to be treated, regardless of actual treatment or adherence to the research protocol. After being assigned to a study condition, subjects may not actually use the treatment or intervention, may use less than intended doses, may drop out from the research program and therefore have indeterminate outcomes, or may even cross over between study conditions. These subjects may differ systematically from those who follow the protocol, and their removal can invalidate random assignment, introduce bias, and lead to inappropriate interpretation of statistical tests. The concept of ITT originated in pharmaceutical and randomized clinical trials but applies to behavioral interventions as well.

ITT is associated with effectiveness trials, which try to emulate real-world circumstances by offering an intervention but not assuming protocol compliance. In contrast to ITT, per-protocol analysis and efficacy subset analysis are associated with efficacy trials, which typically are more rigorous in ensuring that an intervention is not only offered to participants but complied with as well. When conducted on groups with nonrandom biases or errors, efficacy subset analysis can result in a Type I error probability higher than usually accepted alphas of .05 or .10, leading to inappropriate rejection of null hypotheses. The exclusion of such subjects from analysis of intervention outcomes may result in an overestimation of the intervention's effectiveness. On the other hand, ITT analysis can weaken the statistical power of a study and can require the application of methods for handling missing data or crossover subjects. It is possible to conduct both types of analysis on the same data for different research questions: ITT for effectiveness and per-protocol analysis for efficacy in different groups of subjects.

Some have traced the emergence of the debate surrounding ITT versus per-protocol analysis to the Federal Drug Administration and others' criticism and defense of the Anturane Reinfarction Trial in the early 1980s. In this trial, a number of participants were determined to be ineligible after randomization, and although statistically equal numbers were removed from the experimental and control groups, more deaths were removed from the experimental group, leading to criticism of the reported outcome of the trial. ITT also was controversial in analysis of results from the 3-year Concorde Trial of AZT for asymptomatic HIV-infected patients that began in 1988. In this case, ethical considerations caused the Concorde protocol to change after 1 year of recruitment, so that some participants who had been randomly assigned to the placebo group were able to switch to AZT during the study. Following the ITT rule of "as randomized, so analyzed," the study suggested that AZT made no difference in progression to AIDS among asymptomatic people with HIV, countering accepted practice and stirring much controversy and concern among researchers and patients.

Research using the ITT principle can be enhanced if efforts are made in study planning and implementation to keep subjects as adherent to the protocol as possible and to minimize study attrition. Documentation of subjects' actual use of assigned treatment is crucial and often is reported by means of detailed flowcharts and tables of study compliance and attrition at all stages.

—Jane K. Burke-Miller

See also Effectiveness; Efficacy; Missing Data Methods; Randomization

Further Readings

Ellenberg, J. H. (1996). Intent-to-treat analysis versus as-treated analysis. *Drug Information Journal, 30*, 535–544.

Fisher, L. D, Dixon, D. O., Herson, J., Frankowski, R. F., Hearron, M. S., & Peace, K. E. (1991). Intention to treat in clinical trials. In K. E. Peace (Ed.), *Statistical issues in drug research and development.* New York: Marcel Dekker.

James, J. S. (1993, April 23). *AZT, early intervention, and the Concorde controversy*. AIDS Treatment News, No. 173. Retrieved November 2, 2006, from http://www.aegis.com/pubs/atn/1993/ATN17301.html.

Lachin, J. M. (2000). Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials, 21*, 167–189.

May, G. S., DeMets, D. L., Friedman, L. M., Furberg, C., & Passamani, E. (1981). The randomized clinical trial: Bias in analysis. *Circulation, 64*, 669–673.

Sheiner, L. B. (2000). Is intent-to-treat analysis always (ever) enough? *British Journal of Clinical Pharmacology, 54*, 203–211.

Interaction

See Effect Modification and Interaction

Internal Consistency

See Interrater Reliability

International Classification of Diseases

To classify means to arrange different entities in classes, groups, or categories, according to similar characteristics, based on defined criteria. Classification usually involves the categorization of relevant natural language for the purposes of systematic analysis within a single field of concepts. In the case of diseases and

health problems, there are many possible axes along which to organize a classification, including anatomy, etiology, and pathology, and the axis used will depend on the intended use of the compiled data.

The International Statistical Classification of Diseases and Related Health Problems (the complete name of the Classification also known by the acronym of ICD) is a global standard and international public tool for organizing and classifying information about diseases and other related health problems. It is a detailed list of known diseases, injuries, external causes of injuries, signs, symptoms, factors influencing health status, and contact with health services. The ICD is an epidemiological and statistical instrument developed to facilitate the understanding of health information on diseases and health-related problems and to help identify and monitor health situations and define health policies, priorities, resources, and programs.

The ICD is published by the World Health Organization (WHO) and is revised periodically. The currently used edition is the Tenth Revision (ICD-10), approved in 1990 by the World Health Assembly and published for the first time in 1992. The WHO assumed leadership and coordination of the preparation and publication of the successive revisions of the ICD, beginning with the ICD-6 in 1950.

History

The Hindu text known as *Sushruta Samhita* (600 AD) is possibly the earliest effort to classify diseases and injuries. The first statistical study of diseases and causes of death published in the world is considered to be the "Natural and Political Observation Made Upon the Bills of Mortality" of John Graunt, published in London in 1662. Graunt prepared a list of 83 causes of death, mixing etiology, pathology, circumstances, and other criteria, and recorded the number of citizens who died of each. However, the first internationally recognized attempt to systematically classify diseases is generally considered to be the *Nosologia Methodica* of François Bossier de Lacroix, better known as Sauvages (1706–1777). More then one century later, in 1891, the International Statistical Institute created a committee, chaired by the Chief of the Statistical Services of the City of Paris, Jacques Bertillon, to prepare a classification of causes of death, which was presented and approved at the 1893 meeting. The basic structure was based on that proposed by William Farr (1808–1883) in the beginning of the international discussions on classification structure. The scheme was, for all purposes, epidemiological or statistical, to organize diseases and conditions as follows:

- Epidemic diseases
- Constitutional or general diseases
- Local diseases arranged by site
- Developmental diseases
- Injuries

This general structure, which can still be identified in ICD-10, was reviewed in every new revision but is maintained because it is still considered more useful than any other tested.

The *Bertillon Classification of Causes of Death*, as this classification was first called, was adopted by several countries, and in 1898 the American Public Health Association recommended its adoption by Canada, Mexico, and the United States. It was also suggested by the Association that the classification should be revised every 10 years. Following that suggestion, the French Government convoked, in 1900, the First International Conference for the Revision of the International List of Causes of Death (Bertillon's classification). In August 1900, the First Revision was adopted.

The name *International Classification of Diseases—ICD* was defined at the Sixth Revision. ICD-6 incorporated several important changes, including a major enlargement of the number of categories, which were now to be used to classify morbidity as well as mortality; the inclusion of rules for selecting an underlying cause of death; the inclusion of the WHO's Regulations Regarding Nomenclature; Standards and Definitions related to maternal and child health; a recommended International Form of Medical Certification of Cause of Death; and special short lists for data tabulation.

The next major changes came with ICD-10. These included almost doubling the number of codes from that used in ICD-9, the adoption of an alphanumeric system, the definition of an updating system, and the definition of the concept of the family of international classifications.

Uses and Implementation

ICD is an instrument for recording and analyzing mortality and morbidity data and comparing data

collected in different areas or countries at different times. It is used around the world for health statistics, reporting, research, reimbursement systems, and automated decision support in medicine.

The transformation of medical terminology from words into codes permits easy storage and retrieval, and tabulation of data for analysis. Normally mortality and morbidity data are tabulated with a single code per death (the underlying cause) or morbid episode (the main condition). To maintain comparability, the translation of words into codes, the *codification*, and the selection of a single diagnosis from a death certificate or other medical record requires specific training.

The implementation of a new ICD Revision requires a number of actions, including the preparation of a version in the national language (the original version is published in English); purchase and distribution of printed and electronic versions; training coders; revision of validation and consistency tables; adjustments in data processing systems; preparation of new short lists for easier data tabulation and presentation, according to the country's needs; review of forms and publications; and informing and advising users.

International Classification of Disease (10th Revision)

To provide a common basis of classification for general statistical use, some adjustments must be made on a strictly logical arrangement of morbid conditions. That is the reason why the Tenth Revision of ICD, as with previous revisions, is a multiaxial classification, where the chapters are organized according to different approaches: etiology, anatomical site, age, circumstance of onset, and the quality of medical information. ICD-10 is presented in three volumes:

- Volume 1, the Tabular List, contains the Report of the International Conference for the Tenth Revision, the classification itself at three and four character levels, the classification of morphology of neoplasms, special tabulation lists, definitions, and the nomenclature regulations.
- Volume 2, the Instruction Manual, contains notes on certification and classification, use of ICD, guidance on coding, tabulations, and historic material.
- Volume 3, the alphabetical Index, contains three sections, one for diseases and nature of injuries, one for external causes of injuries, and one with a table of drugs and chemicals for poisonings.

Several changes in organization were made in the revision of ICD-9 to ICD-10. While ICD-9 had 17 sections and two supplementary classifications (External causes of injury, and Factors influencing health status and contact with health services), ICD-10 has 21 chapters, including those two supplementary in the core classification. Also, the former Section VI, Diseases of the Nervous System and Sense Organs, was divided in three chapters: Diseases of the nervous system, Diseases of eye and adnexa, and Diseases of the ear and mastoid process. The ICD-10 chapters, with the range of codes in parentheses, are as follows:

I: Certain infectious and parasitic diseases (A00–B99)

II: Neoplasms (C00–D48)

III: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50–D89)

IV: Endocrine, nutritional, and metabolic diseases (E00–E90)

V: Mental and behavioral disorders (F00–F99)

VI: Diseases of the nervous system (G00–G99)

VII: Diseases of the eye and adnexa (H00–H59)

VIII: Diseases of the ear and mastoid process (H60–H95)

IX: Diseases of the circulatory system (I00–I99)

X: Diseases of the respiratory system (J00–J99)

XI: Diseases of the digestive system (K00–K93)

XII: Diseases of the skin and subcutaneous tissue (L00–L99)

XIII: Diseases of the musculoskeletal system and connective tissue (M00–M99)

XIV: Diseases of the genitourinary system (N00–N99)

XV: Pregnancy, childbirth, and the puerperium (O00–O99)

XVI: Certain conditions originating in the perinatal period (P00–P96)

XVII: Congenital malformations, deformations, and chromosomal abnormalities (Q00–Q99)

XVIII: Symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified (R00–R99)

XIX: Injury, poisoning, and certain other consequences of external causes (S00–T98)

XX: External causes of morbidity and mortality (V01–Y98)

XXI: Factors influencing health status and contact with health services (Z00–Z99)

The letter U is not used, being reserved for the provisional assignment of new diseases (U00–U49) or for research needs (U50–U59). When used internationally, codes assigned to the letter U are considered as the 22nd chapter of ICD-10.

Each chapter is organized in blocks that are kept as homogeneous as possible, with three-character categories describing diseases, injuries, or other conditions. Within each block, some of the three-character categories are used for single diseases or conditions, selected because their public health importance, severity, or frequency, while others are used for groups of diseases with common characteristics. Usually there are provisions for "other" conditions, allowing the inclusion of different but infrequent conditions and "unspecified" conditions.

Three chapters (XIII, XIX, and XX) offer fifth-character levels as subclassifications along an axis different from the fourth character.

Due to different use of terminology, particularly in mental health, the corresponding Chapter V provides a glossary description to clearly indicate the content of each code. Similarly, Chapter XX, for external causes, presents a set of definitions for transportation accidents.

National Adaptations of ICD

Since ICD-6 began to be used for morbidity, in 1950, many considered that it was based too heavily on epidemiology, statistics, and pathology; some countries began to prepare national adaptations to have more codes for different clinical options related to the health care provided in hospitals. In 1977, because of the increasing use of coded data with ICD in morbidity and the need for a better accuracy, the United States called a national conference of experts to discuss a new adaptation of the Ninth Revision of the Classification, which was to be used starting in 1979. The resulting expansion, called ICD-9-CM (Clinical Modification), was considered to meet the country's needs and was completely compatible with the original ICD-9. Together with the expanded classification of diseases, an updated classification of procedures was published as part of ICD-9-CM. Since then, the "CM" version is used not only in the United States but also in several other countries.

For the Tenth Revision, in addition to ICD-10-CM, three other countries had already developed their clinical adaptations—Canada (ICD-10-CA), Australia (ICD-10-AM), and Germany (ICD-10-GM)—to meet their own requirements relating to different health care practices, uses of data, and needs.

Updating Process

The main objectives of the ICD update and revision process are to maintain the classification as both user friendly and scientifically reliable, while using modern knowledge management and continuously synthesizing scientific advances in health care.

Prior to the Tenth Revision, updates were not published between revisions, which occurred in 10-year cycles. From ICD-9 to ICD-10 the cycle was of 15 years. In 1989, the WHO ICD-10 International Conference recommended the definition of an updating mechanism so that changes could be implemented between revisions. To that effect, two separate bodies, the Mortality Reference Group (MRG) and the Update Reference Committee (URC), were established in 1997 and 1999, respectively, to initiate and follow up on that process.

The MRG is composed of members from the different WHO Collaborating Centers for Classification of Diseases and makes proposals relating to the application and interpretation of the ICD to mortality, as well as recommendations to the URC on proposed ICD updates. The URC receives proposals from the MRG and other sources and revises and submits recommendations on proposed ICD updates for mortality and morbidity to the WHO and the Collaborating Centers for a final decision. These recommendations reinforce the process of *updating* the ICD-10 rather than creating the foundation of an ICD-11. This continuous process is facilitated by consideration of reports sent from individual countries to their corresponding WHO Collaborating Center on any significant problems in the use of the ICD-10.

Changes to the ICD vary in nature from minor corrections, which are updated in the classification's tabular list, instruction manual, or alphabetic index every year, to major alterations that take place every 3 years.

Minor changes include the following:

- Correction or clarification of an existing index entry that only changes the code assignment to a code within the same three-character category

- Enhancements to the tabular list or index such as the addition of an inclusion term to an existing code, the addition of an exclusion note, or the duplication of an existing index entry under another main term
- Changes to a code description that enhance the description rather than change the concept
- Changes to a rule or guideline that does not affect the integrity of morbidity or mortality data collections
- Corrections of typographical errors

Major changes include the following:

- Addition of a code
- Deletion of a code
- Movement of a code to another category or chapter
- Change to an existing index entry that changes the code assignment from one three-character category to another three-character category (movement of terms)
- Changes to a rule or guideline that affect the integrity of morbidity or mortality data collections
- Introduction of a new term into the index

Once approved, the updates are posted on WHO's Web page, translated into different languages, and incorporated in new printed editions and electronic versions.

International Classification of Disease (11th Revision)

One of the topics addressed at the annual WHO-FIC Network's meeting in Reykjavik in October 2004 was the WHO's proposal to start the preparation of the ICD-11. As of November 2006, under the coordination of WHO, the tentative timeline to be followed is to have an *alpha version* for internal discussion by 2008, a *beta version* for worldwide discussion and field trials by 2010, a final version by 2012, and the version for the World Health Assembly approval by 2014, the implementation being envisaged to start by 2015.

The most relevant tasks required to prepare ICD-11 are the following:

- Consult with Member States, multiple parties, and professional organizations to ensure a comprehensive response to different aspects of health care
- Create an Internet platform in multiple languages to enable participation of all interested parties using transparent knowledge management and sharing mechanisms
- Convene expert groups in different areas (e.g., oncology, gastroenterology, sleep disorders, mental health, and others) that are the subject of significant specialty interest
- Explore the congruence within the WHO Family of International Classifications: ICD with derived classifications such as ICD-O (adaptation for Oncology) and related classifications such as ICECI (International Classification of External Causes of Injury) and ICPC-2 (International Classification for Primary Care, second edition), as well as ICF (International Classification of Functioning, Disability, and Health) and ICHI (International Classification of Health Intervention—under development in 2006) for conceptual overlaps, and joint use
- Explore changes in national classifications schemes and ICD modifications, known collectively as ICD-XM (this includes the set of different national "Clinical Modifications," including ICD-9-CM/ICD-10-CM, ICD-10-AM, ICD-10-CA, and ICD-10-GM, respectively, from United States, Australia, Canada, and Germany) to respond to the need expressed in national classifications because these will indicate the user needs and advances in science
- Evaluate the ICD-10 implementation process and ICD-10 update process
- Explore information technology (IT) and standards requirements such as terminology, links and mappings, indexes, rules
- Conduct a staged development process including relevance, coverage, utility, translatability, and links with other IT applications
- Prepare a package of training and implementation tools such as coding software and linkage to IT systems, translation tools, and bridge-coding with ICD-10
- Develop a clear communication and dissemination strategy

One of the additional tasks is to develop a kind of enlargement of the ICD to address the needs for morbidity use, trying to harmonize in some way the national modifications of the ICD, including the versions labeled CM (United States), AM (Australia), CA (Canada), and GM (Germany).

—Roberto Augusto Becker

See also International Classification of Functioning, Disability and Health; Secondary Data; World Health Organization

Further Readings

World Health Organization. (2001). *International classification of functioning, disability and health*. Geneva, Switzerland: Author.

World Health Organization. (2004). *International statistical classification of diseases and related health problems* (2nd ed., 10th rev., Vols. 1–3). Geneva, Switzerland: Author.

Web Sites

International Classification of Diseases, WHO: http://www.who.int/classifications/icd/en.

International Classification of Functioning, Disability, and Health

The World Health Organization (WHO) endorsed the publication and worldwide use of the *International Classification of Functioning, Disability, and Health* (ICF) in May 2001. The ICF is the successor classification of the *International Classification of Impairments, Disabilities, and Handicaps* (ICIDH), which was released for field trial purposes in 1980. Responding to criticisms of the ICIDH, and the fact that it was little used for data collection or epidemiological purposes, WHO initiated a 10-year international collaborative revision exercise, with input from hundreds of health professionals, epidemiologists, health statisticians, health systems analysts, and members of disability advocacy groups. To ensure cross-cultural and linguistic applicability of the ICF, the drafting process was informed by data from a set of innovative cultural applicability field tests (reported in Üstün, Chatterji, Bickenbach, Trotter, & Saxena, 2000).

The ICF is a classification of dimensions of human functioning and disability associated with health conditions. WHO's 1947 Constitution obliges it to collate health statistics on mortality and disability in an internationally comparable format. The ICF, as a companion classification to WHO's considerably more well-known *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), provides a complete and consistent "language of health" for international health data. Comparable health data covering the full range of the health experience are an essential prerequisite to health outcomes research and health systems performance assessment, the primary epidemiological tasks of the WHO.

ICF's Conceptual Model

The underlying rationale of the ICF is that the lived experience of health is primarily a matter of the range and extent of functioning across all domains. Classifying human functioning requires a distinction between functioning at the body level (e.g., visual acuity, digestive functions, metabolic functions, and muscle functions) and the level of the whole person (thinking, communicating, walking, maintaining interpersonal relationships, attending school, and working). These functions, simple and complex, identify basic bodily functions and human capacities to perform actions and display behaviors that constitute a person's state of health. A third level of functioning completes the picture of the lived experience of health by characterizing functioning as an outcome of the interaction between body and person level capacities and the complete physical, human-built, social, and attitudinal environment. ICF is therefore a classification of the full range of human functioning and decrements in function—that is, disabilities—at the body and person levels, as well as the actual performance of these capacities, within and as modified by the complete context of the individual's world (Figure 1).

The concept of "disability" in the ICF therefore refers to decrements or difficulties encountered in functioning at all three levels of functioning. Decrements in body functions are called *impairments*,

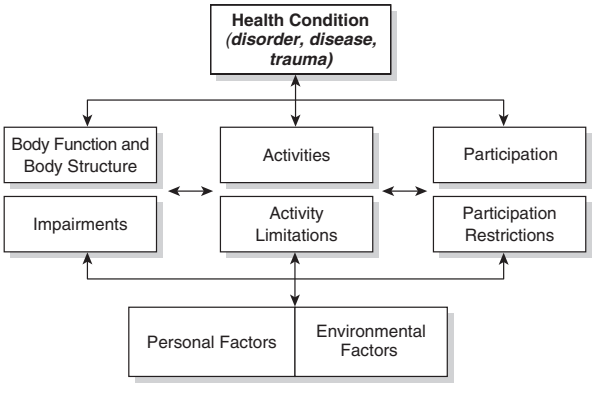


Figure 1 ICF Model of Functioning, Disability, and Health

decrements in person level capacities are called *activity limitations*, and difficulties in performing in context all human actions and displaying behaviors constitutive of the lived experience of health are called *participation restrictions.* Impairments and activity limitations are fully within the domain of health (indeed a fully operationalizable definition of *health* for measurement purposes can be given in terms of functions), whereas participation restrictions, as outcomes of person-environment interactions, are outside the domain of health, although associated directly with health states.

ICF as a Health and Disability Classification

The ICF is a classificatory tool for clinical, administrative, research, and epidemiological collection of health and disability information. ICF consists of three separate classifications: a classification of body functions and body structures, a classification of person level activities and varieties of participation, and a classification of environmental factors (Table 1). (The overall model also incorporates as part of the "context" of functioning and disability, personal factors—such as gender, race, age, fitness, lifestyle habits, social background, education, profession, and past and current experience—but these are not classified in the ICF.)

Each classification is a hierarchical arrangement of item terms (1,400 overall for all three classifications), each of which is operationally defined. Examples are provided in each definition, as well as inclusions and exclusions. For data management, each item is identified by a coding number that designates the dimension of functioning and disability, as well as placement within the hierarchical structure. Although ICF is not itself an assessment tool, coded qualifiers are provided that can be used for scaling extent of decrement and other information.

Epidemiological Applications of ICF

Although debates about models of disability have raged for decades, the multilevel, person-environment interactive conception of disability is now the standard model of disability and, in various versions, is widely employed clinically and in research across health and health-related disciplines. The ICF, however, represents the only complete, systematic, and extensively field-tested classification of functioning and disability that builds on the interactive approach. As such, it is the only scientifically viable tool for health and disability data collection and analysis. It is not surprising then that since 2001, ICF has been adopted around the world, both as an organizational model of health and disability services and as a classification platform for assessment and measurement tools and data management and analysis.

The potential impact of ICF on both descriptive and research epidemiology may be assessed by comparing the ICF approach with what has long been, and in many parts of the world continues to be, the standard one used for health and disability data collection. Disability data are often collected by means of population health surveys, or derived from administrative data collections, in which disability is divided into broad categories: Blind, deaf, crippled, and retarded are the traditional ones. Alternatively, disease conditions are used as proxies for functional limitations. The resulting prevalence information is of little use for resource allocation or public health planning because it is impossible to tease apart the various factors that were involved in the creation of the disability that is actually experienced by individuals. Risk factor analysis is equally problematic in this approach. The traditional approach, in a word, identifies the "problem of disability" as a confused muddle of body-level functioning decrements, person-level capacity problems, and restrictions in the actual day-to-day context of the individual's life.

In particular, the crucial role of the person's environment is either ignored or misdescribed in the traditional approach. Environmental factors can worsen a functional difficulty by creating obstacles that limit the nature and extent of a person's participation in life activities. Conversely, environmental factors such as assistive technology or policies of workplace accommodation can ameliorate (if not eliminate) the impact of a capacity limitation on a person's actual lived experience. Disability epidemiology must be able to separately measure the impact on a person's life of, on the one hand, his or her health problem and, on the other, the role of the person's physical and social environment in worsening or ameliorating the impact of that health problem on the person lived experience. Only then can sensible conclusions about intervention strategies and prevention policies be derived.

The ICF is also a potentially powerful research tool for epidemiology. If, for example, researchers propose

Table 1 The ICF Classifications: First Level

Body Functions	
Chapter 1	Mental functions
Chapter 2	Sensory functions and pain
Chapter 3	Voice and speech functions
Chapter 4	Functions of the cardiovascular, hematological, immunological, and respiratory systems
Chapter 5	Functions of the digestive, metabolic, and endocrine systems
Chapter 6	Genitourinary and reproductive functions
Chapter 7	Neuromusculoskeletal and movement-related functions
Chapter 8	Functions of the skin and related structures
Body Structures	
Chapter 1	Structures of the nervous system
Chapter 2	The eye, ear, and related structures
Chapter 3	Structures involved in voice and speech
Chapter 4	Structures of the cardiovascular, immunological, and respiratory systems
Chapter 5	Structures related to the digestive, metabolic, and endocrine systems
Chapter 6	Structures related to the genitourinary and reproductive systems
Chapter 7	Structures related to movement
Chapter 8	Skin and related structures
Activities and Participation	
Chapter 1	Learning and applying knowledge
Chapter 2	General tasks and demands
Chapter 3	Communication
Chapter 4	Mobility
Chapter 5	Self-care
Chapter 6	Domestic life
Chapter 7	Interpersonal interactions and relationships
Chapter 8	Major life areas
Chapter 9	Community, social, and civic life
Environmental Factors	
Chapter 1	Products and technology
Chapter 2	Natural environment and human-made changes to environment
Chapter 3	Support and relationships
Chapter 4	Attitudes
Chapter 5	Services, systems, and policies

to formulate hypotheses about the compression of morbidity, and then to test these claims, it is essential that researchers be able to pull apart the intrinsic health contributions to disability from the extrinsic or environmental contributions. If we notice that the incidence of chronic disability conditions is occurring later in life, it is an open research question whether this is because of changes in the patterns of the underlying health condition or changes in the environment that lessen the impact of these health decrements. Most likely, it is a complex mixture of both determinants, but researchers must be able to collect data that clearly separate the impact of these determinants so that measurement is possible.

Because the ICF's model describes decrements of functioning on a continuum, and ICF classifications offer the language needed to describe this full spectrum of human functioning, the ICF greatly expands our ability to understand disability and the complex processes that create disability as a lived experience. As a clinical tool, the ICF facilitates collection of data that can be systematically analyzed for health administrative purposes. Additionally, insights into patterns of disabilities across populations can be analyzed in terms of demographic factors and social determinants for a more complete picture of disability at the population level. The range of epidemiological uses for ICF is indeed extensive, and as the developing literature on the applications of ICF is indicating, the ICF is becoming a powerful tool for epidemiology.

—Jerome E. Bickenbach

See also Disability Epidemiology; International Classification of Diseases; World Health Organization

Further Readings

Bickenbach, J. E., Chatterji, S., Badley, E. M., & Üstün, T. B. (1999). Models of disablement, universalism and the ICIDH. *Social Science & Medicine*, *48*, 1173–1187.

Bickenbach, J. E., Chatterji, S., Kostansjek, N., & Üstün, T. B. (2003). Ageing, disability and the WHO's International Classification of Functioning, Disability and Health (ICF). *Geneva Papers on Risk and Insurance: Issues and Practice*, *28*, 294–303.

Leonardi, M., Bickenbach, J., Üstün, T. B., Kostanjsek, N., & Chatterji, S. (2006). The definition of disability: What is in a name? *Lancet*, *368*, 1219–1221.

Trotter, R. T., II, Üstün, B., Chatterji, S., Rehm, J., Room, R., & Bickenbach, J. (2001). Cross-cultural applicability research on disablement: Models and methods for the revision of an international classification. *Human Organization*, *60*, 13–27.

Üstün, T. B., Chatterji, S., Bickenbach, J. E., Trotter, R. T., II, & Saxena, S. (2000). *Disability and culture: Universalism and diversity.* Bern, Switzerland: Hogrefe & Huber.

Üstün, T. B., Chatterji, S., Kostansjek, N., & Bickenbach, J. E. (2003). The World Health Organization's International Classification of Functioning, Disability and Health (ICF) and functional status information in health record. *Health Care Financing Review*, *24*, 77–88.

World Health Organization. (1992). *International statistical classification of diseases and related health problems* (10th rev., ICD-10). Geneva, Switzerland: Author.

Web Sites

International Classification of Functioning, Disability, and Health: http://www3.who.int/icf/icftemplate.cfm.

Internet Data Collection

Collecting data through the Internet has become increasingly popular with researchers in epidemiology and public health. Internet data collection offers the advantages of saving time and money on research projects. However, researchers must also be aware of how concerns such as privacy, security, and error relate specifically to Internet data collection. This entry focuses on three specific topics: types of Internet data collection, techniques and protocols for Internet data collection, and privacy issues.

Types of Data Collection

There are two general categories of Internet data solicitation—*active* and *passive.* In active electronic data solicitation, the researcher directly contacts a specific population and asks for a response. In contrast, passive electronic data solicitation does not target specific respondents.

Active Data Solicitation

A researcher using active Internet data collection methods actively targets specific individuals or identifiable, restricted populations and directly asks for their opinion, viewpoint, or feedback. Examples of the type of population that might be targeted by active data solicitation include everyone who attended an

infection control training session, all the pediatric physicians on a listserv, and all new enrollees in a WIC program (the federal nutritional supplement program for women, infants, and children). Active data solicitation includes techniques such as Web and e-mail surveys where the researcher sends an e-mail with an embedded electronic survey or sends a link to a Web site containing the survey. In some instances, the link for the Web site could be mailed or presented (e.g., at a conference) for individuals to access directly. Regardless of how potential responders get to the survey, active data solicitation requires the researcher to get respondents to use their own time to provide information to the researcher. In addition, all respondents must have sufficient access to a computer and the Internet to be able to respond. A current uncertainty that often arises with this method lies in understanding the response rate, that is, the relationship between the number of potential respondents and the number of those who actually participated. The researcher knows how many people responded to the survey, which yields the numerator, but what is the denominator? To put it another way, how can you determine the number of individuals who are part of the population you targeted but who did not respond to your survey? This is less of a problem when a survey is e-mailed to specified recipients, although there may still be issues such as those raised by inactive e-mail addresses. However, when a survey is posted on a listserv, calculating the response rate presents a major problem. In many cases, the researcher does not know how many people actively read and participate in a particular listserv, so it is unclear what number should be used for the denominator in calculating the response rate. In addition, standards for acceptable response rates for Internet surveys still need to be established.

Passive Data Solicitation

Passive electronic data solicitation, in contrast, does not target specific respondents. For example, a researcher could post a link to his or her survey on a popular news Web site so that anyone accessing the Web site could complete the survey. In this example, the researcher does not specifically solicit information from a known person or restricted population; in fact, the respondents are self-selected rather than targeted by the researcher, although there is a general concept of targeting in the choice of the Web site where the link is posted. For instance, some Web sites are more likely to be accessed by persons with a liberal or progressive political point of view, others by persons with a conservative political point of view. As with active data solicitation, respondents are still required to spend their own time and effort to complete the survey.

Passive data collection suffers from two major biases and also presents a methodological difficulty. The first type of bias is due to the fact that only people who use the Internet and visit the Web site where the survey was posted are eligible to answer the survey. This relates both to issues of Internet access and use and to the type of Web site chosen to post the survey, as in the liberal/conservative political example of the foregoing. The second type of bias is due to the fact that the survey participants are self-selected among users of the Web site, and people who choose to respond to an unsolicited Internet survey may differ in many ways from people who do not choose to respond. In addition, the problem of calculating the response rate is even more difficult than in the active solicitation example because the researcher has no good way of determining what the denominator should be and there are no currently accepted standards for the response rate for this type of survey.

Surveillance and Data Mining

In addition to active and passive data solicitation, a third set of procedures—which may be classified as surveillance and data mining methods—can be used to collect data from the Internet. These methods are distinguished by the fact that they do not require the "respondent"—the person whose actions, preferences, and so on are represented by the data being collected—to do any additional work, yet such methods can generate a vast amount of information for the researcher. Surveillance involves the act of pursuing data while events are occurring or shortly afterward. An example of this category is public health officers monitoring biological terrorism Web sites and Internet blogs to determine how many people visit the sites and what content is posted on them. Data mining, on the other hand, is the analysis of patterns in information without regard to time of collection. For example, the sales records of an on-line bookstore could be analyzed to discover relationships among book purchases or between book and ancillary product purchases. Many types of statistical procedures are used

in data mining, from the very simple techniques such as means and correlations to computationally complex methods such as exhaustive regression. While the definitions of surveillance and data mining are not always distinct, they share a characteristic that distinguishes them from both active and passive Internet data solicitation: Both surveillance and data mining collect and analyze data for a purpose other than that for which it was originally created.

Tools and Techniques

Most researchers constructing a Web survey use a pre-existing software package created for that purpose. Many software packages exist that help the researcher create surveys for the Internet, including commercial software packages such as Snap Surveys and SurveyMonkey. Researchers at the Pennsylvania State University's Survey Research Center have made a number of presentations available on-line that explain the nuances of traditional and electronic surveys in more detail.

The basic principles of survey design apply to on-line surveys. In addition, certain technical guidelines apply to most on-line surveys independent of which software is used. These include the following:

- keep the survey to one screen width (i.e., do not make the viewer scroll horizontally);
- restrict individual questions to one screen length, particularly in longer surveys, so the respondent does not have to scroll down to see all of a question;
- limit open-ended questions;
- do not start with open-ended questions; and
- view the survey on different computer types *and* screen sizes since you have no control over the hardware your respondents will use.

In addition, specific issues regarding follow-up, sampling error, coverage, measurement, and nonresponse are addressed in the books by Best and Dillman listed in the Further Readings section.

Privacy Issues

There are two types of privacy issues related to Internet data collection: privacy concerns related to the collection and storage of the data, which are common to any research project that includes identifiable information, and privacy concerns regarding transmission of data to and from the participant and researcher.

Many people believe that they are "anonymous" when they use the Internet, and they are surprised to learn how much information can be collected about them and their activities. For instance, when someone visits a Web site, information that can be collected and stored includes the domain name (e.g., "sage .com," "stanford.edu") of the site accessed, the computer Internet Protocol (IP) addresses of the accessing computer, the type of browser used to access the site (e.g., Internet Explorer, Netscape, Mozilla), and the operating system of the accessing computer (Windows XP, Linux, MAC OS X). Other information, such as dates and times the site was accessed, specific Web pages within a site that were accessed, and links that were followed, can also be collected automatically and analyzed.

Organizations such as the U.S. Census Bureau collect this information about those who access their Web sites and also provide a Web page clarifying what information they collect and why. Many commercial Web sites also collect this information. The relevance for the researcher interested in collecting data over the Internet is that information about every person who takes a survey or visits the research Web site can be collected automatically, and some of this information can be used to identify individuals. For example, because each IP address is specific to a computer, a list of IP addresses provides a list of the computers that accessed the Web site. If an individual accessed the Web site from his or her home computer, the computer would be identified by its IP address—in the same way that a street address identifies a house. Not only does this raise confidentiality concerns, many people may be reluctant to complete a survey if they are not assured that procedures are in place to ensure the confidentiality of their responses.

In addition, when the information is collected by the researcher or submitted by the participant using the Internet, transmission of data must be secured. This is usually accomplished by encrypting electronic data, authenticating Internet sources, and controlling access to data once they are collected. Researchers are responsible for informing themselves about data security issues when collecting data using the Internet, just as they would be with more traditional methods such as face-to-face interviews or chart reviews. Different methods are available for collecting and storing electronic information, and it is the researcher's responsibility to see that these problems are treated appropriately, in the same way that, for instance,

hospital or clinical charts must be secured and kept confidential.

Although this discussion has focused specifically on Internet data collection, good practices for survey data collection in general apply equally on- or offline. The researcher will still need to test the survey usability, including the readability and accessibility of the survey, before beginning data collection. As with any survey, the researcher must define what is needed from the data and what he or she expects to find before starting data collection. This will assist in planning the survey development and data collection process so that it is neither too broad—providing far more data than can be used—nor too narrow—not giving the researcher enough data on which to base an understanding of the subject matter. With all survey methods, maintaining the confidence and trust of the participant should be a high priority for the researcher.

—Leslie McIntosh

See also Bias; Measurement; Questionnaire Design; Sampling Techniques; Survey Research Methods

Further Readings

Best, S. J., & Krueger, B. S. (2004). *Internet data collection*. Thousand Oaks, CA: Sage.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York: Wiley.

Dillman, D. A., & Bowker, D. K. (2001). The Web questionnaire challenge to survey methodologists. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 159–178). Lengerich, Germany: Pabst Science.

Sampson, S. E. (1998). Gathering customer feedback via the Internet: Instruments and prospects. *Industrial Management & Data Systems*, *98*, 71–82.

U.S. Census Bureau. (2001). *Federal audit clearinghouse Internet privacy policy statement*. Washington, DC: Author.

Web Sites

Snap Surveys: http://www.snapsurveys.com.

SurveyMonkey: http://www.surveymonkey.com.

Survey Research Center at Pennsylvania State University: http://www.ssri.psu.edu/survey/educ.htm.

U.S. Census Bureau: http://www.census.gov.

Interquartile Range

The interquartile range (IQR) is a resistant measure of spread for a quantitative data set. A "resistant measure" is not influenced by outliers. A "measure of spread" indicates how consistent or variable the data set is. Other measures of spread that are not resistant to outliers are the standard deviation and the range (maximum – minimum). The IQR measures the number of units in which the middle 50% of the data lie and is used in one technique to determine outliers in a data set.

The IQR is the difference between the third and first quartiles, Q3 and Q1, respectively, that is, $IQR = Q3 - Q1$. The IQR is always a positive value. The Q3 value is the 75th percentile, while Q1 is the 25th percentile; hence, the difference between these values gives the distance that contains the middle 50% of the observations. Different statistical packages may calculate Q1 and Q3 differently, which leads to slightly different values of the IQR. The difference in values between statistical packages is usually insignificant.

Consider Data Sets A and B in Tables 1 and 2, respectively, (created by the author for this entry).

By observation, one sees that Data Set B is more consistent than Data Set A. Data Set A is more variable. The first and third quartiles for the data sets are $Q1A = 5$, $Q3A = 55$, $Q1B = 112$, and $Q3B = 117$. Hence, $IQRA = 55 - 5 = 50$ and $IQRB = 117 - 112 = 5$. The middle 50% of the observations for Data Set A are contained within 50 units, while for Data

Table 1 Data Set A, Which Contains More Variability: IQR = 50

2	4	6	18	20	45	52	58	60	100

Table 2 Data Set B, Which Is More Consistent: IQR = 5

110	112	112	114	115	116	116	118	120	130

Set B it only takes 5 units. Since IQRA > IQRB, Data Set A is more variable than Data Set A.

One can observe that any changes in the minimums or maximums in the data sets will not affect the quartiles, which means that the IQR will not be affected. Hence, the IQR is resistant to outliers.

Outliers and the IQR

The IQR is also used to determine outliers in a data set. The rule of thumb is this: If an observation is 1.5 IQRs away from either the first or third quartile, then the observation is considered an outlier.

Determining whether Data Set A contains any outliers:

$$\begin{aligned}\text{Lower fence} &= Q1 - 1.5IQR = 5 - 1.5 \times 50\\ &= 5 - 75 = -70.\\ \text{Upper fence} &= Q3 + 1.5IQR = 55 + 1.5 \times 50\\ &= 55 + 75 = 130.\end{aligned}$$

Since Data Set A does not contain any values below the lower fence or above the upper fence, Data Set A does not contain any outliers.

Determining whether Data Set B contains any outliers:

$$\begin{aligned}\text{Lower fence} &= Q1 - 1.5IQR = 112 - 1.5 \times 5\\ &= 112 - 7.5 = 104.5.\\ \text{Upper fence} &= Q3 + 1.5IQR = 117 + 1.5 \times 50\\ &= 117 + 7.5 = 124.5.\end{aligned}$$

Data Set B does not have any value below the lower fence, but the value 130 is above the upper fence. Hence the value 130 is considered an outlier.

A data set may have multiple outliers.

—Marjorie E. Bond

See also Inferential and Descriptive Statistics; Measurement; Measures of Variability

Further Readings

Agresti, A., & Franklin, C. (2007). *Statistics: The art and science of learning from data.* Upper Saddle River, NJ: Pearson Prentice Hall.

Moore, D. S. (2007). *The basic practice of statistics* (4th ed.). New York: W. H. Freeman.

Interrater Reliability

Researchers and practitioners rely on a variety of instruments for measurement, such as scales, surveys, and recordings. For an instrument to be useful, it must be both reliable (i.e., measurements made using it are consistent and can be replicated) and valid (i.e., it measures what the researcher thinks it is measuring). This entry discusses one aspect of reliability, interrater reliability, in the general context of reliability in measurement.

Reliability essentially refers to repeatability. In asking how reliable an instrument is, we are asking whether we would get the same results if we were to take the same measurement of the same entity over and over again. Reliability does not require that an instrument deliver perfect measurements; rather, it assumes that some error occurs when we use an instrument repeatedly but that the error is random rather than systematic. As a result, after multiple measures using the same instrument, the random errors should cancel each other out and we would have an estimate of the true quantity of whatever we were measuring. Reliability also means that the quality being measured does not change from one occasion of measurement to another; thus, for example, we could evaluate the reliability of a scale in measuring children's heights by taking several measurements within the same day but perhaps not by taking several measurements over a period of months (because their actual height could have changed during that time).

Reliability is a prerequisite for validity. If an instrument yields wildly different results for a presumably unchanging entity on different occasions, it is not possible to evaluate the validity of the instrument because the unreliable nature of the scores precludes interpreting their meaning. On the other hand, if a survey or screening instrument yields essentially the same results over and over again, then we can go on to evaluate whether the instrument is valid, that is, whether it is in fact measuring what we are hoping it is measuring. However, assessing reliability is not always as simple as taking repeated measurements with the same instrument because other factors may cause repeated measurements to be invalid. For instance, if someone is asked the same question repeatedly, he or she may change his or her response because asking the question may have caused him or her to reflect on the issue and change his or her

answer from the first to the second asking, or he or she may have become tired of answering the same questions repeatedly and ceased giving honest answers. Because true reliability is impossible to measure without the impact of this survey effect, a variety of alternative means have been developed to capture the idea of repeated measurements.

Interrater reliability is just one possible way of assessing the reliability of a measurement instrument. As its name suggests, interrater reliability is a method of comparing the observations of multiple "raters" or judges. It is often used in psychological and behavioral evaluations (e.g., in judging if a child engages in disruptive behavior in a classroom setting) and in evaluating the accuracy of medical procedures such as reading X rays or evaluating mammograms. Rather than making one individual perform multiple ratings of the same person or event, interrater reliability uses multiple people to observe a single set of responses or actions of an individual and then examines the extent to which different judges agree. If the ratings of the judges do not agree, then the measure is not valid and the instrument used to collect the information may need revision, or the judges may need more training to use it correctly. If the judges do agree, this is supportive evidence that the measure may be a valid one. Of course, we do not expect perfect agreement, and several statistical methods have been developed to evaluate interrater reliability.

Methods for Computing Interrater Reliability

There are a variety of statistical methods for computing interrater reliability. Three of these methods are discussed in the following. The first two, percent agreement and kappa, are relatively simple means for evaluating interrater reliability and can be easily calculated by hand. These methods are useful only if the data are categorical. Both methods require separate calculations for each pair of judges (if there are more than two) and for each item. These are generally considered measures of consensus. The third method, correlation, is generally reserved for continuous measures and typically requires a computer to calculate.

Percent Agreement

Probably the least complex method for calculating interrater reliability is calculating the percent of times that the different judges agree. Using the notation used in Table 1, percent agreement (P) would be calculated as

$$P = \frac{A + D}{T}.$$

For instance, the data in Table 2 might represent ratings by two judges for the presence or absence of tuberculosis from a chest X ray. Cell A represents the number of cases rated as positive by both judges, in this case 80; cell D the number of cases rated as negative by both judges, in this case 20; and cells B and C as the number of cases rated positive by one judge and negative by the other (10 and 5). For this example, $P = 0.870$ or 87% because $(80 + 20)/115 = .870$.

Cohen's Kappa

Cohen's kappa (or kappa) is often favored over percent agreement as a measure of interrater reliability because it removes the effect of agreement due to chance. Kappa scores are therefore generally lower than percent agreement scores computed on the same data but provide a more accurate estimation of how well the raters agree, given the specific task they were performing and the instrument they were using, and are more commonly used to report reliability.

While the percent agreement value is easy to interpret, interpretation of the kappa statistic is more subjective. Kappa values range from −1.0 to 1.0, where a negative scores means that the two judges agreed less than would have been predicted by chance, a zero means there is no agreement beyond that expected by chance, and a 1.0 means that the two judges agreed perfectly. As with the interpretation of correlation coefficients, the decision as to whether, for example, 0.6 represents strong or moderate agreement is somewhat subjective and may depend on what, specifically,

Table 1 Distribution of Ratings by Judges

	Judge 1		
Judge 2	+	−	
+	A	B	R1 = A + B
−	C	D	R2 = C + D
	C1 = A + C	C2 = B + D	T = A + B + C + D

Note: C1 = Column 1, R1 = Row 1, etc.

Table 2 Distribution of Ratings by Judges (Example)

	Judge 1	
Judge 2	+	−
+	80	5
−	10	20

is being examined. However, certain rules of thumb have been developed designating values below 0.4 as fair to poor and values above 0.8 as representing strong agreement.

The calculation for the kappa statistic (κ) is very similar to that used to calculate Person's chi-square. Referring to Table 1 and recalling that P represents percent agreement (above) or observed agreement, E represents expected agreement by chance alone:

$$\kappa = \frac{P - E}{1 - E},$$

where

$$E = \left(\frac{C1}{T}\right)\left(\frac{R1}{T}\right) + \left(\frac{C2}{T}\right)\left(\frac{R2}{T}\right).$$

Using the data from Table 2, $E = .635$ and $P = .870$ (from the previous example). Therefore, kappa $= (.870 - .635)/(1 - .635) = .644$.

Correlation

A third method, Person's correlation coefficient (Pearson's r), typically requires computer calculation and is most appropriate for continuous interval variables. A related statistic, the Spearman rank-order correlation, may be used for data in which the categories are ranked in some natural order or were developed by categorizing a continuous variable. For nominal data such as that in Table 2, the phi statistic is a nonparametric analog to correlation and is appropriate. The formula for phi is

$$\Phi = \frac{(AD - BC)}{\sqrt{r1r2c1c2}}.$$

For these data, phi $= .647$. Correlation measures how much of the variance in one judge's prediction is "explained" by the variance in another judge's prediction. As with kappa, judgment is required when interpreting correlations because there are no absolute rules about how high a correlation represents acceptable reliability. Although statistical tests are commonly performed for correlation coefficients, they partially depend on sample size and only test the null hypothesis that the correlation is 0, rejection of which is not sufficient for establishing the reliability of a measure.

Limitations of Methods for Assessing Reliability

When using the percent agreement or Cohen's kappa methods to assess reliability, each item must be assessed separately. Consequently, there is no way to directly measure overall agreement between raters if multiple items were rated. This is particularly a problem if exact agreement on one item was less important than an overall agreement on, for example, a multi-item scale. Similarly, if more than two judges provided ratings, separate analyses would need to be done for each pair of judges. There are methods using correlation measures of reliability that allow for these multiple comparisons such as intraclass correlation and Cronbach's alpha.

Interclass correlation (ICC) can be used to assess the reliability of both single ratings and mean ratings and is useful when there are more than two raters. Based on an ANOVA, ICC provides a ratio of homogeneity of ratings between subjects (numerator) to the homogeneity of ratings within subjects (denominator).

Cronbach's alpha is generally used to measure the internal consistency of multiple measures; however, it can also serve to provide a summary of interrater reliability identifying how similarly a set a judges rate multiple measures representing a single construct.

Each type of interrater reliability assessment is sensitive in different ways depending on the distribution of the event under consideration in the population. For example, if the event is common, then there is a greater likelihood that different raters may agree by chance alone. Percent agreement is influenced by the incidence of exposure. The more common the exposure is, the more likely it is that two different judges may agree by chance alone. Very different results could come from this situation particularly if there is, for example, a tendency to agree that someone is positively exposed more than if someone is negatively exposed.

Assumptions in Assessing Interrater Reliability

Perhaps the most important assumption for measures of interrater reliability is that each judge's observations are independent. This is important as these estimates are, in actuality, assessing if the observers are equal, not if tests are equal. Similarly, these estimates assume that the errors in observations are uncorrelated. Finally, these estimates assume that the coding performed by the judges is consistent.

When to Use Interrater Reliability

In addition to providing a summary measure of agreement between judges, interrater reliability also offers the opportunity to identify problems with measurement particularly when divergent results between different judges is assessed in regards to the source of the opposing observations.

In practice, interrater reliability can be used in a variety of situations. It may be used for evaluation of a method of health care delivery. For example, if a researcher wanted to determine whether a new method for children with asthma to use in checking airflow was better than an older method, half of the children could be taught the new method while the other half used only the older method. Two judges could observe all the children assessing their airflow and rate them on how successfully they performed this task. If there is generally high agreement that those children taught the new method checked their airflow more successfully than those taught the old method, then the new method could be adopted.

Interrater reliability can also be used for screening purposes. For example, to assess the potential validity of a brief questionnaire to be completed by patients to assess diabetes risk, a research may give 100 patients this screening survey to complete. Two judges would review the surveys and categorize each patient in terms of risk. Then the categorization of each patient by each judge would be compared to measure the level of interrater agreement. Strong agreement between the raters would suggest that the screening items are validly measuring diabetes risk and that the judge's assessments of risk are objective.

In fields such as psychology, interrater reliability is often used when making observations about an individual. For example, two judges may observe a child at play to measure the child's exhibition of aggressive tendencies, for example, by counting the number of aggressive acts performed by the child in an hour.

Finally, although the examples presented here used quantitative methods to estimate interrater reliability, this can also be a useful tool in qualitative research. For example, when a researcher is conducting a content analysis of educational health pamphlets, the thematic issues that seemed strongest as determined by one judge can be compared with those identified by a second judge.

—Eve Waltermaurer

See also Kappa; Pearson Correlation Coefficient; Reliability

Further Readings

Bullock, G., Kovacs, G., Macdonald, K., & Story, B. A. (1999). Evaluating procedural skills competence: Interrater reliability of expert and non-expert observers. *Academic Medicine*, *74*, 76–78.

Endicott, J., Tracy, K., Burt, D., Olson, E., & Coccaro, E. F. (2002). A novel approach to assess inter-rater reliability in the use of the Overt Aggression Scale-Modified. *Psychiatry Research*, *112*, 153–159.

Rosenzweig, S., Brigham, T. P., Snyder, R. D., Xu, G., & McDonald, A. J. (1999). Assessing emergency medicine resident communication skills using videotaped patient encounters: Gaps in inter-rater reliability. *Journal of Emergency Medicine*, *17*, 355–361.

Saito, Y., Sozu, T., Hamada, C., & Yoshimura, I. (2006). Effective number of subjects and number of raters for inter-rater reliability studies. *Statistics in Medicine*, *25*, 1547–1560.

Intervention Studies

Intervention studies are broadly defined as those that include an action or component intended to promote health or prevent disease by influencing, affecting, or manipulating environmental, behavioral, or etiological aspects of a disease. In other words, an intervention is designed with the express intention of improving health at the individual or group level. One of the most famous interventions in modern biomedicine occurred in 1854 when John Snow, a medical doctor, advocated removing the pump handle from a public water pump located on London's Broad Street. Snow believed that the Broad Street pump was the source of contaminated water contributing to a cholera

epidemic in the city. Snow's earlier observations of cholera convinced him that the disease was transmitted through direct physical contact with contaminated individuals or other contaminated sources. Therefore, he believed, the simple act of removing the handle from the Broad Street public pump, which was located within the community where most of the cholera infections were clustered, would prevent further exposure to local residents.

Snow's idea was controversial among the mainstream public as well as among many of his medical colleagues, for it was commonly accepted at the time that cholera was spread by miasma (i.e., bad air). However, removing the handle from the Broad Street pump, which rendered it unusable and forced local residents to find alternative water sources, coincided with a dramatic decrease in cholera outbreaks in the London neighborhood. It was later determined that the Broad Street pump was adjacent to a sewage catchment area that had been contaminated by *Vibrio cholerae* bacteria, and the cholera-laden sewage had subsequently seeped into the public water supply. Although it has been suggested that the epidemic was already receding before Snow removed the handle from Broad Street's pump, the phrase "removing the pump handle" remains a popular metaphor among public health practitioners more than 100 years after Snow's intervention challenged both standard medical authority and popular opinion of his day. The Broad Street pump episode came to reflect the importance of the emerging discipline of epidemiology as a tool for enhancing public health through identification of disease-causing agents and development of intervention studies able to systematically measure the efficacy of treatments to attenuate or prevent disease.

Conducting Interventions

Intervention studies are designed to measure the efficacy of a procedure or other action on a particular health problem. Interventions are conducted by first identifying a population vulnerable to the particular health issue of interest—the target population. Individuals who belong to the target population are then recruited to participate in the study based on a series of inclusion and exclusion criteria. The inclusion criteria include specific conditions or factors that the investigative team is seeking to influence, or factors for which they will control (e.g., age, gender, or socioeconomic status). For example, eligibility to enroll in an HIV prevention intervention study to address heterosexual transmission might be limited to only those women who reported being HIV-negative or serostatus unknown and who reported unprotected sexual intercourse with at least one male partner who was HIV-positive or an injection drug user. Therefore, women who reported being HIV-positive would be ineligible to participate, as they would not meet an inclusion criterion. In such a study, the investigative team might choose to include or exclude women who reported that they were trying to become pregnant.

Evaluating an intervention is usually based on some variation of the classic control group design model, in which, on enrollment in the study, each participant is assigned to either the control group (also known as the comparison group) or the intervention group (also known as the experimental group). The control group does not receive the intervention, and the experimental group receives the intervention. Outcome data are then collected from both groups, and analyses of the data form the basis for evaluating the intervention's efficacy. For example, a longitudinal study of an obesity reduction intervention might include overweight individuals who are assigned at baseline to the control group or to the group to receive an intervention to increase exercise and make healthier food choices. Baseline data collected from all the participants (i.e., the comparison group and the experimental group) would include the variables of interest, such as weight, body mass index, and blood pressure. After a specified time (say, 6 months after baseline), data would again be collected from all participants on the variables of interest and compared to see whether there were differences (e.g., weight loss or a lowering of blood pressure) between those participants who received the obesity reduction intervention (the experimental group) and those participants who did not receive the intervention (the control group).

Types of Intervention Studies

Intervention studies include nonrandomized studies (NRSs) and randomized studies, which are often set up as randomized controlled trials (RCTs). Two disadvantages of NRSs are that they lend themselves to selection bias and that they do not provide definitive evidence on effectiveness. Consequently, they are open to critique with regard to evaluation of effectiveness and selection bias. Despite these threats to generalizability and validity, NRSs are appropriate when

studying small, special, hidden, or transient populations, such as injection drug users or tourists on holiday. RCTs are considered the gold standard when evaluating the effectiveness of an intervention. RCTs are often used in large pharmaceutical trials to test the efficacy of a new drug. However, noncompliance and investigator bias are threats to the validity of RCTs that the investigators should address early in the process of developing a study proposal.

Noncompliance to randomization among study participants can be grouped into three categories: never-takers, always-takers, and defiers. Never-takers are participants who are assigned to the experimental group but fail to receive the intervention—commonly because they miss the appointment. Always-takers are participants who are assigned to the control group who manage to receive the intervention. Defiers are participants who do not follow the instructions for the group to which they are assigned but instead engage in behaviors opposite to those that they have been asked to adhere to.

Investigator bias, which is often unintentional and unrecognized, creeps into a study when, for example, the research team knows which participants have been randomized into the experimental group and treats these individuals differently than those participants in the comparison group. Interventions that are designed as blind studies can reduce investigator bias. In a single-blind study, the researchers know the intervention protocol, methods, and which participants have been assigned to a particular group (control or experimental), but the participants do not know their assignment. In a double-blind study, neither the researchers nor the participants know the intervention protocol, methods, and participant assignments. In a triple-blind study, the intervention protocol, methods, and participant assignments are unknown to the researcher, the participants, and the researchers who contribute to conducting the data analysis.

Ethics and Intervention Studies

Intervention studies must be both scientifically rigorous and ethically sound. Being mindful of ethical concerns has become more routinized and standardized with the inception of institutional review boards (IRBs). Yet adhering to IRB guidelines is only one of several important aspects of conducting ethical research, especially among vulnerable populations such as underserved minorities and children. For example, how do researchers address the ethics of assigning medically underserved minority participants with chronic illness to a standard-care control group when the standard of care the participants currently receive is suboptimal compared with that available to members of the general population who can access better health care? A study that has generated dialog between researchers, IRBs, ethicists, and the general public is the Wisconsin Cystic Fibrosis Newborn Screening Trial, a longitudinal epidemiologic study conducted between 1985 and 1998.

One of the largest public health studies ever conducted in the United States, the trial involved routine newborn screening of more than 650,000 infants for cystic fibrosis (CF) by identifying elevated levels of immunoreactive trypsinogen (IRT). The newborns were alternatively assigned to a screening diagnosis group or a symptom diagnosis group, and therefore each infant had a 50% chance of being in the screening group or the symptom group. The parents of those infants assigned to the screening group with elevated IRT levels were informed of the results and encouraged to have the infant undergo a CF diagnostic test (usually within 6 weeks after birth). Parents of children subsequently diagnosed with CF were asked to enroll their infants in the study to provide routine CF care to evaluate the effect of an early screening intervention in the treatment of CF.

The infants assigned to the symptom group did not have their IRT data fully processed or automatically disclosed to their parents. Instead, the IRT data were stored in a database and released to the parents when one of the following criteria was met: (1) the parents requested the IRT data, (2) the child was later diagnosed with CF, or (3) the child turned 4 years old. All children in the symptom group who were later diagnosed with CF were then offered enrollment in the study. It has been argued that all infants who received the screening should have had their IRT data fully processed and automatically disclosed to their parents, as early detection of CF can help ameliorate the cumulative effects of the disease. Although this controversial study was found to have scientific merit, it has helped draw attention to the issues surrounding parental consent and rights to disclosure.

Syndemics and Intervention Studies

An emerging area of importance in public health promotion and intervention research is the study of

syndemics (i.e., linked, connected, or co-occurring epidemics within a population). This area of interdisciplinary research and scholarship includes social epidemiologists, anthropologists, psychologists, and other social and behavioral scientists and practitioners who seek to bring the best of research and policy to bear on disease prevention. In the 1990s, Merrill Singer identified the co-occurrence of epidemics within particular populations. Akin to a domino effect, in which a line of dominoes stacked in a row fall down in sequence after the first domino is triggered to fall into the second, so one epidemic may be linked to or triggered by another epidemic, which is then linked to or triggered by yet another epidemic. For instance, HIV and hepatitis C are syndemic among low-income injection drug users. Studies informed by an understanding of population syndemics may yield more efficacious interventions, which in turn will inform evidence-based public health policy and practice.

—Kathleen Ragsdale

See also Applied Epidemiology; Drug Abuse and Dependence, Epidemiology of; Epidemiology, History of; Ethics in Human Subjects Research; Randomization; Social Epidemiology

Further Readings

Centers for Disease Control and Prevention. (2006). *Syndemics prevention network*. Retrieved August 3, 2006, from http://www.cdc.gov/syndemics/index.htm.

Fisher, A. A., James, R., & Foreit, J. R. (2002). *Designing HIV/AIDS intervention studies: An operations research handbook.* New York: Population Council.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods, 7*, 178–193.

Lesaffre, B. S., & Verbeke, G. (2005). Clinical trials and intervention studies: The intervention study. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science.* New York: Wiley. Retrieved May 17, 2006, from http://www.wiley.com//legacy/wileychi/eosbs/pdfs/bsa099.pdf.

School of Public Health, University of California, Los Angeles. (2001, August). *Removal of the pump handle*. Retrieved May 17, 2006, from http://www.ph.ucla.edu/epi/snow/removal.html.

Taylor, H., & Wilfond, B. (2004). Ethical issues in newborn screening research: Lessons from the Wisconsin Cystic Fibrosis Trial. *Journal of Pediatrics, 145*, 292–296.

Interview Techniques

Interviews are used to collect both quantitative and qualitative data. They are particularly important in qualitative studies, where they are the principal form of data collection. Interview techniques used in qualitative data collection vary from unstructured interviews or narratives to open-ended, semistructured interviews. Both unstructured and semistructured interviews may be distinguished from formal or structured interviews, in which a fixed list of questions is used and the emphasis is on collecting data in a standardized manner. Structured interviews are used to collect data for many surveys, including the Behavioral Risk Factors Surveillance System and the National Health Interview System; however, because the data collection process is much more straightforward in structured interviews, this entry concentrates on techniques used in unstructured and semistructured interviews.

Regardless of the type of interview, the quality of a research project relies heavily on the researcher's ability to elicit accurate information from the participant. A successful qualitative interview is similar to a warm and personal sharing of a confidence with a trusted friend. Even in structured interviews, it is important to establish a relationship of trust with the subject. Several procedures may assist in establishing a successful interview process. First, whenever possible, the participants should choose the setting. Wherever they have selected, it should be private with little opportunity for interruption. A small table nearby will allow a tape recorder and microphone to be placed with less intrusiveness. The researcher should begin with small talk to relax the participant and start the interview with the consent procedures and demographic information. This simply gets the participant used to talking with the interviewer. Depending on the type of interview selected, once the participant is relaxed and comfortable with the interview process, the researcher should begin administering the structured interview questions or, in an unstructured interview, allow the participant to lead the interview and focus on telling his or her story.

The Unstructured Interview

The unstructured interview is used when the researcher knows little about the topic and is learning about the

subject matter as the interviews progress. Rather than using a set of prepared questions, the researcher encourages the participant to tell their story with minimal interruption, especially during the first interview.

The questions that are asked make a difference in the quality of the information. If questions asked are theoretically based, the answers from the participants will be theoretical "reports" rather than life stories. Each participant has a story concerning the topic of interest; therefore, the researcher's role is to ask questions that invite the participant to tell the story he or she most wants to tell.

Even if the researcher aims to invite others to tell their story, it's not always clear in advance which questions will serve as an invitation. So to start, researchers need to work their way toward some sense of the broad parameters of the participant's life experiences, which make this group of people's life experiences interesting in the first place.

The participants should be provided with a context for the interview by an explanation or briefing before and a debriefing afterward. The context is introduced before the interview starts with a brief purpose of the interview, the use of the tape recorder, and asking participants if they have any questions before the interview begins.

When the researcher is conducting an unstructured, interactive interview, the main technique is to listen intently. Assuming an active listening stance, accompanied by encouraging nods of the head and "Mmmmm" and "I see" spoken in a noncommittal way, encourages the participant to continue with the story. The problem with the researcher's asking several questions is that the participant does not fully focus on his or her story, stops the narration, and waits for the researcher to ask the next question. In the unstructured interview, the topic is still unknown, so the researcher will not know what questions to ask; therefore, it is a threat to validity to move into the question-asking mode during the interview. In semistructured interviews, some questions with open-ended stems such as "Tell me about … " and "Please explain … " may be used to elicit the best responses.

While telling their stories, participants may relive their experiences, including the emotional responses to life events. For this reason, if the research topic is stressful, the researcher needs to be prepared for sadness or anger and possibly tears. It may happen that the participant wants to tell the researcher something stressful or upsetting but backs off the topic, only to approach it later and tell the researcher about the event. This is an example of how the participant controls the pace of the interview. Requesting participants to tell their story in a certain order or before they are ready may upset them, and they may refuse to continue with the interview. Therefore, the researcher needs to be comfortable with silence, emotion, and allowing participants to set the pace.

At the end of the interview, there may be some tension or anxiety because the subject has been open about often personal and emotional experiences or may have a feeling of emptiness that much has been given and little has been received in return. The researcher may avoid this by rounding off the interview and mentioning some main points obtained by the interview. The last question of the interview should be, "Is there anything else I should have asked you?" Often this lends itself to some of the most revealing data. At the end of the interview, the researcher may ask participants if they might be contacted again (as was included on the consent form) should the researcher think of other questions to cover. This allows for any gaps in the data to be resolved and murky areas to be explained.

The Semistructured Interview

The semistructured interview is used when the researcher knows most of the questions he or she wants to ask but cannot predict the answers. This structure ensures the researcher will obtain all the information required about "how" an event occurred, at the same time giving the participants freedom to tell their story and illustrate concepts.

Questions that the researcher has thought through ahead of time tend to work best when they are logical within the domain of the topic and address only one aspect at a time. The interviewer may use the funnel approach or the nonfunnel approach when asking questions. In the funnel approach, the researcher starts with broad, open questions and narrows them down to more specific probes. This is helpful when interviewers want to avoid imposing their own frame of reference on the respondent. The inverted funnel approach starts with specific questions and moves to the more general. This technique is helpful when the participant is not motivated to speak easily. "Double-barreled" questions that include two concepts or time frames in one question tend to confuse the participant and stop the flow of the interview. For example,

A: Tell me what methods you used to cope with your wife's cancer both when you first found out about the diagnosis and then again when it recurred.

B: Well the first time was scary and then again the second time we were at the doctor's office when we found out. You mean the first time or the second time?

It is best to ask about one concept at a time, allow the participants to explore it thoroughly, and then move to the next event or time frame. Probes or additional questions will elicit further information. The quality of the study relies on the quality of the questions; however, the quality of the interviewer as an instrument of the study is important also.

To invite a full and rich life story from participants, several general principles of good interviewing technique may be used. For example, it is important to give the participant all of one's attention without fidgeting or rushing through the topic. Likewise, researchers need to be able to tolerate silence in comfort. During silent periods, participants may be thinking or working to keep their emotions under control. They will continue when they are ready.

One of the most important tactics in interviewing is communicating the idea that the participant's views are acceptable and important. The interviewer must accept the participant's answers at face value and communicate that to the participant. However, an astute researcher continuously observes participants' nonverbal body language and looks for mismatching signals between what participants say and how they behave while they are speaking. When a mismatch occurs, the researcher gently probes for clarification while maintaining warmth and acceptance. The interview is not a reciprocal interaction of two equal partners. There is a definite asymmetry of power. The researcher defines the situation, introduces the topic of interest, and through broad questions and more specific probes, guides the course of the interview. The researcher listens for gaps, silences, or contradictions and reiterates the invitation through probes and questions that encourage a fuller telling of the story.

Researcher Traits

There are three personality traits that are assets to the interviewer: flexibility, intelligence, and emotional security. Flexibility enables the interviewer to assume an active or passive role when needed to facilitate communication and keep the interview on track. Intelligence helps the researcher determine the objectives of the interview, remember what was said, probe appropriately, and ask for clarification of gaps in the story. The emotionally secure researcher is able to communicate warmth and put the participant at ease.

The interview situation may be characterized by positive feelings of a common intellectual curiosity and a reciprocal respect. It is a conversation in which data arise in an interpersonal relationship, coauthored by the researcher and the participant, producing new, trustworthy, and interesting knowledge.

—Mary M. Lopez

See also Ethics in Human Subjects Research; Qualitative Methods in Epidemiology; Survey Research Methods

Further Readings

Oishi, S. (2002). *How to conduct in-person interviews for surveys* (2nd ed.). Thousand Oaks, CA: Sage.
Rubin, H. (2004). *Qualitative interviewing: The art of hearing data* (2nd ed.). Thousand Oaks, CA: Sage.
Weingraf, T. (2001). *Qualitative research interviewing: Bibliographic narrative and semi-structured methods.* Thousand Oaks, CA: Sage.

Intimate Partner Violence

Intimate partner violence (IPV), also called *domestic violence*, is defined by the American Medical Association as abuse (i.e., physical violence, sexual assault, or psychological abuse) to an individual perpetrated by a current or former intimate partner. While this general definition is widely accepted, there are many nuances in the inclusion criteria for severity and context of IPV. These nuances, in turn, lead to varying estimates of IPV prevalence, risk factors, and outcomes. This entry discusses these estimates, examines health-related and other outcomes of IPV, considers issues related to prevention and intervention, and discusses concerns related to the safety of study participants in IVP research.

Definitions

Most research definitions of IPV focus on acts. For physical assault, these acts may include pushing,

slapping, shoving, throwing, grabbing, biting, shaking, poking, punching, hitting, kicking, burning, threatening physical harm, and using a weapon such as a knife or gun. Context is also important: Some studies consider defensive acts to be IPV, but others do not. For sexual assault, these acts may include threat or attempted rape, rape, demands that sex be videoed, and touching in an unwanted manner. Emotional or psychological abuse and controlling behaviors may include constantly putting down and insulting, lying, saying no one would want the partner, calling him or her crazy, isolating the partner from family and friends, blaming him or her for anything that is not perfect, and causing problems at the partner's work. While most definitions of IPV focus on acts of IPV, some focus more on the experience of the victim, such as "do you feel like you are walking on eggshells at home?" Finally, some research simply asks individuals if they are abused and leaves it up to the respondent to define *abuse*.

Prevalence of IPV

The estimated prevalence of IPV depends on the definition used and the population sampled. Most studies focus on the prevalence of IPV against women. Internationally, the annual prevalence of physical or sexual assault against women by a partner is approximately 20%; however, the range is substantial, with country prevalences ranging between 3% and more than 50%. Country estimates of physical or sexual assault ever occurring to a woman ranged from 15% to 70%, and 4% to 50% for severe assaults. The lowest prevalences tend to be in Western Europe, the United States, Canada, and some developed Asian countries (e.g., Japan). The areas with high prevalence tend to be developing countries in Africa, South America, and Asia. Worldwide, women are at greater risk of assault from a partner than from a stranger.

In the United States, measures of IPV that limit the focus to acts of violence find that women and men experience equivalent violence by their heterosexual partners: 1% to 10% report any physical assault in the past year. However, violence that results in injury is far more likely to occur to women. About 1 million women report severe physical assaults annually. These findings have raised substantial conversation and disagreement in the research community. The growing hypothesis is that two or more distinctly different types of IPV are being measured. One is bilateral violence (i.e., both partners being violent), which may be considered acceptable behavior within the social norms of some communities. The other is violence used as a means to gain and maintain control of a partner. This second form is more often perpetrated by men against women.

Risk and Protective Factors

Factors that describe a person's risk of being victimized can be categorized into four groups: *victim* characteristics, experiences, and support; *perpetrator* characteristics, experiences, and support; *relationship* characteristics; and *societal norms*. Most IPV research has focused on violence by men against women, and the following observations are based on this body of knowledge.

While victim characteristics may share some common elements, relatively few have predictive value. This makes intuitive sense given that victims do not cause violence. Women victims tend to be young and are more likely to be from a family where her mother was victimized and thus may see victimization as the norm (e.g., "don't all husbands hit their wives?"). Women who are poorer, from a minority group, or have been victimized in the past are also more likely to be victimized. Women are less likely to be victimized if they have a strong social support network. Men are more likely to perpetrate violence if they are young, poorly educated, unemployed, drink alcohol to excess or use drugs, have psychiatric disorders (e.g., depression, personality disorder), have low self-esteem, have a history of child abuse, believe in gender-specific roles, carry a knife or gun, or have a need to be controlling their environment. Relationship risk factors include the woman having more education, higher job attainment, or more income than her male partner; economic stress; poor communication skills; and a history of marital tension. The primary societal risk factor for IPV is a normative acceptance: the lack of negative community consequences for IPV and emphasis on a traditional patriarchic social structure.

Health Outcomes Related to IPV

Intimate partner violence has multiple health-related sequelae. While injuries and homicide are obvious outcomes, IPV impacts a continuum of women's health issues. Victims of IPV are more likely to

be depressed, have anxiety, experience posttraumatic stress disorder, have low self-esteem, experience somatization, and commit suicide. Multiple studies have found that mental acuity is markedly reduced for victims of IPV compared with others. A myriad of physical symptoms are associated with IPV: loss of appetite, frequent or serious bruises, nightmares, vaginal discharge, eating binges or self-induced vomiting, diarrhea, broken bones, sprains or serious cuts, pain in the pelvis or genitals, fainting, abdominal pain, breast pain, frequent or serious headaches, difficulty in passing urine, chest pain, problems with sleeping, shortness of breath, and constipation. Recent research has found that IPV is associated with multiple chronic diseases, such as heart disease and diabetes, likely caused by an elevated propensity for risky health behaviors. For instance, women who experience IPV report higher levels of smoking, a sedentary lifestyle, and poor nutrition compared with other women.

Less research has been conducted on the health sequelae for men who experience violence in a relationship. The limited data suggest that men who are either a victim or perpetrator of IPV experience a similarly wide range of health sequelae. Male perpetrators more frequently report mental and physical health problems than men in nonviolent relationships. Men reporting victimization or both victimization and perpetration of IPV are at the highest risk of having multiple health problems. These findings lead to the obvious conclusion that violence is bad for your health.

Other Outcomes Related to IPV

Women who are abused are more likely to miss work and have a higher job turnover than women who are not abused. Abused women are more likely to have an unintended pregnancy. Children in a home with IPV are more likely to have health problems that can last into adulthood, become abusers or victims, and have behavioral problems. Homelessness can be a consequence of IPV for those women and children who do not have the resources to reestablish themselves in a similar lifestyle. From an economic perspective, IPV is very costly. Based on a National Institute of Justice report, the annual medical costs for adult IPV were estimated to be $1.8 billion. The total U.S. annual societal cost was estimated at $67 billion, including tangible property loss and quality of life impact.

Prevention and Intervention for IPV

Multifaceted prevention programs have been implemented in the United States during the past four decades, many of them with roots in the women's movement. Social marketing campaigns educate communities about IPV and the importance of structuring society on the premises of mutual respect, equality, and trust for all. Prevention programs target high school students who are starting to date. Educational programs focus on the need for mutual respect and often include exercises in role modeling to help students practice the skills needed to negotiate difficult relationship issues. Legislation has also been passed to criminalize IPV in many states. Until the 1980s, in some states in the United States it was impossible for men to be prosecuted for raping their wives. The convention was that men should be able to have sex with their wives whenever they want. Similarly, it was difficult to prosecute a man for assaulting his wife although a similar act would result in criminal charges if committed against a stranger. These campaigns and legislative changes have reduced the reported physical assaults in relationships. However, some scholars suggest that this decrease in assault may be somewhat offset by a transition in the mode of violence from physical to emotional abuse.

Intervention programs for IPV typically focus on settings where abused women seek care, such as police stations, battered women's shelters, and medical clinics or hospitals. Epidemiologists are particularly interested in studying IPV intervention programs in medical care settings due to the numerous health events abused women have, but screening all women for IPV victimization has been controversial. Screening is recommended by the AMA and many other medical organizations but has not received a recommendation from the U.S. Preventive Services Task Force, charged with evaluating the effectiveness of screening and counseling programs in primary care. The task force is waiting for more evidence from randomized trials before issuing guidance.

Randomized trials focus on screening followed by a variety of interventions, ranging from physician or provider response training to multifaceted interventions. The multifaceted interventions are based on the theory of chronic disease management, which recognizes that the best outcomes are achieved when (1) patients are educated and given the support to set

goals and make choices, (2) multidisciplinary teams work together, (3) progress is actively monitored, (4) a range of therapeutic modalities are available to support active coping styles, and (5) triage is used to determine the intensity of support according to the severity of abuse and difficulty in seeking safety.

Safety Issues in the Study of IPV

When epidemiologists study IPV, as much attention must be paid to the safety of study participants as to the integrity of the scientific methodology. Some victims of IPV are at risk of serious harm and even homicide, and thus tremendous care must be taken to make sure that research does not increase the risk to them in the process of trying to help. Multiple safety protocols must be developed and implemented, from the clinic setting to follow-up research calls at home. Women are often at most risk when they are leaving an abusive relationship or making changes to be safer. For this reason, education of the abused woman about safety and available resources is an essential first step. The research staff need to be particularly well trained so that the protocol can be modified instantly when circumstances change (e.g., the abusive partner unexpectedly joins the woman at the clinic during study data collection).

The prevalence and disparate negative outcomes of IPV underscore the large burden of suffering associated with relationship violence. The complexity of IPV requires careful study designs and intervention programs. Safety issues heighten the need to provide unusually stringent safety protocols to minimize harm to women. Despite these risks, it is imperative that epidemiologic researchers work with domestic violence advocates and social scientists to develop and measure interventions that reduce the public health burden of IPV.

—Louise-Anne McNutt

See also Child Abuse; Screening; Violence as a Public Health Issue; Women's Health Issues

Further Readings

Campbell, J. C., Jones, A. S., Dienemann, J., Kub, J., Schollenberger, J., O'Campo, P., et al. (2002). Intimate partner violence and physical health consequences. *Archives of Internal Medicine*, *162*, 1157–1163.

Coker, A. L., Smith, P. H., Thompson, M. P., McKeown, R. E., Bethea, L., & Davis, K. E. (2002). Social support protects against the negative effects of partner violence on mental health. *Journal of Womens Health and Gender-Based Medicine*, *11*, 465–476.

García-Moreno, C., Jansen, H., Ellsberg, M., Heise, L., & Watts, C. (2005). *WHO multi-country study on women's health and domestic violence against women: Initial results on prevalence, health outcomes and women's responses*. Geneva, Switzerland: World Health Organization.

Tjaden, P., & Thoennes, N. (2000). *Full report of the prevalence, incidence, and consequences of violence against women: Findings from the National Violence Against Women Survey*. Publication No. NCJ183781. Washington, DC: U.S. Department of Justice. Retrieved August 8, 2007, from www.ncjrs.org/txtfiles1/nij/183781.txt.

Web Sites

Centers for Disease Control and Prevention, Injury Center. Intimate partner violence: Overview: http://www.cdc.gov/ncipc/factsheets/ipvfacts.htm.

Item Response Theory

See Measurement

J

JENNER, EDWARD
(1749–1823)

Edward Jenner is best known as the inventor of the smallpox vaccination. Although little of the basic science of the smallpox virus or human immunity was known at the time, Jenner tested a hypothesis formulated by epidemiologic observation in a clinical trial (with an *n* of 1) and established its validity. His research provided the model for the next 150 years of human medical research.

Jenner was born, and spent most of his career as a country doctor, in the English county of Gloucestershire. Orphaned at age 5, he was sent to boarding school at age 8, where he was inoculated with smallpox and reportedly traumatized by the experience. At the age of 12, he began surgical training.

Jenner observed that most milkmaids and cowmen did not get smallpox, a widespread disease that killed many infants and small children and often left the few who did survive deaf, blind, and horribly scarred. Jenner established that these immunized individuals had all suffered from cowpox, a relatively mild disorder characterized by skin blisters, and hypothesized that having had cowpox somehow prevented one from acquiring smallpox. In 1796, Jenner inoculated cowpox lymph derived from a cowpox vesicle on the skin of dairymaid Sarah Nelmes into a young boy named James Phipps. Phipps's cowpox took a normal course, and he recovered. A month later, Jenner inoculated the boy with the smallpox virus and no reaction occurred. After repeating the inoculation a few months later, smallpox still did not develop.

Jenner was not the first to suggest that cowpox infection provided immunity to smallpox or the first to attempt cowpox inoculation to prevent smallpox. He was however the first scientist to demonstrate by experiment that naturally acquired cowpox protected against smallpox. Jenner coined the word *vaccination* for his treatment (*vacca* means cow in Latin). Louis Pasteur eventually adopted this term for immunization against any disease.

Jenner eventually gave up his life as a country doctor and spent considerable time and money promoting vaccination. His 1798 book *An Inquiry Into the Causes and Effects of the Variolae Vaccinae* received favorable as well as critical reviews. Controversy dominated the rest of Jenner's life. Although he was often ridiculed for his work on immunization, scientific advances in the areas of germ theory, viruses, and human immunity support Jenner's main conclusions. Ultimately, Jenner's vaccination enabled the eradication of smallpox. In 1980, the World Health Assembly declared the world free from endemic smallpox as a direct result of Jenner's discovery.

—Emily E. Anderson

See also Disease Eradication; Smallpox; Vaccination

Further Readings

Bazin, H. (2000). *The eradication of smallpox: Edward Jenner and the first and only eradication of a human infectious disease* (A. Morgan & G. Morgan, Trans.). San Diego, CA: Academic Press.

Mullin, D. (2003). Prometheus in Gloucestershire: Edward Jenner, 1749–1823. *Journal of Allergy and Clinical Immunology*, *112*(4), 810–814.

Web Sites

The Jenner Museum: http://www.jennermuseum.com.

Journals, Epidemiological

A journal is a regularly published scholarly, peer-reviewed or refereed publication in print or online format. There are many journals important to the field of epidemiology. This entry reviews the history of epidemiological journals and identifies those considered by the Medical Library Association to be essential and minimal core journals.

History

Journals explicitly focused on epidemiology are a relatively new phenomenon; prior to their development, epidemiological articles were published in a variety of public health and clinical journals. When, in 1965, the *American Journal of Epidemiology* changed its name from *The American Journal of Hygiene*, it stated in an editorial that

> so far as we are aware there is no journal in the English language which has the word epidemiology in its title. Epidemiology is both a method and a substantive field. . . . a journal devoted to epidemiology will fill a distinct need. ("Change in Name," 1965, p. 1)

The editorial further stated that the change in name reflected a change in scope, extending coverage to studies of the epidemiology of noninfectious conditions to make the journal become more broadly representative of the field of epidemiology.

Today, several dozen journals are published that focus specifically on epidemiology, clinical epidemiology, and biostatistical methods in disease surveillance and description. Many focus on particular fields within epidemiology, reflecting the development of narrow fields of investigation within epidemiology: This specialization is reflected in titles such as *Genetic Epidemiology* (Wiley), *Opthalmic Epidemiology* (Taylor & Francis), and *Paediatric and Perinatal Epidemiology* (Blackwell). The PubMed Journal Database, which includes the journals indexed in the National Library of Medicine's widely used PubMed database, includes 57 journals in a variety of languages classified as belonging to the subject "epidemiology," including both general journals such as the *American Journal of Epidemiology* and specialized journals on epidemiologic subtopics such as *Neuroepidemiology* and *Genetic Epidemiology.*

Core Journals

According to the Public Health/Health Administration (PH/HA) Section of the Medical Library Association, the following are the "Essential Core" journals in the field of epidemiology (with publishers in parentheses). These are defined by PH/HA as those titles "essential for a collection that supports a program with subject specialization in this area." In the list of journals provided below, those marked with an asterisk (*) are considered "Minimal Core," designating them as first-purchase recommendations for those developing an epidemiology journal collection.

* *American Journal of Epidemiology* (Oxford University Press)

Annals of Epidemiology (Elsevier Science)

Cancer Causes and Control (Springer-Verlag)

Cancer Epidemiology, Biomarkers & Prevention (American Association for Cancer Research)

Epidemiologic Reviews (Oxford University Press)

* *Epidemiology* (Lippincott Williams & Wilkins)

International Journal of Epidemiology (Oxford University Press)

* *Journal of Clinical Epidemiology* (Elsevier Science)

* *Journal of Epidemiology and Community Health* (BMJ Publishing)

Lancet Infectious Diseases (Elsevier Science)

In addition to the above titles, nearly all medical and public health journals include some articles pertinent to epidemiology. Clinical epidemiology is indispensable to clinical medicine as its principles form the basis for the evaluation of clinical trials and evidence-based medicine. Hence, major medical journals such as *The Journal of the American Medical Association (JAMA), The New England Journal of*

Medicine, and *The Lancet* are also of primary importance to epidemiologists, as are the major public health journals such as *American Journal of Public Health* and *Public Health Reports.* Another important source of epidemiological information is *Morbidity and Mortality Weekly Reports (MMWR)*, published by the Centers for Disease Control and Prevention.

Minimal Core Journals

The *American Journal of Epidemiology* is currently published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. It began publication in 1921 as the *American Journal of Hygiene;* as noted above, the title was changed in 1965. It publishes primarily empirical research findings, methodological developments, and opinion pieces related to epidemiology. *The American Journal of Epidemiology* is one of the top-tier journals in all public health. In 1978, the *Journal of Epidemiology and Community Health* arose from the *British Journal of Social and Preventive Medicine*, first issued in 1947 as the *British Journal of Social Medicine.* The focus of this journal is, as its title implies, the methodology of research of health in communities. The *Journal of Clinical Epidemiology* also began under another title, published as the *Journal of Chronic Diseases* from 1955 to 1987. In 1988, the editors of this journal changed the name to more closely reflect what the journal had become, an outlet for publishing the methodological research in clinical epidemiology and health services research. And last, in 1990 a new journal, *Epidemiology*, was introduced by Blackwell Scientific, to meet the need of an expanding body of epidemiologic research looking for a place to publish. This fledgling publication quickly became one of the most respected journals in the field, with a focus leaning toward, but not limited to, environmental epidemiology.

—Marie T. Ascher

See also Epidemiology, History of; Journals, Public Health; Publication Bias

Further Readings

Change in Name [Editorial]. (1965). *American Journal of Epidemiology*, *81*(1), 1.

Public Health/Health Administration Section of the Medical Library Association. (2006). *Core public health journal project: Version 2.0* [Electronic version]. Retrieved January 11, 2007, from http://publichealth.yale.edu/phlibrary/phjournals/v2.

Journals, Public Health

Publishing in public health takes place in many forms, but the most common prestigious format is an article in an academic public health journal. A journal is a regularly published scholarly, peer-reviewed or refereed publication. Journals traditionally have been produced in print format, but most publishers now offer online electronic versions of their journals as well. The realm of public health journals is as diverse and multidisciplinary as the field of public health itself. Articles relevant to public health researchers and practitioners are found in the major medical journals, in journals in the social sciences and psychology, in the statistics literature, in the environmental and nutrition literature, and in journals specific to the field of public health and its specific disciplines.

Public health professionals read journals to keep up with current trends and events and to get help in making decisions and developing programs based on the best available information. The articles are generally written by their peers in public and private institutions, academics, government, or policy organizations and think tanks.

Authors looking to publish articles likely take several factors into consideration when deciding on appropriate journals for their work: the appropriateness of topic to the subject scope of the journal, the prestige and reputation of the journal, and the readership of the journal. Subject-specific journals are unlikely to have the same level of prestige and reputation as journals with broader coverage such as the *New England Journal of Medicine* or, less broadly, the *American Journal of Public Health.* However, a carefully selected subject-specific journal may be a better venue for an article, and the article may be more readily accepted. The author may also consider how the journal evaluates articles—for example, whether a peer-review or refereeing process is used. Similarly, those developing personal or institutional libraries of public health journals will seek journals with an appropriate subject scope, intended audience, and reputation to meet their needs.

Core Public Health Journals

Several techniques can be used to determine the prestige and reputation of a journal. One commonly used approach is to look up information about journals in *Journal Citation Reports (JCR)*, published annually by Thomson ISI. This publication includes various pieces of information about journal titles, which can be used to infer their prestige, such as the number of times articles in the journal are cited by other articles, especially in the first year of publication, and most famously, the "journal impact factor," which is an average number of times articles published in the previous 2 years were cited in the *JCR* coverage year.

The main category *JCR* (Science Edition) uses for public health is "Public, Environmental & Occupational Health." The top 10 journals, ranked by impact factor in this collective public health category are *American Journal of Epidemiology, Cancer Epidemiology and Biomarkers, WHO Technical Report Series, Annual Review of Public Health, Environmental Health Perspectives, Epidemiology, International Journal of Epidemiology, American Journal of Public Health, American Journal of Preventive Medicine*, and *Tobacco Control.* Other categories listed in *JCR* (Science Edition) of interest to public health researchers and practitioners may include allergy, behavioral sciences. geriatrics and gerontology, health care sciences and services, infectious diseases, nutrition and dietetics, psychology, statistics and probability, substance abuse, tropical medicine, water resources, and others. There is also a social sciences edition of *JCR.*

In addition, a list of Core Public Health Journals has been developed by the Public Health/Health Administration Section of the Medical Library Association. This list is divided into 14 subcategories: key journals for all public health, biostatistics, environmental health sciences, epidemiology, health education/behavioral sciences, health services administration, biomedical and laboratory practice, international public health, maternal and child health, occupational safety and health, public health practice, public health dentistry, public health nutrition, and public health nursing. This list is designed for librarians and public health professionals developing journal collections, as well as for potential authors looking for top-tier journals in which to publish. The developers of the list took into consideration factors related to prestige, reputation, impact factor, and scope, as well as price and access.

Indexing

As varied as the journals of public health are, similarly there are several bibliographic databases that index the content of these journals. Most important of these databases is *MEDLINE*—or the free version *PubMed*—produced by the National Library of Medicine. *PubMed* contains more than 16 million citations to journal articles in the biomedical literature back to the 1950s. If one works for an employer with online electronic journal subscriptions, better access will be achieved via an institution-specific designated URL. Other bibliographic databases that are commonly searched for public health journal articles include *Cumulative Index to Nursing and Allied Health Literature (CINAHL)*, *Agricola, ERIC, POPLINE, Ageline, PsycINFO, Public Affairs Information Service (PAIS), Embase, Scopus, EconLit*, and *BIOSIS.*

—Marie T. Ascher

See also Publication Bias

Further Readings

JCR: Journal Citation Reports, Science Edition 2004. (2005). Philadelphia: Thomson Corporation Institute for Scientific Information.

Public Health/Health Administration Section of the Medical Library Association. (2006). *Core public health journal project: version 2.0* [Electronic version]. Retrieved January 8, 2007, from http://info.med.yale.edu/eph/phlibrary/phjournals/v2/index.html.

Web Sites

PubMed: http://www.pubmed.gov.

K

Kaplan-Meier Method

The Kaplan-Meier (or product limit) estimator $\hat{S}(t)$ is a nonparametric (or distribution free) estimator of a survival distribution $S(t)$. It was derived by Kaplan and Meier in 1958 as a direct generalization of the sample survivor function in presence of censored data.

In clinical applications, the Kaplan-Meier method is very often used to estimate the probability of dying from specific causes or the probability of occurrence or recurrence of a disease. In general, the Kaplan-Meier method can be used to estimate the probability of occurrence of any event.

The Kaplan-Meier method is generally used to summarize the survival experience of groups of individuals in terms of the empirical survivor function.

Typically, not all individuals under study fail during the observation period. Some individuals may leave the study early while still alive, and some other individuals may finish the study alive. These individuals are called censored.

The Kaplan-Meier estimator does not require any assumptions about the functional form of the distribution of failures and accounts for censored observations. For small data set, the Kaplan-Meier curve can be easily calculated by hand. Most statistical software contains routines for the calculation of the Kaplan-Meier estimator.

Consider a sample of N individuals who are followed up in time prospectively. During the observation period, suppose that K of these individuals die. We also assume that $N - K$ individuals are censored.

Let $t_1 \leq t_2 \leq \cdots \leq t_K$ be the ordered failure times for the K individuals who die during the observation period.

To construct the Kaplan-Meier estimator of the survival distribution, we start by dividing the observation period into small intervals $I_1 = [t_0, t_1)$, $I_2 = [t_1, t_2)$, $\ldots$, $I_j = [t_{j-1}, t_j), \ldots, I_K = [t_{K-1}, t_K)$, each one corresponding to the survival time of the noncensored individuals. For each interval $I_j (j = 1, \ldots, K)$,

$d_j =$ the number of individuals who die in the interval I_j;

$c_j =$ the number of individuals censored in the interval I_j;

$r_j =$ the number of individuals who are alive and at risk at the beginning of the interval; and

$h_j =$ the hazard of failure, or the conditional probability of an individual surviving through I_j, given that he was alive at the beginning of I_j; this quantity can be well approximated by $\hat{h}_j = d_j / r_j$, the ratio of number of failures over the number of individuals at risk during the interval I_j.

The observed proportion of failures d_{1j}/r_{1j} represent an estimate of the hazard of failure (or instantaneous failure rate) $h(t)$.

At the beginning of the observation period, t_0, all individuals are alive, so that, $d_0 = 0$ and $r_0 = N$. At each step, we calculate $r_j = r_{j-1} - d_{j-1} - c_{j-1}$ to update the number of individuals at risk.

The Kaplan-Meier estimate of the survival distribution $S(t)$ is obtained by the product of all the $1 - \hat{h}_j$:

$$\hat{S}(t) = \prod_{t_j < t} (1 - \hat{h}_j).$$

The Kaplan-Meier estimate $\hat{S}(t)$ is a left continuous, not increasing, step function that is discontinuous at the observed failure times t_j. The intervals I_j may vary in length and depend on the observed data. Observations that are censored at t_j are assumed to occur after t_j. Censored observations contribute to the risk set till the time they are last seen alive. If a failure and a censoring time are tied (i.e., occur at the same point in time), we assume that the failure occurs just before the censoring.

An estimate of the variance of the Kaplan-Meier curve is given by the Greenwood's formula:

$$\hat{\text{Var}}[\ln \hat{S}(t)] = \sum_{j:t_j < t} \frac{d_j}{r_j(r_j - d_j)}$$

(here ln indicates the natural logarithm) from which it is derived that

$$\hat{\text{Var}}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{r_j(r_j - d_j)}.$$

To calculate an approximate $100(1-\alpha)\%$ confidence interval for $\hat{S}(t)$, we first calculate a confidence interval for $\ln \hat{S}(t)$ and then we exponentiate to obtain the upper and lower bound of the confidence interval for $\hat{S}(t)$. Note that at extreme values of t such a confidence interval may include unreasonable limits outside the range [0,1].

Example

Consider the following survival times corresponding to the time to death, in days, of 15 patients with advanced squamous cell lung cancer:

$$72, 411, 228, 126, 118, 10, 82, 110,$$
$$314, 100^c, 42, 8, 144, 25^c, 11.$$

The superscript c indicates that the observation is censored. Thus, two patients left the study alive at 25 and 100 days, respectively. All other patients died during the observation period. To construct a Kaplan-Meier estimate of the probability of dying from lung cancer, we follow the steps described above.

Table 1 Lung Cancer Data

t_j	r_j	d_j	c_j	$(1-\hat{h}_j)$	$\hat{S}(t)$
0	15	0	0	1	1
8	15	1	0	0.9333	0.933
10	14	1	0	0.9286	0.866
11	13	1	0	0.9230	0.800
25 *	12	0	1	1	0.800
42	11	1	0	0.9091	0.7273
72	10	1	0	0.9000	0.6545
82	9	1	0	0.8888	0.5818
100 *	8	0	1	1	0.5818
110	7	1	0	0.8571	0.4987
118	6	1	0	0.8333	0.4156
126	5	1	0	0.8000	0.3325
144	4	1	0	0.7500	0.2494
228	3	1	0	0.6666	0.1662
314	2	1	0	0.5000	0.083
411	1	1	0	0.0000	0.00

Source: Adapted from Kalbfleish and Prentice (2002, Appendix A, p. 378).

Note: Asterisk (*) represents censored observations.

Table 1 summarizes the steps necessary to calculate the Kaplan-Meier estimate of the time to death for these individuals.

The median survival time is calculated from a Kaplan-Meier curve as the time at which the probability of dying is 50%. In the lung cancer data example, the median survival time is 100 days.

To obtain a 95% confidence interval for $\hat{S}(72)$, we first compute a confidence interval for $\ln \hat{S}(72)$

$$\ln S(72) \pm 1.96\sqrt{\sum_j \frac{d_j}{r_j(r_j - d_j)}}$$
$$= [-0.4761 - 0.0299],$$

and then we exponentiate the lower and the upper bound of this interval to obtain a 95% confidence interval for $\hat{S}(72)$:[0.6212 – 1.00].

—Emilia Bagiella

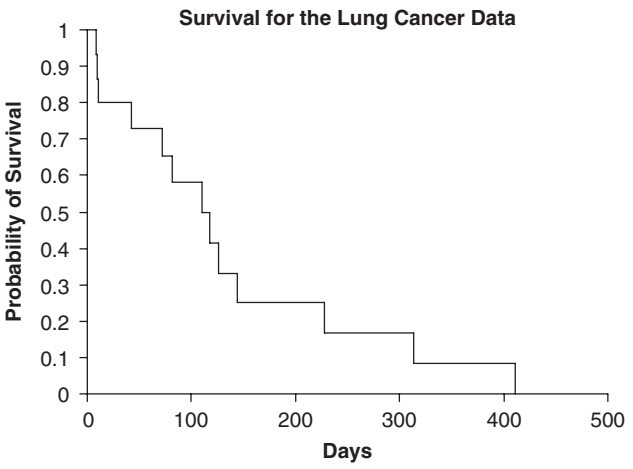


Figure 1 Kaplan-Meier Method Estimates

Source: Adapted from Kalbfleish and Prentice (2002).

See also Censored Data; Hazard Rate; Survival Analysis

Further Readings

Kalbfleish, D. J., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). Hoboken, NJ: Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association, 53*, 457–481.

Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika, 60*, 279–288.

Kappa

The kappa statistic is a measure of agreement, corrected for chance, for a categorical variable. For example, if two radiologists each assess the results for the same set of patients, the kappa is one way to measure how well their conclusions agree. The kappa may be used if the rating system used to grade each patient is binary or categorical. With either a large number of ordinal categories (such as a scale from 0 to 20) or a continuous rating scale, Pearson's correlation coefficient would provide a better assessment of agreement than the kappa.

The formula for the kappa is $\kappa = (p_o - p_e)/(1 - p_e)$, where p_o is the proportion of observed agreement (the sum of the observed values of the cells on the diagonal over the total number of observations), and p_e is the proportion of agreement expected by chance (the sum of the expected values of the same cells on the diagonal over the total number of observations). Notice that the denominator shows the difference between perfect agreement and the amount of agreement expected by chance, representing the best possible improvement of the raters over chance alone. This is contrasted with the numerator, the difference between the observed proportion of agreement and that expected by chance. As a result, the kappa statistic may be interpreted as the proportion of agreement beyond that which is expected just by chance, and kappa values range from less than 0 (less agreement than expected by chance) to 1 (perfect agreement). Kappa values between 0 and 0.4 represent marginal reproducibility or agreement, values between 0.4 and 0.75 show good agreement, and values more than 0.75 indicate excellent agreement.

As the number of possible categories for each rating increases, the associated kappa values tend to decrease. Fortunately, if the categories are ordinal (such as a score from 1 to 10), this can be combated by use of the weighted kappa. In the weighted kappa, the most weight is given to observations with identical ratings, then less weight is given to ratings one unit apart, still less weight is given to observations two units apart, and so on. The user defines how much weight is allotted to each possibility (identical ratings, one unit apart, two units apart, and so on) as well as defining how far apart ratings can be and still contribute toward the weighted agreement.

In the ordinary kappa statistic, only perfect agreement between Raters A and B would count toward

Table 1 Comparison of the Unweighted and Weighted Kappa Statistics

		Rater A			
		I	II	III	IV
Rater B	I	5	(5)	2	2
	II	3	4	4	0
	III	2	2	6	3
	IV	1	0	2	5

agreement. In the weighted kappa statistic, the most weight is given to perfect agreement (dark gray) with less weight given to cells with near perfect agreement (light gray). Specifically, the five observations that Rater A coded as "II" and Rater B coded as "I" would count as disagreement for the ordinary kappa statistic but count as partial agreement in the weighted kappa statistic.

The usual versions of the kappa and weighted kappa statistics allow only two raters to assess each observation. The multirater kappa is an alternative when more than two raters assess each observation. The weighted kappa and the multirater kappa may be interpreted in the same way as the generic kappa statistic.

—Felicity Boyd Enders

See also Interrater Reliability; Pearson Correlation Coefficient

Further Readings

Banerjee, M. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics, 27*(1), 3–23.

Keys, Ancel (1904–2004)

Ancel Keys was an American scientist who did pioneering work on the relationship between diet and health, particularly on the relationship between dietary fat and heart disease. Keys and his wife Margaret were the first American promoters of the Mediterranean diet, coauthoring *Eat Well and Stay Well*, an immensely popular cookbook. The discovery of a link between diet and heart disease landed Keys on the cover of *Time* magazine in 1961 and garnered him the nickname "Mr. Cholesterol."

Keys was born in Colorado Springs in 1904, an only child. His family moved to San Francisco just before the devastating April 1906 earthquake, then across San Francisco Bay to Berkeley. While in elementary school, he was identified as one of the 1,528 "gifted" children studied by Stanford psychologist Lewis Terman.

Keys attended the University of California, Berkeley, where he received a B.A. in economics and political science, an M.S. in biology, and a Ph.D. in oceanography and biology. He earned a second Ph.D. in physiology at Cambridge. In 1936, Keys became a professor of physiology at the University of Minnesota. In 1939, Keys founded the Laboratory of Physiological Hygiene. A new quantitative human biology, physiological hygiene combined physiology, nutrition, epidemiology, and prevention research.

In 1944, the U.S. government commissioned Keys to study human performance during nutritional deficiency states and to design lightweight but nutritionally adequate rations for paratroopers. Keys proposed an ambitious project to study the physiology of starvation, selecting 36 conscientious objectors, all volunteers. Keys and his colleagues brought the subjects to a baseline weight, gradually cut their daily diets from 3,500 calories to a semistarvation diet of 1,600 calories, and followed up with a rehabilitation diet. Keys then recorded the physiological changes associated with progressive food deprivation. Out of this research, Keys developed the emergency K-ration, used extensively by the U.S. military.

Immediately following World War II, Keys was perplexed by a set of seemingly counterintuitive observations: U.S. businessmen, presumably among the best-fed persons in the world, had high rates of heart disease, while in postwar Europe, cardiovascular disease rates had decreased sharply in the wake of reduced food supplies. In 1947, Keys helped establish cardiovascular epidemiology by launching the Twin Cities Study, a study of Minnesota businessmen, a few months before the better-known Framingham Heart Study began. Keys identified the relationship between dietary fat, blood cholesterol, and heart disease.

More ambitious was the Seven Countries Study, launched in 1958, which followed a sample of men in 16 distinct populations from seven nations throughout North America, Europe, and Asia. An extensive effort to characterize diet in detail via the collection and biochemical analysis of food samples distinguished the Seven Countries Study from the Framingham Study. Keys and his colleagues established that the risk of chronic disease differed greatly between populations and individuals and that these differences correlated with culturally determined lifestyle and dietary habits.

Keys retired in 1972 and maintained an active lifestyle until his death at age 100.

—Todd M. Olszewski

See also Cardiovascular Disease; Cholesterol; Chronic Disease Epidemiology; Framingham Heart Study; Nutritional Epidemiology

Further Readings

Keys, A., Aravanis, C., Blackburn, H., Buzina, R., Djordjević, B. S., Dontas, A. S., et al. (1980). *Seven countries: A multivariate analysis of death and coronary heart disease.* Cambridge, MA: Harvard University Press.

Keys, A., Brozek, J., Henschel, K., Mickelsen, O., Taylor, H. L., Simonson, E., et al. (1950). *The biology of human starvation.* Minneapolis: University of Minnesota Press.

Koch, Robert
(1843–1910)

Robert Koch is considered one of the founders of modern bacteriology and a key contributor to the etiology of diseases, along with Louis Pasteur. He isolated several disease-causing bacteria, including those for anthrax (1877), tuberculosis (1882), and cholera (1883), and developed Koch's postulates criteria for ascertaining the microbial causes of a specific disease.

Robert Koch was born in Clausthal, Germany, in 1843, one of 13 children. He received a medical degree from the University of Göttingen in 1866. Following this, Koch served as a physician in several German towns, was a field surgeon during the 1870 to 1872 Franco-Prussian war, and then became a medical officer in Wollstein, Germany. It was during this latter part of his career that Koch did most of his research, in a laboratory he developed in Wollstein.

Koch's first major scientific breakthrough occurred when he isolated anthrax bacillus and proved that it caused disease. He did this by injecting healthy mice with spores of *Bacillus anthracis* that had been obtained from the spleens of animals infected with anthrax. Mice injected with these spores later developed anthrax, while mice injected with spores from healthy animals did not. This was the first time that a specific microorganism was causally related to a specific disease.

Following this discovery, Koch developed a set of criteria to prove that a disease is caused by a specific microorganism. These four criteria are commonly referred to as Koch's postulates. Koch argues that for a microorganism to cause a specific disease, each of the four of Koch's postulates had to be fulfilled. While these criteria are not literally believed today, their development contributed significantly to the establishment of the germ theory of disease.

In 1882, Koch isolated the tuberculosis bacillus and then inoculated uninfected animals with it. This induced tuberculosis in the animals and thus established the etiologic role of the bacterium in the causation of disease. He later did further work on tuberculosis by investigating the possible protective effect of injecting a person with dead tuberculin bacilli and then subsequently injecting them with live tuberculosis bacilli and suggesting that he may have discovered a cure for the disease. Although it was not successful as a cure, findings from this work have been important in the development of the tuberculin test currently used today to detect tuberculosis infection in individuals.

Finally, Koch traveled to Egypt and India where he identified the cholera bacillus and determined that its mode of transmission was waterborne. Following this discovery, he did some work investigating vector-borne diseases such as malaria.

Koch not only did work isolating bacteria but also developed many microbiology techniques. These included methods of staining bacteria and investigation using the microscope. Furthermore, Koch introduced the solid culture medium for the cultivation of bacteria. In this way, Koch was an important contributor to the methodology of bacteriology.

Koch received the Nobel Prize in Physiology or Medicine in 1905 for his discoveries in tuberculosis. Koch was married twice during his lifetime and had one child, a daughter. He died in 1910 in Baden-Baden, Germany.

—Kate Bassil

See also Koch's Postulates; Pasteur, Louis

Further Readings

Brock, T. D. (1998). *Robert Koch: A life in medicine and bacteriology.* Washington, DC: American Society for Microbiology Press.

Web Sites

Nobel Foundation: http://nobelprize.org/nobel_prizes/medicine/laureates/1905/koch-bio.html.

Koch's Postulates

Koch's postulates, also known as Henle-Koch postulates, were published by Robert Koch in various forms between 1878 and 1884 to set forth a method of demonstrating that a bacillus causes a particular disease. These postulates follow the process that Koch went through in demonstrating that anthrax and tuberculosis bacilli cause disease. Koch's postulates state that, to establish that an organism causes disease,

- the organism must be present in all cases of the disease;
- the organism must be grown in pure culture outside a diseased animal;
- when inoculated with the organism, healthy test animals must develop the same symptoms as were present in the original cases; and
- the organism must be present in the experimentally infected animals.

Koch believed that satisfying these postulates provided definitive proof that the organism was a necessary and sufficient cause of disease. If fulfilled, these postulates provide powerful evidence that an agent causes disease; however, all these conditions need not be fulfilled to establish causation. Koch noted in his investigations that healthy animals sometimes would not develop disease after being inoculated with the pathogen, leaving the third postulate unfulfilled. Such asymptomatic infections occur in many diseases with well-established causes, such as cholera and influenza. The second postulate, that the organism must be grown in a pure culture outside the diseased animal, cannot be fulfilled for viruses since they are intracellular parasites.

Even when not fulfilled in the strictest sense, Koch's postulates still provide guidelines for establishing disease causation. In modern practice, the first and third postulates are more accurately stated as follows:

- The agent must be significantly more common in individuals with the disease than those without.
- Individuals exposed to the agent must be significantly more likely to develop the disease than those who are not.

There is some debate about the extent to which Koch was influenced by the work of Jacob Henle. Those who feel that Koch was heavily influenced by Henle's work on disease causation refer to the postulates as the Henle-Koch postulates or even as the Henle postulates. While the extent to which Koch was influenced by previous investigators is open to some debate, it is clear that the postulates were significant in Koch's groundbreaking work showing the role of the anthrax and tuberculosis bacilli in the causation of disease. The postulates therefore generally bear only his name.

Modern epidemiological research often focuses on diseases, such as diabetes and cancer, that are not necessarily caused by microorganisms. For these diseases other methods for establishing causation, such as Hill's Considerations for Causal Inference, are used instead of Koch's postulates. An expansion of Koch's postulates was proposed by Alfred S. Evans, who attempted to unify criteria of causation used in the investigation of chronic and acute diseases.

—Justin Lessler

See also Causation and Causal Inference; Etiology of Disease; Hill's Considerations for Causal Inference; Koch, Robert

Further Readings

Carter, K. C. (1985). Koch's postulates in relation to the work of Jacob Henle and Edwin Klebs. *Medical History, 29*, 353–374.

Evans, A. S. (1976). Causation and disease: The Henle-Koch postulates revisited. *Yale Journal of Biology and Medicine, 49*, 175–195.

Kurtosis

Kurtosis is a measure of the thickness of the tails of a statistical distribution and the sharpness of its peak. Kurtosis is also called the fourth moment about the mean and is one of the two most common statistics used to describe the shape of a distribution (the other is skewness).

There are three types of kurtosis: mesokurtosis, platykurtosis, and leptokurtosis. Many times, these are referred to as zero, negative, and positive kurtosis, respectively. The positive and negative descriptors refer to whether the peak of the distribution is "sharper" or higher than a normal distribution or if the peak is "flatter" or lower than the normal distribution.

Kurtosis statistics compare the distribution under study with the normal distribution. A distribution that

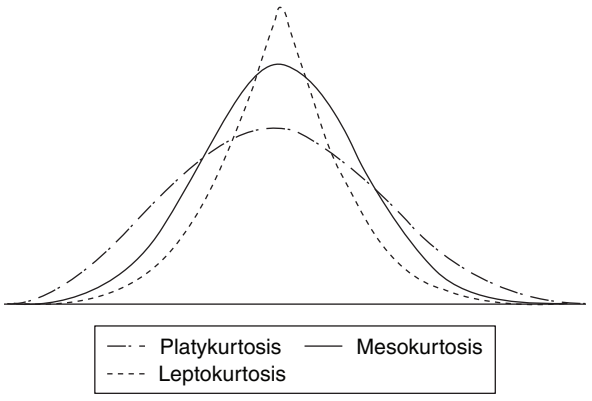


Figure 1 Common Distributions With Type of Kurtosis

Table 1 Illustration of the Three Types of Kurtosis

Distribution	*Type of Kurtosis*	
Normal	Zero	Mesokurtosis
Student's *t*	Negative	Platykurtosis (for sample sizes greater than seven)
Uniform	Negative	Platykurtosis
Exponential	Positive	Leptokurtosis
Laplace	Positive	Leptokurtosis
Weibull	Depends on the parameters of the distribution	

resembles the normal distribution in terms of relative peakedness of the distribution is said to be mesokurtic or have zero kurtosis. If a distribution has a higher peak where the width of the peak is thinner and the tails are thinner than the normal distribution, then the distribution is said to be leptokurtic or have positive kurtosis. If a distribution has a lower peak where the width of the peak is wider and the tails are thicker than the normal distribution, then the distribution is said to be platykurtic or have negative kurtosis. A platykurtic distribution may even have a concave peak instead of a rounded peak (see Figure 1).

When analyzing data using an analysis package, a researcher needs to know whether the program uses the kurtosis statistic or the kurtosis *excess* statistic. For the kurtosis statistic, a value of 3 indicates a normal distribution. However, kurtosis excess is a measure of how far the kurtosis statistic is from 3. So a normal distribution has a kurtosis excess of 0. Using the kurtosis excess statistic, the sign of the statistic matches the description of the kurtosis (-1 is negative kurtosis, $+1$ is positive kurtosis).

The most commonly known distributions and their type of kurtosis are given in Table 1.

Many times kurtosis is used to help assess whether a distribution being studied meets the normality assumptions of most common parametric statistical tests. While the normal distribution has a kurtosis of $+3$, it is important to realize that, in practice, the kurtosis statistic for a sample from the population will not be exactly equal to $+3$. How far off can the statistic be and not violate the normality assumption? Provided the statistic is not grossly different from $+3$, then that decision is up to the researcher and his or her opinion of an acceptable difference. For most typically sized (small) samples, the kurtosis statistic is unreliable. To accurately measure the kurtosis of a distribution, sample sizes of several hundred may be needed.

—Stacie Ezelle Taylor

See also Inferential and Descriptive Statistics; Skewness

Further Readings

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness & kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*(1), 183–189 (doi:10.1111/1467-9884.00122).

Reineke, D. M., Baggett, J., & Elfessi, A. (2003). A note on the effect of skewness, kurtosis, and shifting on one-sample *t* and sign tests. *Journal of Statistics Education, 11*(3). Retrieved June 5, 2007, from http://www.amstat.org/publications/jse/v11n3/reineke.html.

L

Latency and Incubation Periods

The starting point of a communicable disease is the exposure of the host to the infectious agent. When communicable diseases are considered as a whole, two different processes have to be distinguished in their evolution over time: infectiveness and disease. Infectiveness consists of two successive periods: the latency period and the period of communicability. The disease process encompasses the incubation period and the clinical signs and symptoms. Whereas the knowledge of infectiveness is of paramount importance for microbiological, pharmacological, and public health purposes, the main interest in the disease process hinges on the clinical care of the patient.

The incubation period is the time that elapses between the initial exposure of the host to the infectious agent and the first appearance of the clinical manifestations (signs or symptoms) associated with the disease. In a vector, it is the time between entrance of an infectious agent into the vector and the time when that vector can transmit the infection (a period known as the extrinsic incubation period).

The duration of the incubation period is determined by the time needed by the infectious agent to grow enough in number in the host to produce symptoms. During the incubation period, the infectious agent can be transferred from one host to another. In many diseases, the communicable period begins before the inception of signs and symptoms, as, for example, in viral hepatitis and some exanthematic infections (those characterized by skin eruptions), such as measles, rubella, scarlet fever, and chickenpox.

The incubation period is known for most diseases. It varies among diseases, and for each particular disease, although in a much lesser degree, depending on the infective dose of the agent, namely, the mean number of microorganisms needed to cause infection. (The infectiveness of an infectious agent is usually expressed as the infectious dose 50 [ID_{50}], and it is defined as the infectious dose needed to produce infection in 50% of the susceptible hosts.) Some examples of incubation periods are as follows: influenza (from 1 to 3 days), diarrhoea caused by *Escherichia coli* (from 3 to 8 days, with a median of 3 to 4 days), measles (from 7 to 18 days, with a median of 10 days), hepatitis A (from 15 to 50 days, depending on the infective dose, with an average of 28 to 30 days), and leprosy (from 9 months to 20 years, with a probable average of 4 years for tuberculoid leprosy and 8 years for lepromatous leprosy).

—Carlos Campillo

See also Host; Vector-Borne Disease

Further Readings

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans*: *Dynamics and control*. Oxford, UK: Oxford University Press.

Benenson, A. S. (Ed.). (1995). *Control of communicable diseases manual* (16th ed.). Washington, DC: American Public Health Association.

Thomas, J. C., & Weber, D. J. (2001). *Epidemiologic methods for the study of infectious diseases*. New York: Oxford University Press.

Latent Class Models

Many quantities of interest in sociology, psychology, public health, and medicine are unobservable but well-conceived characteristics such as attitudes, temperament, psychological diagnoses, and health behaviors. Such constructs can be measured indirectly by using multiple items as indicators of the unobservable characteristics. The latent class (LC) model can classify individuals into population subgroups based on these unobservable characteristics, by using their responses to questionnaire items that are related to those characteristics. Suppose, for example, we are interested in the construct of nicotine dependence and want to classify respondents into groups corresponding to different types of nicotine dependence we believe are indicated by a number of behaviors relating to cigarette use. We can collect data on several items pertaining to cigarette use and then apply an LC model to this data to identify two or more nicotine dependence types to which smokers might belong. We can use this LC model to classify not only the subjects in our study but also subjects in other studies. The LC model has been widely applied in behavioral and biomedical applications, particularly in the area of substance use prevention and treatment. In physical or psychiatric research, the LC model has been a popular strategy for development and evaluation of diagnostic criteria. The LC model can also be a useful way to address many problems of categorical data analysis in population-based epidemiologic studies.

The Mathematical Model

The LC model explains the relationship among manifest items by positing the assumption that the population comprises different classes related to the construct of interest. In other words, the population is assumed to consist of mutually exclusive and exhaustive groups, called latent classes, and the distribution of the items varies across classes. The LC model comprises two types of parameters: the relative size of each class and the probability of a particular response to each item within a class. To specify an LC model, let $\mathbf{Y} = (Y_1, \ldots, Y_M))$ be M discrete items measuring latent classes, where variable Y_m takes possible values from 1 to r_m. Let $C = 1, 2, \ldots, L$ be the variable of LC membership, and let $I(y = k)$ denote the indicator function that takes the value 1 if $y = k$ and 0 otherwise. If the class membership were observed, the joint probability that an individual belongs to Class 1 and provides responses $\mathbf{y} = (y_1, \ldots, y_M)$ would be

$$P(Y = y, C = l) = \gamma_l \prod_{m=1}^{M} \prod_{k=1}^{r_m} \rho_{mk|l}^{I(y_m = k)},$$

where

$\gamma_l = P(C = 1)$ represents the probability of belonging to Latent Class 1 and

$\rho_{mk|1} = P(Y_m = k|C = 1)$ represents the probability of response k to the mth item given a class membership in 1.

Therefore, the marginal probability of a particular response pattern $\mathbf{y} = (y_1, \ldots, y_M)$ without regard for the unseen class membership is

$$P(Y = y) = \sum_{l=1}^{L} \gamma_l \prod_{m=1}^{M} \prod_{k=1}^{r_m} \rho_{mk|l}^{I(y_m = k)}.$$

Here, we have assumed *local independence*—that is, the items are assumed to be unrelated within each class. This assumption is the crucial feature of the LC model that allows us to draw inferences about the unseen class variable.

Model Identifiability

Model identification is imperative in estimating parameters of an LC model. The parameters of an LC model are said to be *locally identifiable* if the likelihood function is uniquely determined by the parameters within some neighborhood of a particular value of parameters. A necessary (but not sufficient) condition to make the LC model locally identifiable is that the number of possible response patterns of manifest items must be greater than the number of free parameters in the LC model. Even if this necessary condition is satisfied, it is impossible to say a priori whether or not this model is indeed identifiable. A necessary and sufficient condition for local identifiability is that the first derivative matrix of the log-likelihood function with respect to the parameters evaluated at a particular value must have full column rank. When an LC model is not identifiable, the simplest way to achieve identification is to reduce the number of parameters to be estimated by fixing or constraining parameters.

Estimation

The most common approach to estimate unknown parameters in an LC model is maximum likelihood (ML) estimation using the EM (E-step [expectation] and M-step [maximization]) algorithm. EM is an iterative procedure in which each iteration consists of two steps, the E-step and M-step. Iterating these two steps produces a sequence of parameter estimates that converges reliably to a local or global maximum of the likelihood function. However, researchers should apply several sets of starting values to ensure that the ML estimates represent the best solution (i.e., global maximum). On convergence, standard errors for the estimated parameters are obtained by inverting the Hessian matrix (i.e., the negative second derivative matrix of the log-likelihood function).

As an alternative to ML, one can apply the Bayesian method via a Markov chain Monte Carlo (MCMC) method to estimate unknown parameters in LC models. The most popular MCMC method for the LC model is closely related to EM; it is an iterative procedure whereby each iteration consists of two steps, the I-step (imputation) and the P-step (posterior). Repeating this two-step procedure creates a sequence of iterates converging to the stationary posterior distribution of the LC model parameters. MCMC may produce greater flexibility in model fit assessment and various hypothesis tests without appealing to large-sample approximations.

Model Selection

It is important that a model be assessed adequately at the outset of an analysis. Numerous statistical methods are being adopted to evaluate model fit. However, different methods often suggest different solutions, yielding ambiguity in model selection. The choice of the number of latent classes can be driven, to summarize the distinctive features of the data in as parsimonious a fashion as possible, by a balanced judgment that takes into account the substantive knowledge and objective measures available for assessing model fit.

The log-likelihood-ratio statistics, including the likelihood-ratio test (LRT), are used to assess the absolute model fit of an LC model by comparing the predicted response pattern frequencies with the observed frequencies appearing in the data. For large samples with a relatively small number of response patterns, LRT may be a valid statistic to assess the absolute model fit with the asymptotic chi-square approximation. However, the difference in LRT for testing the relative fit of an L-class model against an $(L+1)$-class alternative does not have limiting chi-square distribution. Therefore, LRT cannot be used to assess the relative model fit of two competing models with different numbers of latent classes. If we do not give up the concept of overall LRT, several solutions to the problems in the asymptotic assumption are available. For example, the adjusted LRT and the parametric bootstrap are popular methods because they are easy to access via software packages such as Mplus and Latent GOLD.

There are numerous other model selection criteria widely used to compare models with different numbers of classes. For example, goodness-of-fit measures based on information theory (e.g., Akaike information criterion, consistent Akaike information criterion, and Bayesian information criterion) can be used to compare LC models with different number of classes. One drawback to this approach is that these methods assess relative model fit only; this says nothing as to whether the best model in a set of competing models actually fits well.

Within a Bayesian framework, there is a tool using the posterior probability check distributions (PPCD) that can be used to assess model fit. The PPCD for the LRT that does not rely on any known distribution can be constructed in the following way: (a) simulate the random draw of parameters from their Bayesian posterior distribution under the current LC model, (b) draw a hypothetical new data set from the same model using the simulated parameters, (c) fit the current model and the competing model to the simulated data set, and (d) compute the LRT difference based on output from (c). Repeating (a) to (d) many times produces a sample of LRT from the PPCD. The area to the right of the observed LRT can be regarded as a Bayesian p value.

—Hwan Chung

See also Bayesian Approach to Statistics; Likelihood Ratio; Regression

Further Readings

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*, 215–231.

Lanza, S. T., Collins, L. M., Schafer, J. L., & Flaherty, B. P. (2005). Using data augmentation to obtain standard errors and conduct hypothesis tests in latent class and latent transition analysis. *Psychological Methods*, *10*, 84–100.

Lanza, S. T., Flaherty, B. P., & Collins, L. M. (2003). Latent class and latent transition analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology* (pp. 663–685). Hoboken, NJ: Wiley.

Rubin, D. B., & Stern, H. S. (1994). Testing in latent class models using a posterior predictive check distribution. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 420–438). Oxford, UK: Oxford University Press.

Latino Health Issues

As a diverse group, comprising people who differ in national origin, generation, level of acculturation, socioeconomic class, region of residence, and gender, it is not surprising that there is considerable variation in the overall health of the Latino population of the United States. (*Latino* is an umbrella term that covers people of diverse points of origin in Latin American countries.) This diversity among Latinos notwithstanding, reviews of the health status of Latinos relative to the dominant non-Latino white population affirm the importance of health disparities as a prevailing feature of the national health pattern. The average life span of Latinos in the country is 20 years shorter than for non-Latino whites. In terms of specific health differences, Latinos are disproportionately likely to die of violence (homicide being the second leading cause of death for Latinos 15 to 24 years of age), develop late-onset diabetes (11% compared with 5% in the general population), develop cervical cancer (with an incidence rate of 17 per 100,000 for Latinas compared with just under 9 per 100,000 for non-Latinas), suffer from asthma (hospitalization rates for Latino youth in some states are five times higher than for whites), and become infected with HIV/AIDS (although Latinos account for approximately 14% of the population of the United States and Puerto Rico, they comprise 18% of AIDS cases diagnoses since the beginning of the epidemic). Among some Latino subgroups, especially Puerto Ricans and Mexican Americans, there are comparatively high rates of illicit drug, and Mexican Americans have comparatively high rates of heavy drinking compared with non-Latino whites. Latinas are more likely to report teen pregnancy and have the highest teen birth rate of all ethnic populations in the United States. Of equal significance as these various health disparities, Latinos are the least likely population in the United States to have health insurance and one of the most likely to encounter linguistic and cultural barriers in accessing effective and appropriate health care. For example, research has shown that monolingual Spanish-speaking Latino patients are significantly less likely than non-Hispanic whites to have had a physician visit or an influenza vaccination during the year, while rates for English-speaking Latinos are similar to those of non-Hispanic whites. Despite these challenges, in terms of a number of critical health indicators (e.g., various cancers, cardiovascular and pulmonary disease), Latinos tend to exhibit lower rates of disease and ill health than the dominant population.

Sociodemographic Factors in Latino Health

Latinos now constitute the largest ethnic minority population in the country—5 years earlier than the U.S. Census Bureau had projected they would do so. This population continues to grow at an unprecedented rate of 5.7% each year nationally. It is now estimated that there are 37 million Latinos in the United States, up almost 12% since the completion of the 2000 census. Latinos now comprise almost 13% of the total U.S. population or roughly one of every eight people in the country. As a result, the overall health of the Latino population and the specific health needs of Latinos have a significant impact on health in the country generally.

Generally, Latinos are a comparatively young population, which is a key factor in their exhibiting lower rates of morbidity and mortality relative to diseases commonly associated with aging (e.g., heart disease, cancer). The median age of Latinos in the United States is 25 years, and nearly 40% of Latinos are less than 20 years of age. Moreover, the proportion of Latino children in the total number of children in the United States has increased at a faster pace than that of any other ethnic group, growing from 9% of the child population in 1980 to 19% by 2004. If current trends continue, one in four people living in the United States by 2050 will be Latino, affirming the importance of the health of this population relative to overall health in the United States.

In assessing the health status of the Latino population, it must be stressed that public health researchers generally acknowledge that mortality rates among Latinos tend to be underestimated for several reasons, including misreporting of ethnicity on death certificates and other health-related documents used as health indicators and the return migration of Latinos with life-threatening illnesses to their country of origin. Similarly, morbidity rates for Latinos often are underestimated because of a failure to collect full ethnicity data in many studies or in surveillance monitoring, use of methodologies such as telephone interviewing that do not effectively reach the lowest income populations (because of geographic mobility, inability to pay telephone bills on time, doubling up of families in residential settings, homelessness, etc.), failure of survey research to reach segments of the Latino population, language barriers, and lack of access to health care despite significant health problems. Moreover, because of diversity within the overall Latino population, national statistics on Latinos provide only a rough image of the sociodemographic characteristics and health profiles of Latinos in any particular state. Country of origin (for individuals born outside the United States, or of ancestors for those born in the United States), in particular, differentiates health patterns among Latino subgroups. For example, the frequency of female-headed households among Puerto Ricans is double that of Cubans, the percentage of Latinos of Mexican origin who live in a nonmetropolitan area is more than double that of Puerto Ricans, the rate of asthma among Puerto Ricans is more than double that of Mexicans or Cubans, and the frequency of AIDS-related mortality among Puerto Ricans is more than three times the rate among Mexican Americans.

While Latino health in the United States broadly reflects the overall health profile of the U.S. population, the health disparities found among Latinos reflect broad social disparities relative to the dominant non-Latino white population. Most notably in this regard is the issue of poverty. Poverty has been found to be closely tied with health status, and it is generally considered a risk factor for poor health and disease, especially among children. Moreover, children who grow up in poverty are more likely to do poorly in school, become teen parents, and be unemployed as adults. The poverty rate among Latino is three times the rate of non-Latino whites (approximately 26% vs. 8%). Additionally, the median income for Latino households is about $15,000 a year below that of non-Latino white households. Among Latino subgroups, rates of poverty are highest among Puerto Ricans and lowest among Cubans; nonetheless, even among Cubans, the rate of poverty is about twice as high as among non-Latino whites.

Rates of unemployment among Latinos are almost double that of non-Latino whites. Moreover, while there is a growing Latino middle class, the majority of employed Latinos are disproportionately concentrated in low-salaried jobs with limited room for economic advancement, and, of equal importance, Latinos often hold jobs that do not provide health insurance. Several economic indicators show that there has been a drop in occupational status of Latinos and a growing gap in the respective occupational status of Latinos and non-Latino whites. Data from the 2000 census indicate that about one quarter (27%) of Latino children under the age of 18 years live in poverty, compared with only 9% of non-Latino white children. Poverty rates among Latinos, however, vary significantly by ethnic subgroup. While 16% of Cuban children live in poverty, the rate for Puerto Rican children is 44%. Although rates of poverty are especially high among immigrant Latinos, even among third- and fourth-generation children of Mexican origin, rates of poverty are 2.5 times those of non-Latino white children.

Poverty and related structural violence, such as discrimination and racism, affect health in a number of direct and indirect ways, including unhealthy diets, increased exposure to physical and emotional stressors in the environment (e.g., street violence), heightened exposure to environmental toxins (e.g., lead paint, rodent infestation) and street violence, comparatively high rates of immune system impairments, and limitations on access to preventive care and treatment for existing conditions. For example, Latinos are four times as likely as non-Latino whites to suffer from tuberculosis. Rates of tuberculosis tend to reflect living conditions and residential crowding. Because of poverty, many Latino families live in substandard, poorly ventilated housing, with a higher number of people per square foot of living space than non-Latino whites. These are the conditions under which a communicable disease such as tuberculosis is most likely to spread.

While, as noted, health disparities have been found to be strongly influenced by poverty, they are especially common among poor people who live in areas

with a high percentage of poor people in the local population. Thus, poor people in cities with smaller impoverished populations are at lower risk of dying than those in cities with large impoverished populations. Comparisons of findings from the 1990 and 2000 census counts indicate that the size of the Latino population in big cities is increasing, especially in large metropolitan areas such as New York and Los Angeles that already have large concentrations of Latinos. Thus, the number of Latinos who live in neighborhoods in which a majority of residents were Latino grew faster (76%) than in neighborhoods in which Latinos constitute a minority of residents (51%) between 1990 and 2000. In many urban areas with Latino populations, Latinos are disproportionately found in more densely populated inner city zones with comparatively high rates of poverty, a place where disease syndemics involving multiple interacting diseases of diverse kinds are most common.

Moreover, studies of differences in location among the poor show that the sociophysical environment in which people live—that is, their experience of their surrounding community, including issues of danger, stress, comfort, and appeal—is a critical determinant of their health. Feelings of hopelessness and powerlessness in a community have been found to be good predictors of health risk and health status. These are precisely the kinds of sentiments that have been recorded in a number of studies of Latino youth.

This array of factors underlies the disproportionate health burden of Latinos and directly contributes to the specific health problems, discussed below, that are especially common in this population.

Diabetes

The fifth leading cause of death in the United States, diabetes, a disease without a cure, is particularly common among Latinos. Mexican Americans over the age of 20 years, for example, are 1.7 times more likely to develop diabetes than non-Latino whites. Among residents of Puerto Rico, this rate is even somewhat higher. Comparing Latino subgroups, approximately 26% of Puerto Ricans, 24% of Mexican Americans, and 16% of Cubans between the ages of 45 and 74 have diabetes. Diabetes produces several debilitating conditions in different parts of the body, including diabetic retinopathy or damage to the small blood vessels of the retina of the eye and renal disease and kidney failure. Among Mexican Americans with diabetes, the prevalence of retinopathy is between 32% and 40% while they are 4.5 to 6.6 times more likely to develop end-state renal failure than non-Latino whites.

Asthma

More than 17 million people in the United States suffer from asthma, of whom about a third are below 18 years of age. Asthma has been identified as one of the most common chronic illnesses of childhood in the United States. Recent national surveys report an overall lifetime asthma prevalence of 12.2% for children (below 18 years of age), although there are dramatic differences across ethnic groups and subgroups. Among Latinos, Puerto Ricans have the highest lifetime asthma rate (19.6%), more than three times that for Mexican Americans (6.1%) and almost double the rate of non-Latino whites (11.1%). While alarming, reported rates of asthma among Puerto Ricans probably underestimate the prevalence of this disease in this population because many symptomatic children go undiagnosed for long periods of time. It is estimated that the actual rate of childhood asthma among Puerto Ricans may reach 30%. Research shows that children with asthma are more likely to miss school than children without asthma from the same neighborhoods. An accumulating body of research indicates that low-income Latino families that have young children with asthma lack the necessary information, training in asthma management, and medical resources needed for good asthma control. The problem, however, is not simply one of the health care system failing to adequately prepare Latino parents in asthma management. Research suggests that Latino children are given fewer beta2-agonists (a standard component of asthma management), and Latino children received fewer inhaled steroids from their physicians than non-Latino white children. Even among Latino children in private care, a significant association was found between Latino ethnicity and low inhaled steroid use.

Cancer

Cancer is the second leading cause of death in the United States, with just less than half a million new cancer diagnoses each year. Notably, about half of all new cases of cancer are diagnosed among people 65 years of age and above, a group that while growing

among Latinos is still comparatively smaller than in the general populations. Nationally, cancers that are most common among Latinos are of the cervix, esophagus, gallbladder, and stomach, with noticeable rises in female breast and lung cancer rates. Even though breast cancer incidence rates are lower among Latinas than non-Latino white women, Latinas are more likely to die from the disease. Divergent patterns between cancer incidence and mortality rates among Latinas appear to be primarily due to differences in rates of cancer screening (e.g., mammography). Latinas are less likely to obtain cancer screening than whites, which results in later-stage cancer diagnoses and less curable diagnoses, producing lower survival rates from breast cancer among Latinas. A similar pattern exists with cervical cancer. In addition to barriers to health care access, several cultural factors, including the belief that cancer is not curable, may contribute to comparatively low cancer screening rates among Latinos.

Cardiovascular and Pulmonary Health Issues

The number one cause of death in the United States, regardless of ethnicity, is cardiovascular disease, including coronary heart disease, hypertension, and stroke. These diseases, especially coronary heart disease and stroke, kill almost as many people in the United States as all other diseases combined, and they are among the leading causes of disability in the country. A national focus on lowering the rates of such diseases has contributed to a recent decline in rates of heart disease and stroke in the overall population of the country. Nationally, about one fourth of deaths among Latinos stem from cardiovascular conditions, and the rate is expected to rise in coming years as the Latino population ages. Latinas, in particular, are overrepresented among heart disease cases in the country.

Acquired Immune Deficiency Syndrome (AIDS)

Approximately 20% of people in the United States living with HIV infection are Latino, and it is expected that the overall number of new AIDS cases among Latinos will surpass that of non-Latino whites, raising critical questions about the adequacy and appropriateness of AIDS prevention and care available to Latinos. This concern is sharpened by recognition that while improved prevention and treatment methods contributed to a general decline in AIDS cases for all ages, genders, and ethnicities during the 1990s, the decline was not distributed evenly, resulting in a widening gap between Latinos and non-Latino whites. While the decline nationally in new AIDS cases among non-Latino whites between the years 1993 and 2001 was 73% compared with previous years, among Latinos it was only 56%. Similarly, the number of deaths attributed to AIDS during this period fell by 80% among non-Latino whites but only by 63% among Latinos. The estimated AIDS prevalence (i.e., accumulated cases) among non-Latino whites during this period rose by 68%, in contrast, among Latinos prevalence jumped by 130%, further affirming significant AIDS disparities between Latinos and the majority population.

—Merrill Singer

See also Cultural Sensitivity; Health Communication; Health Disparities

Further Readings

Aguirre-Molina, M., Molina, C., & Zambrana, R. (2001). *Health issues in the Latino community*. New York: Jossey-Bass.

Asencio, M. (2002). *Sex and sexuality among New York's Puerto Rican youth*. Boulder, CO: Lynne Rienner.

Diaz, R. (1998). *Latino gay men and HIV: Culture, sexuality and risk behavior*. New York: Routledge.

LaVeist, T. (Ed.). (2002). *Race, ethnicity, and health: A public health reader*. New York: Jossey-Bass.

LaVeist, T. (2005). *Minority populations and health: An introduction to health disparities in the U.S.* New York: Jossey-Bass.

Perez, D. (2006). *Understanding Latino health policy and barriers to care*. New York: ProQuest.

Smedley, B., Stith, A., & Nelson, A. (Eds.). (2002). *Unequal treatment: Confronting racial and ethnic disparities in health*. Washington, DC: National Academies Press.

LEAD

Lead and its compounds have been used in countless ways for thousands of years, despite the equally long history of knowledge about the dangers of lead. Exposure to lead-containing goods and the environmental contamination from their manufacture and

use produces considerable mortality and morbidity among workers, children, and the general public. Lead's toxicity appears to have no threshold for harm: Blood-lead levels (BLLs) $< 10\mu$g/dl (well below those associated with clinical symptoms of lead poisoning) are associated with neurological deficits in children, while slightly elevated BLLs are implicated in increased rates of hypertension and kidney disease in adults. At higher BLLs, clinical signs of lead poisoning appear, including chronic or acute gastrointestinal symptoms and neurological conditions ranging from palsies to paralysis and encephalopathy.

Sources of Exposure to Lead

There are three primary avenues of exposure to lead: (1) environmental sources that are more or less shared by all persons in a given population; (2) occupational exposures from manufacturing or handling lead and lead products; and (3) pediatric exposures, due to the special environmental, behavioral, and metabolic features of early childhood.

Until a few decades ago, people living in industrialized nations routinely faced lead levels higher than those in industrial settings where lead was used, as a result of exposure via air pollution, water supply systems, adulterated foods, medicines, and other sources. Assays of historical lead pollution show a steady increase in bioavailable lead in the general environment from early-modern times, accelerating dramatically in the middle of the 20th century with increased consumption of leaded gasoline. In addition to such airborne sources, lead made its way into foods via insecticide residues and pigments and into drinking water via solid-lead or lead-soldered water supply pipes. Lead-tainted alcohol is implicated in widespread chronic illness from the 18th century through the early 20th century, and alcohol distilled illegally in lead-soldered automotive radiators (e.g., ''moonshine'') continues to produce occasional outbreaks of saturnism.

At the beginning of the 20th century, the health burden of these universal exposures remained largely hidden beneath the crushing mortality and morbidity from lead in the workplace. From 1910 to 1930, federal mortality statistics reported more than a hundred lead-poisoning deaths annually. But these reports drastically underreported lead-poisoning cases; Leake, a New York physician, complained in 1927 that "there are many plants. . . from which no lead cases are reported, except those failing to 'get by' the coroner" (Leake, 1927). In the Progressive Era (ca. 1890–1913), the fatal conditions in America's lead factories prompted a number of government-sponsored investigations, such as those conducted by Alice Hamilton. The resulting social and political pressure, together with a constricting labor market and the adoption of workers' compensation laws, prompted manufacturers to adopt basic improvements in ventilation and processes. Workers' compensation laws also transformed factory culture, as an empowered force of industrial hygienists sought to control the physical and fiscal costs of occupational sickness. Together, these factors dramatically lowered American workers' exposure to bioavailable lead.

It took considerably longer to fully assess and respond to the ubiquitous dangers lead posed children. Young children in urban and suburban environments were exposed to most of the "universal" sources adults faced, prompting one astute pediatrician, Ruddock, in 1924 to alert his peers that "the child lives in a lead world."

Young children's normal hand-to-mouth activities increase their exposure to any lead dust in their environment (whether from painted surfaces or from air pollution), and teething and oral exploratory activities can lead to the direct ingestion of lead paint. In addition to these behaviors, children's stature and their age-specific metabolic characteristics amplify both the level of exposure and the rate of lead absorption. Despite this, childhood lead poisoning was nearly invisible prior to the 1930s. Federal mortality statistics from 1910 to 1930 typically reported fewer than five pediatric lead-poisoning deaths annually.

Several factors combined to produce this statistical silence. First was the association of lead poisoning with occupational exposure and the dominance of occupational exposure in defining where lead poisoning was to be found and what symptoms and causative factors physicians looked for. Because the typical nursery looked nothing like the typical paint factory, medical professionals assumed that lead poisoning would not occur in the nursery. Second was the dominance of acute infectious diseases as the main health hazard of the young: Several hundred children dying from the toxicants in their environment barely registered compared with the numbers killed annually by contagious diseases. The statistical picture was also confused by a third factor: Lead poisoning's wide

symptomatology easily allowed misdiagnosis by doctors who attributed the gastrointestinal and neurological symptoms seriously lead-poisoned children presented to infectious causes. Gradually, a more accurate picture of childhood lead poisoning developed, though this realization was slowed considerably by the fourth factor: As medical interest in pediatric lead poisoning rose, the disease quickly came to be perceived as one of poverty, and of race—just one more tragic problem of the ghetto—and treated with similar complacency; as one researcher, Conway, put it in 1940, "like the poor, lead poisoning is always with us."

Historical Changes in Lead Exposure

Given what we now know about the effects of lead absorption, and with due consideration for the work yet to be done, the dramatic reduction since the 1960s in the amount of bioavailable lead in the environment, and the corresponding drop in average BLLs, must be considered one of the 20th century's greatest public health achievements. Lead production and reliance on lead products have not waned; in fact, annual global production of lead rarely drops far below 3 million metric tons since it peaked in 1977, at 3.5 million. In the United States, the amount of lead consumed (i.e., in the economic sense) remains at near record highs, between 1.3 and 1.7 metric tons annually, with approximately 80% going to storage batteries. But lead consumption (in the metabolic sense) has fallen dramatically, due primarily to the elimination of lead-laced consumer goods, such as ethylized gasoline, lead-based paints, and the widespread use of lead solder in canned foods and plumbing. The impact of leaded gasoline's phase out on average BLL was practically instantaneous, while the impact of shifting from lead-based house paints (well underway in the 1940s as other pigments became economically attractive) has been incremental, with the gradual removal or encapsulation of lead-painted surfaces.

Tracing the body burden of lead in populations is impossible for years prior to the standardization and wide availability of accurate BLL analysis (now routinely done with atomic absorption spectroscopy) and sufficiently robust screening programs, beginning roughly in the 1970s. But case-finding programs in American cities in the 1950s and 1960s suggest that *average* BLLs for urban Americans were >20 μg/dl of whole blood. The distribution of risk was skewed heavily toward the urban poor: for example, a 1956 Baltimore study found that of 333 babies attending well-baby clinics, more than 40% had BLLs >50 μg/dl and 21 of these "asymptomatic" children carried a lead burden in excess of 80 μg/dl. Beginning with the Second National Health and Nutritional Examination Survey (NHANES II), conducted from 1976 to 1980, a much clearer picture of average lead burdens emerged; and together with the subsequent NHANES III and "Continuous NHANES," the series provide dramatic evidence of the progress in reducing average BLLs since the 1970s. Geometric mean BLLs in late 1970s ranged from 10 to 15.8 μg/dl, and the mean for children aged between 1 and 5 years was 15 μg/dl. By 1991, the average BLL had declined to 3.6 μg/dl; by 1999, it stood at 1.9 μg/dl. The prevalence of BLLs >10 μg/dl fell dramatically as well. Where NHANES II found that 88% of children aged 1 to 5 years had elevated BLLs, by the end of the century that number had fallen from 98% to 1.6%. Racial and ethnic disparities remain, however. NHANES data from 1999 to 2002 show that the percentage of black children aged 1 to 5 years with elevated BLLs (3.1%) was nearly twice (1.6%) that for all children aged 1 to 5 years.

It is only as average BLLs decline that the health effects of lower exposure levels can be determined. The impact of this fact is clear in the federal government's gradual lowering of the "threshold" for harm in pediatric cases. In 1970, the Surgeon General defined BLLs >70 μg/dl as "undue lead absorption"; a year later the Centers for Disease Control and Prevention (CDC) established 40 μg/dl as the level above which children should be treated for lead. In 1975, this number was reduced to 30; it was lowered to 25 μg/dl 10 years later and to 15 μg/dl in 1990. Since 1991, the CDC has defined 10 μg/dl as "elevated." The agency insists that this is not a "definitive toxicologic threshold," but a useful level for assessing and managing risks. For example, the Department of Health and Human Services "Healthy People 2010" program intends to eliminate all BLLs >10 μg/dl among children aged 1 to 5 years by 2010. The adequacy of even that goal is challenged by recent studies suggesting that lead's impact on IQ scores increases more sharply between 0 and 10 μg/dl than it does above that "threshold."

—Christian Warren

See also Environmental and Occupational Epidemiology; Hamilton, Alice; National Health and Nutrition Examination Survey; Pollution; Urban Health Issues

Further Readings

Bellinger, D. C., & Bellinger, A. M. (2006). Childhood lead poisoning: The torturous path from science to policy. *Journal of Clinical Investigation*, *116*, 853–857.

Centers for Disease Control and Prevention. (2005). *Preventing lead poisoning in young children.* Atlanta, GA: Author.

Conway, N. (1940). Lead poisoning: From unusual causes. *Industrial Medicine*, *9*, 471–477.

Leake, J. P. (1927). Lead hazards. *Journal of the American Medical Association*, *89*, 1113.

Ruddock, J. C. (1924). Lead poisoning in children with special reference to pica. *Journal of the American Medical Association*, *82*, 1682.

Warren, C. (2000). *Brush with death: A social history of lead poisoning.* Baltimore: Johns Hopkins University Press.

Life Course Approach

A life course approach to epidemiology is the study of the long-term effects on health and disease risk of biological, behavioral, social, and psychological exposures that occur during gestation, childhood, adolescence, and adulthood. The approach recognizes that exposures and disease pathways may operate independently, cumulatively, and interactively throughout an individual's life course, across generations, and on population-level disease trends. It acknowledges that exposures that occur during certain critical or sensitive developmental periods may have particular long-term health effects. Within chronic disease epidemiology, the life course approach has both challenged and expanded the prevailing adult lifestyle model of chronic disease risk. While many conceptual and methodological challenges remain, this renewed perspective within epidemiology has catalyzed a reconceptualizing of pathways of disease etiology and contributed to an increasingly multilevel and integrative understanding of the social, psychological, and biological determinants of health.

Historical Background

The idea of childhood origins of risk for adult diseases was present in epidemiology and public health thinking during the first half of the 20th century. This paralleled the development of a life course approach in disciplines, such as developmental psychology, demography, sociology, anthropology, human development, and the biological sciences. After World War II, however, the adult risk factor model of disease, with its focus on clinical and biochemical markers in adults, became the dominant model within chronic disease epidemiology. The identification of the major adult lifestyle risk factors for coronary heart disease (CHD) solidified the success of this approach.

Nonetheless, since the late 1970s and 1980s, a growing body of evidence has documented the potential effects of poor growth, undernutrition, and infectious disease in early life and on the risk of adult cardiovascular and respiratory disease. The most influential of the research from this period originated from David Barker and colleagues in the United Kingdom, who used historical cohort studies to investigate the long-term effects of in utero biological programming associated with maternal and fetal undernutrition. The subsequent proliferation in the 1990s of research on the fetal origins of CHD, which often focused on the proxy of birth weight, met with initial resistance from proponents of the primacy of well-established adult risk factors. Life course epidemiology emerged partly as a synthesis of the two approaches, demonstrating that early life factors may act cumulatively or interactively with adult exposures to influence health and disease in later life.

Conceptual Models

Several conceptual models have emerged within the life course approach to health. One of these, the critical period model, focuses on the importance of the timing of exposure. It holds that biological programming occurring during critical periods of growth and development in utero and in early infancy has irreversible effects on fetal structure, physiology, and metabolism. This model is the basis of the fetal origins of adult disease hypothesis, elaborated by Barker and colleagues and now known as the developmental origins of health and disease (DOHaD), to reflect an expanded developmental time frame.

Research in DOHaD has suggested that the highest risk of CHD, central distribution of body fat, insulin resistance, type 2 diabetes, and the metabolic syndrome may be associated with a phenotype of lower birth weight coupled with a higher body mass index in childhood or adulthood. Many of these studies have explored the potential risk associated with accelerated postnatal catch-up growth. Researchers hypothesized that predictive adaptive responses by

the fetus and infant lead to irreversible changes that may, when coupled with environments in childhood and adult life that differ from the one anticipated during early life, increase the risk of adult disease. Populations in the developing world undergoing urbanization and the nutritional transition to Western lifestyles may thus be particularly susceptible to this developmental mismatch risk pattern. Studies have also explored gene-environment interactions, including potential epigenetic modification of fetal genes by the in utero environment.

Another life course model is the accumulation of risk model, which posits that risk may amass gradually over the life course, although the effects may be greater during certain developmental periods. This concept of accumulation of disease risk suggests that as the number and/or duration of insults increase, there is increasing damage to biological systems. Accumulation of risk may occur among independent factors. If the factors are correlated, they may be related through risk clustering or via chains of risk with additive effects or trigger effects, in which only the final link of the chain produces the outcome. The life course perspective has contributed increasingly to the study of social determinants of health and health inequalities, since risk often clusters together in socially patterned ways. For instance, children of low socioeconomic position may also experience low birth weights, poor diets, passive smoke exposure, poor housing conditions, and exposure to violence. Risk accumulation is also reflected in the weathering hypothesis, which describes early health deterioration due to repeated experience with socioeconomic adversity and political marginalization, and a related concept of allostatic load, or the physiological burden imposed by stress.

Research and Policy Implications

Interdisciplinary life course research draws from such disparate fields as clinical medicine, social epidemiology, the social sciences, human development, and the biological sciences. The complex methodological challenges include difficulties inherent in using large prospective birth cohorts with repeated measures and analytical issues associated with modeling latency periods, hierarchical data, latent (unobserved) variables, dynamic disease trajectories, and multiple time-dependent interactions. Multilevel models, path analysis, and Markov models are just a few of the analytical tools being used to model social and biological pathways and temporal and geographical patterns of disease distribution. As the body of evidence in life course research continues to grow, policymakers may be increasingly challenged to prioritize social and behavioral interventions in pregnancy and early childhood and to address macrolevel determinants of adult health and disease.

—Helen L. Kwon and Luisa N. Borrell

See also Aging, Epidemiology of; Barker Hypothesis; Cardiovascular Disease; Chronic Disease Epidemiology; Geographical and Social Influences on Health; Multilevel Modeling; Social Epidemiology

Further Readings

Barker, D. J. P. (1998). *Mothers, babies, and disease in later life* (2nd ed.). New York: Churchill Livingstone.

Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges, and interdisciplinary perspectives. *International Journal of Epidemiology*, *31*, 285–293.

Kuh, D. L., & Ben-Shlomo, Y. (2004). *A life course approach to chronic disease epidemiology* (2nd ed.). Oxford, UK: Oxford University Press.

Kuh, D. L., & Hardy, R. (2002). *A life course approach to women's health*. Oxford, UK: Oxford University Press.

Life Expectancy

Life expectancy is an estimate of the average number of additional years a person of a given age can expect to live. The most common measure of life expectancy is life expectancy at birth. Life expectancy is a hypothetical measure. It assumes that the age-specific death rates for the year in question will apply throughout the lifetime of those born in that year. The process, in effect, projects the age-specific mortality rates for a given period over the entire lifetime of the population born (or alive) during that time. The measure differs considerably by sex, age, race, and geographic location. Therefore, life expectancy is commonly given for specific categories, rather than for the population in general. For example, the life expectancy for white females in the United States who were born in 2003 is 80.5 years; that is, white female infants born in the United States in the year 2003 would be expected to live, on average, 80.5 years.

Life expectancy reflects local conditions. In less developed countries, life expectancy at birth is relatively low compared with more developed countries. In poor countries, life expectancy at birth is often lower than life expectancy at 1 year of age, because of a high infant mortality rate (commonly due to infectious disease or lack of access to a clean water supply).

Life expectancy is calculated by constructing a life table. The data required to construct a life table are the age-specific death rates for the population in question, which requires enumeration data for the number of people, and the number of deaths at each age for that population. By applying the age-specific death rates that were observed during the period, the average life expectancy for each of the age groups within this population is calculated. Life expectancy at any given age is the average number of additional years a person of that age would live if the age-specific mortality rates for that year continued to apply. For example, U.S. males who were 50 years old in 2003 had an average additional life expectancy of 28.5 years.

The potential accuracy of the estimated life expectancy depends on the completeness of the

Table 1 Expectation of Life by Age, Race, and Sex: United States, 2003

	All Races			*White*			*Black*		
Age	*Total*	*Male*	*Female*	*Total*	*Male*	*Female*	*Total*	*Male*	*Female*
0	77.4	74.7	80.0	77.9	75.3	80.4	72.6	68.9	75.9
1	77.0	74.3	79.5	77.4	74.8	79.8	72.6	69.0	75.9
5	73.1	70.4	75.6	73.4	70.9	76.9	68.7	65.2	72.0
10	68.1	65.5	70.6	68.5	65.9	71.0	63.8	60.2	67.1
15	63.2	60.5	65.7	63.5	61.0	66.0	58.9	55.3	62.1
20	58.4	55.8	60.8	58.7	56.2	61.1	54.1	50.6	57.2
25	53.6	51.2	56.0	54.0	51.6	56.3	49.5	46.2	52.4
30	48.9	46.5	51.1	49.2	46.9	51.4	44.9	41.7	47.7
35	44.1	41.8	46.3	44.5	42.2	46.6	40.3	37.3	43.0
40	39.5	37.2	41.5	39.8	37.6	41.8	35.8	32.9	38.4
45	34.9	32.8	36.9	35.2	33.1	37.1	31.5	28.6	34.0
50	30.5	28.5	32.3	30.7	28.7	32.5	27.4	24.7	29.8
55	26.2	24.3	27.9	26.4	24.5	28.0	23.7	21.1	25.7
60	22.2	20.4	23.7	22.3	20.5	23.7	20.1	17.8	21.9
65	18.4	16.8	19.7	18.4	16.8	19.7	16.8	14.8	18.3
70	14.8	13.4	15.9	14.9	13.4	15.9	13.8	12.0	15.0
75	11.7	10.5	12.5	11.6	10.4	12.5	11.1	9.6	12.1
80	8.9	7.9	9.5	8.8	7.9	9.4	8.8	7.6	9.5
85	6.6	5.9	7.0	6.5	5.8	6.9	6.9	6.0	7.3
90	4.8	4.3	5.0	4.7	4.2	4.9	5.3	4.6	5.6
95	3.5	3.1	3.5	3.4	3.0	3.4	4.1	3.5	4.2
100	2.5	2.2	2.5	2.4	2.1	2.4	3.1	2.7	3.1

Source: Arias (2006, p. 3).

census and mortality data available for the population in question. The completeness of this data varies from country to country. In the United States, official complete life tables based on registered deaths have been prepared since 1900, in connection with the decennial census. Beginning in 1945, annual abridged U.S. life tables have been published based on the annual death registration and estimates of the population. Complete life tables show life expectancy for every year of age, and abridged tables show life expectancy for 5- or 10-year age groups, rather than for single years. The U.S. National Center for Health Statistics is the agency that currently publishes national life tables, as well as state and regional life tables. The United Nations publishes national life tables for many countries in its Demographic Yearbook.

The ''healthy life expectancy'' or ''disability-free life expectancy'' is the average number of years a person is expected to live in good health, or without disability, given current age-specific mortality rates and disease and disability prevalence rates. Calculation of these figures requires reliable health statistics, as well as mortality and census data.

—Judith Marie Bezy

See also Life Tables; Mortality Rates; National Center for Health Statistics

Further Readings

Arias, E. (2006, April). United States life tables, 2003. *National Vital Statistics Reports*, *54*(14), 1–40.

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The methods and materials of demography* (2nd ed.). San Diego, CA: Elsevier Academic Press.

Life Tables

A life table describes, in terms of various life table functions, the probability distribution of duration time in a tabular form. Duration time is the time interval from an initial time point to the occurrence of a point event such as death or incidence of disease. A point event is a transition from one state to another. For example, death is a transition from the alive state (a transient state) to the dead state (an absorbing state). If the initial time point is taken as birth and the point event as death, then the time interval from birth to death is called the *life length* or *age-at-death*, and the life table that describes the distribution of the age-at-death is also called a mortality table (particularly by actuaries). In the case of period life tables, the life length random variable X is further decomposed into two random variables, *current age* A and *future lifetime* Y, and the distributions of all three random variables X, A, and Y are described in a period life table.

Construction of life tables has a long history, dating back to the 1662 Bills of Mortality by John Graunt and the famous 1693 mortality table constructed by Halley for the city of Breslau in what is now Poland. Construction of these two life tables, particularly Halley's table, can be regarded as marking the beginning of modern statistics and public health sciences. Modern life table methods are no longer limited to the study of human mortality and to their use in the calculation of life insurance premiums and annuities. They have been used to study the time to the first marriage and the duration of the marriage (as in the construction of nuptiality tables), the duration of intrauterine device use and birth control effectiveness (birth control life tables), labor force participation (working life tables), renewal of animal and plant populations (life cycles life tables), and so on. Epidemiologists are particularly interested in constructing life tables to measure the probabilities of disease incidence, remission, relapse, and death from competing causes as well as to study the expected duration of stay in healthy and morbid states. Morbidity and mortality life table functions can also be combined to measure the burden of diseases and injuries. These epidemiological uses involve many types of life tables, and constructing them requires considerable technical skills. As can be seen in the following descriptions, life table methods are more sophisticated than most other epidemiological methods.

Types of Life Tables

Life table analysis in each situation requires a new definition of the duration time through appropriate choices of the initial time point and the point event to define a life table type. Many types of life tables can be constructed: If the duration time is a *sojourn*—the length of time a stochastic process remains in a state after entering it, then the associated life

table is called an *attrition life table.* If the observed sojourn is the sojourn in a single transient state until transiting to a single absorbing state, then the associated life table is called an *ordinary life table.* If more than one absorbing state is present, then the observed sojourn is the minimum of the individual sojourns until transiting to each respective absorbing state, and the resulting attrition life tables may include *multiple-decrement life table, net life table, cause-deleted life table*, and *cause-reduced life table.* Construction of these life tables requires consideration of competing risks, as death from one cause automatically prevents an individual from dying from another cause. If the duration time is the *total lifetime*—sum of different sojourns, then the associated life table is known as the *multistate* or *increment-decrement life table.* Whether it is an attrition life table or a multistate life table, two types of life tables can be constructed for use in each case. A *cohort* (or *generation*) *life table* is constructed from data on the initial birth cohort size and vital events as they occur in time in a real birth cohort and reflects the probability distribution of the life length and vital event in question in that cohort. In epidemiology and biostatistics, *modified cohort life tables* have also been constructed for follow-up studies of patient cohorts. *Period* (or *current*) *life table*, on the other hand, is constructed from cross-sectional data, namely, the schedule of vital rates observed during a short calendar time period (usually 1 or 3 years known as the *base period*), and reflects the probability distribution of the life length and vital event in question in a hypothetical cohort in the *stationary population.*

Depending on how finely the age axis is partitioned into intervals, each type of life table defined above can be constructed as an abridged, a complete, or a continuous life table. In an abridged life table, the first two age intervals are of lengths 1 and 4 years, respectively, with the remaining age intervals all having length 5 years. In a complete life table, all age intervals are 1 year long. In a continuous life table, age is treated as a continuous variable and the life table functions can be computed for any age (i.e., any nonnegative real number within the life span). Finally, each type of life table can be constructed for each stratum obtained from cross-classification by calendar time period, geographic region, sex, race, socioeconomic class, occupation, and so on. Table 1 is an abridged period life table for the male population in Canada, constructed from the age-specific population, birth and death data observed in the base period 1990 to 1992.

Period Life Table and Stationary Population

Almost all published life tables are period life tables. Besides providing the basis for probability theory, stochastic processes, mathematical statistics, and population models, period life tables, rather than cohort life tables, are the more common applications for two reasons. (1) Data for constructing period life tables are more readily available, as they are routinely collected by governments, which is not the case with data for constructing cohort life tables. (2) Period life tables can be constructed to provide comparisons of mortality conditions at any past time point up to the present, while cohort life tables cannot be so constructed, as extinction of an entire birth cohort requires the whole life span and so the most recent cohort life table that can be constructed would have to be started a 100 or more years ago. A period life table depicts the distribution of the life length in a hypothetical cohort in the stationary population uniquely determined by the current mortality schedule of the observed population.

A stationary population is a special case of the stable population. It is a population that is closed against migration and characterized by (1) a constant annual number of births and (2) a constant mortality schedule over calendar time. Consequently, the annual number of births always equals the annual number of deaths and so the size of the population is stationary. In a stationary population, a death at any age at a given calendar time is instantly replaced by a birth at that calendar time. Moreover, the age distribution is fixed over calendar time so that the chance of surviving any given age interval is also constant over time. It, therefore, follows that the sequence of instantaneous age-specific death rates (i.e., the *force of mortality*) for any calendar year (period analysis) is identical to the sequence of instantaneous age-specific death rates for any cohort (cohort analysis), and so in a stationary population, a period life table is identical to a cohort life table. That is to say, one can regard a period life table as a hypothetical cohort life table constructed on a stationary population. The age distribution of such stationary population is uniquely determined by the

Table 1 Abridged Life Table for Canadian Males, 1990 to 1992

(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Start of the i*th Age Interval* $[x_i, x_{i+1})$	*Conditional Probability of Dying in Interval*	*First-Order Survival Function*	*Death Density Function at Age*	*Hazard Function at Age* x_i	*Second-Order Survival Function at Age* x_i	*Stationary Population Segment*	*Individual Life Expectancy at Age* x_i	*Third-Order Survival Function at Age* x_i	*Population Life Expectancy at Age* x_i
x_i	q_i	$l(x_i)$	$f(x_i)$	$h(x_i)$	$T(x_i)$	L_i	$e(x_i)$	$Y(x_i)$	$\varepsilon(x_i)$
0	.00758	100,000	.03947	.03947	7,433,284	99,357	74.33	289875994	39.90
1	.00157	99,242	.00094	.00095	7,333,927	395,991	73.90	282491142	38.52
5	.00103	98,946	.00016	.00016	6,937,936	494,480	70.12	253945670	36.60
10	.00131	98,845	.00017	.00017	6,443,456	493,979	65.19	220489298	34.22
15	.00462	98,715	.00056	.00057	5,949,477	492,552	60.27	189504426	31.85
20	.00581	98,259	.00112	.00114	5,456,925	489,872	55.54	160986563	29.50
25	.00580	97,688	.00114	.00116	4,967,053	487,029	50.85	134925085	27.16
30	.00648	97,121	.00117	.00120	4,480,024	484,079	46.13	111305673	24.84
35	.00794	96,492	.00138	.00143	3,995,945	480,611	41.41	90114252	22.55
40	.01052	95,726	.00170	.00177	3,515,334	476,276	36.72	71334841	20.29
45	.01684	94,719	.00248	.00262	3,039,058	469,921	32.08	54948149	18.08
50	.02751	93,124	.00399	.00428	2,569,137	459,739	27.59	40928175	15.93
55	.04632	90,562	.00650	.00718	2,109,398	443,164	23.29	29234364	13.86
60	.07629	86,367	.01054	.01220	1,666,234	416,510	19.29	19801232	11.88
65	.11915	79,778	.01604	.02011	1,249,724	376,338	15.67	12522223	10.02
70	.17784	70,273	.02186	.03110	873,386	321,515	12.43	7231441	8.28
75	.27287	57,775	.02855	.04942	551,871	250,498	9.55	3691536	6.69
80	.39491	42,010	.03351	.07977	301,373	168,162	7.17	1588560	5.27
85	.54434	25,420	.03153	.12404	133,211	90,716	5.24	533752	4.01
90 +	1	11,583	.02293	.19797	42,495	42,495	3.67	120504	2.84

Source: Based on the method developed by the author described in Hsieh 1991a and 1991b.

mortality schedule used to construct the period life table. Thus, different mortality schedules would determine different stationary age distribution and hence different period life tables.

In interpreting the life table functions, the fact that the life table population is a stationary population should always be borne in mind. Thus, the life expectancy function computed from mortality rates observed in the year 1921, say, will underestimate the mean life length of the generation born that year because of the reduction in the force of mortality since then. In general, with the trend toward lower mortality, the expectation of life at birth computed from a period life table would understate the average length of life of a newly born generation. To determine the actual life expectancy for a generation, it is necessary to construct many period life tables at consecutive time periods and to concatenate them to form a generation life table. The direct application of period life tables is for comparative studies. Period life tables provide valid comparisons of mortality among different populations because the life table death rates, or equivalently the life expectancies, arise from the stationary population rather than from the actual observed population and hence are independent of the age distribution of the observed populations and so are devoid of any confounding by differential age distributions among the different observed populations.

Description of Period Life Table Functions

A period life table consists of many columns with age being the first column followed by several life table functions (see Table 1). Each life table function is a function of age and describes the distribution of the life length in a different way—in terms of density function, conditional distribution function, survival function, generalized survival function, and conditional expectation. The purpose of the tabular display is to provide simultaneously several life table function values at each of many different ages. A cohort life table has two life table columns less than a period life table, with the last two columns $Y(x_i)$ and $\varepsilon(x_i)$ excluded.

In Table 1, the first column, age, is the independent variable, which ranges from 0 to $90+$, or $100+$, or $110+$, depending on the type of life table to be constructed. The age axis is partitioned into age intervals $[x_i, x_{i+1})$ with interval length $\Delta x_i \equiv x_{i+1} - x_i = 1$ for all integral values x_i for a complete life table and $\Delta x_i = 5$ except for the first two age intervals in which $\Delta x_i = 1$ for the first age interval and $\Delta x_i = 4$ for the second interval, for $x_i = 0, 1, 5$ (5) 90 for an abridged life table. The remaining columns are the life table functions. The ordinary life table may include up to nine or more main life table functions (nine or more for period life table and seven or more for cohort life table). The unconditional life table functions in Columns 3, 4, 6, and 7 have two interpretations: The first depicts the probability distribution of the life length in a hypothetical cohort, and the second describes the stationary population structure. The remaining conditional life table functions have only one interpretation with Columns 2, 5, and 8 having the first (hypothetical cohort) interpretation and Columns 9 and 10 having the second (stationary population) interpretation.

Column 2, conditional probability of death, $q_i = \Pr\{X < x_{i+1} | x \geq x_i\} = 1 - l(x_{i+1})/l(x_i)$, which is the conditional probability of dying before age x_{i+1} given survival to age x_i (hypothetical cohort interpretation).

Column 3, survivorship function, $l(x_i) = 10^5 \Pr\{X > x_i\}$, which is the expected number surviving to age x_i out of the initial cohort of 100,000 live births (hypothetical cohort interpretation). This implies that $l(x_i)$ is proportional to the survival function $S(x_i) = \Pr\{X > x_i\}$ of the life length X or the expected number alive at age x_i per year in a stationary population supported by 100,000 annual live births (stationary population interpretation), which implies that $l(x_i)$ is proportional to the density function of the current age A and of the future lifetime Y.

Column 4, death density function,

$$f(x_i) = \lim_{\Delta x_i \downarrow 0} \frac{\Pr\{x_i \leq X < x_i + \Delta x_i\}}{\Delta x_i},$$

which is the probability per unit time of dying in the age interval $[x_i, x_i + \Delta x_i)$ $(M-1)$ as Δx_i tends to 0, for a member of the initial cohort of live births (hypothetical cohort interpretation) or the expected proportion of deaths per year in the age interval $[x_i, x_i + \Delta x_i)$ $(M-1)$ as Δx_i tends to 0, among the live births during any fixed calendar time period in a stationary population (stationary population interpretation).

This function describes the left-skewed distribution of the life length X, implying that the mean (74.34 years, obtained from Column 8 below) is less than the

median (77.4 years, obtained from Column 3 above), which in turn is less than the mode (81.2 years, obtained from Column 4). Note that this curve has two peaks, one at the start of life and the other at age 81.2 years and that the area under the whole curve is one.

Column 5, hazard function (force of mortality),

$$h(x_i) = \lim_{\Delta x_i \downarrow 0} \frac{\Pr\{X < x_i + \Delta x_i | X \geq x_i\}}{\Delta x_i},$$

which is the probability per unit time of dying in the age interval $[x_i, x_i + \Delta x_i)$ as Δx_i tends to 0, for a member of the cohort surviving to age x_i (hypothetical cohort interpretation). Note that $h(0) = f(0)$ and $h(x) = f(x)/S(x) > f(x)$ for all $x > 0$, so that $S(x) = \exp(-\int_0^x h(u)du)$. Notice also the almost exponential growth of the hazard function after age 30 years.

Column 6, second-order survivorship function, $T(x_i) = \int_{x_i}^{\omega} l(u)du$, which is the expected number of person-years lived beyond age x_i by the $l(x_i)$ survivors of the initial cohort of 100,000 live births (hypothetical cohort interpretation), which implies that $T(x_i)$ is proportional to the second-order survival function $S^{(2)}(x_i) \equiv \int_{x_i}^{\infty} S(u)du$ of the life length X or the expected population size beyond age x_i in a stationary population supported by 100,000 annual live births (stationary population interpretation), which implies that $T(x_i)$ is proportional to the survival function of the current age A and of the future lifetime Y. The second interpretation of this function can be used to estimate the annual number of births from the observed population size.

Column 7, stationary population segment, $L_i = \int_{x_i}^{x_{i+1}} l(u)du$, which is the expected number of person-years lived in the age interval $[x_i, x_i + \Delta x_i)$ by the $l(x_i)$ survivors of the 100,000 initial birth cohort (hypothetical cohort interpretation) or the expected population size in the age interval $[x_i, x_i + \Delta x_i)$ in a stationary population supported by 100,000 annual live births (stationary population interpretation). The second interpretation of this function can be used to make population projection. This column can also be used in combination with Column 3 to form age-specific life table death rates $m_i \equiv [l(x_i) - l(x_{i+1})]/L_i$ for precise age-adjusted mortality comparisons and other uses.

Column 8, individual life expectancy function, $e(x_i) \equiv E\{X - x_i | X > x_i\} = T(x_i)/l(x_i)$, which is the expected remaining life length conditional on survival to age x (hypothetical cohort interpretation). This function may be used to compare summary mortality levels of different populations.

Column 9, third-order survivorship function, $Y(x_i) = \int_{x_i}^{\omega} T(u)du$, which is the expected future person-years lived by the $T(x)$ people aged x and above in the stationary population supported by 100,000 annual live births (stationary population interpretation).

Column 10, population life expectancy function, $\varepsilon(x_i) \equiv E[Y|A \geq x_i] = Y(x_i)/T(x_i)$, which is the expected remaining lifetime for the $T(x_i)$ people aged x_i and above in the stationary population supported by 100,000 annual live births (stationary population interpretation). This function is a statistically more robust estimator of summary mortality than Column 8, $e(x_i)$.

In addition to these main life table functions, higher-order survival functions and the variances of their empirical life table functions can also be calculated, which are needed for statistical inference (see Application section below).

Construction of Period Life Tables

To construct a life table means to calculate the life table functions from the observed age-specific exposure and decrement data. Large population cohort life tables can be directly constructed from the cohort data by using simple arithmetic operations, division, multiplication, addition, and subtraction, to calculate the life table functions. However, for human populations, construction of a complete cohort life table would require more than 100 years of follow-up data that are rather difficult to obtain. To construct a period (current) life table is a totally different matter, as it requires conversion of the cross-sectional observed data on age-specific deaths and populations or their combinations, death rates, observed in a short base period of 1 to 3 years into the hypothetical cohort (stationary population) life table functions in terms of conditional probabilities and expectations. Construction of period (current) life tables requires a great deal of mathematical skill, because conversion from observed age-specific death rates M_i into conditional probability of death q_i requires accurate approximations and so are the evaluation of death density function $f(x)$, hazard function $h(x)$, and person-year functions $T(x)$ and $Y(x)$ that are all derived from the q_i function. Different life table methods render different solutions to these approximation

problems and so produce slightly different life table function values. The life table shown in Table 1 is constructed using Hsieh's method of 1991 that has been tested to be more accurate than other existing methods. For detailed methods of life table construction and derivation of the standard errors of empirical life table functions, see references listed in Further Readings, particularly those of Chiang, Hsieh, and Keyfitz.

Applications of Life Table Functions

The superiority of the life table method over other epidemiological methods, such as direct and indirect standardization, is that comparisons using standardization methods depend on the choice of the standard population to adjust the age-specific rates and such adjustment can lead to opposing conclusions due to arbitrariness of the choice of the standard population. On the other hand, all period (current) life tables are defined by the stationary population model whose age-sex distribution has a fixed bell shape and so the comparison using period life tables is independent of the age distribution of the observed populations. Thus, the life table method provides a standardized comparison free of age confounding with definitive conclusion. Life table functions have many applications, several of which have been described in the previous sections. Here, we shall concentrate on their applications in epidemiology and public health.

Ordinary cohort and period (current) life tables may be used to compare the intensity of the point event of interest between two strata (such as two geographic regions) by using two-sample Z-test on any of the corresponding life table functions from each stratum. This is because all empirical life table functions are asymptotically Gaussian. For example, to compare mortality for a given age interval $[x, x + \Delta x]$ between two strata, we may use the function ${}_{\Delta x}q_x$. To compare mortality for ages beyond x, we may use the function $e(x)$; and to compare mortality for ages before x, we may use the function $l(x)$; to compare mortality for all ages, we may use the value $e(0)$. The same four life table functions from the net life table may be used to compare mortality from a given cause between the two strata, adjusting for the competing risks.

To evaluate the impact of a given cause on longevity of a given population, we can compare a cause-deleted life table with the ordinary life table in terms of the differences of three ordinary life table functions and the corresponding functions with cause k eliminated: ${}_{\Delta x}q_x - {}_{\Delta x}q_x^{\bullet k}$, $[l^{\bullet k}(x) - l(x)]/1(0)$, and $e^{\bullet k}(x) - e(x)$, which represent potential reduction in age-specific mortality rate, gain in survival probability, and gain in life expectancy, respectively, if cause k were eliminated as a risk of death. We may also compare the percentage change by computing $[{}_{\Delta x}q_x - {}_{\Delta x}q^{\bullet k}x]/{}_{\Delta x}q_x$, $[l^{\bullet k}(x) - l(x)]/l(x)$, and $[e^{\bullet k}(x) - e(x)]/e(x)$.

As the survivorship function increases toward its natural rectangular limit, ordinary cohort and period (current) life tables become less useful as health measures. But they can be used to compute disability-free life expectancy (DFLE) by combining the life table function $l(x)$ (mortality) with the age-specific disability prevalence estimated from cross-sectional survey data (morbidity), as done in Sullivan's method. Disability-adjusted life expectancy (DALE) can also be calculated by weighting different levels of disabilities. Alternatively, DFLE may be obtained by constructing a multistate life table with three states: healthy, disabled, and dead. DFLE may be calculated as the area between the survival curve against the disable state and the survival curve against the dead state to the right divided by the height of the survival curve against the dead state.

Ordinary cohort and period (current) life tables may be used to compute years of life lost (YLL) by summing the products of observed number of death at each age and the age weighted, discounted life expectancy at that age for the loss of life through death. We can also compute years lived with disability (YLD) to account for loss of quality of life through disability by summing the products of observed number of disability incidences at each age, the severity score of the disability, and the duration of disability, the last being obtainable by constructing a multistate life table with two transient states: healthy, disability (illness) and one absorbing state (death), and by subtracting the area under the survival curve for death from the area under the survival curve for disability (or illness). Addition of YLL to YLD yields DALY, disability-adjusted life years that have been used to measure and compare the burden of disease.

Counts of survivors and deaths in an ordinary life table may be treated as point processes and used to compute quantities of use in epidemiology and demography, such as conditional probabilities of

death and expected life lengths, in any regions of the Lexis diagram given surviving a certain region of the Lexis diagram. For multiple decrement and multistate life tables, the point process methods can be used to compute conditional probabilities of stay in each state and expected duration of stay in any transient state given surviving a certain region of the Lexis diagram.

The survivorship function $l(x)$, the second-order survivorship function $T(x)$, and the third-order survivorship function $Y(x)$ in the ordinary and net life tables can be used to compute the Lorenz curve $1-[T(x)+xl(x)]/T(0)$, scaled total-time-on-test $1-T(x)/T(0)$, Gini index $2\int_0^1\{x-1+[T(x)+xl(x)]dx\}/T(0)$ and variances of their empirical estimates for comparative statistical analysis of morbidity and mortality in public health using Central Limit Theorem for Brownian bridge and Brownian motion processes.

—*John J. Hsieh*

See also Birth Cohort Analysis; Direct Standardization; Global Burden of Disease Project; Graunt, John; Hazard Rate; Life Expectancy; Mortality Rates

Further Readings

Bowers, N., Gerber, H., Hickman, J., Jones, D., & Nesbitt, C. (1986). *Actuarial mathematics.* Itasca, IL: Society of Actuaries.

Chiang, C. L. (1968). *Introduction to stochastic processes in biostatistics.* New York: Wiley.

Chiang, C. L. (1984). *The life table and its applications.* Malabar, FL: Robert E. Krieger.

Colvez, A., & Blanchet, M. (1983). Potential gains in the life expectancy free of disability: A tool for health planning. *International Journal of Epidemiology*, *12*, 224–229.

Ebert, T. A. (1999). *Plant and animal populations: Methods in demography.* New York: Academic Press.

Hsieh, J. J. (1985). Construction of expanded infant life tables: A method based on a new mortality law. *Mathematical Biosciences*, *76*, 221–242.

Hsieh, J. J. (1989). A probabilistic approach to the construction of competing-risk life tables. *Biometrical Journal*, *31*, 339–357.

Hsieh, J. J. (1991a). A general theory of life table construction and a precise abridged life table method. *Biometrical Journal*, *33*, 143–162.

Hsieh, J. J. (1991b). Construction of expanded continuous life tables: A generalization of abridged and completed life tables. *Mathematical Biosciences*, *103*, 287–302.

Keyfitz, N., & Caswell, H. (2005). *Applied mathematical demography.* New York: Springer.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modelling population process.* Oxford, UK: Blackwell.

Preston, S. H., Keyfitz, N., & Schoen, R. (1972). *Causes of death: Life tables for national populations.* New York: Seminar Press.

Rogers, A. (1995). *Multiregional demography: Principles, methods and extensions.* Chichester: Wiley.

Salomon, J. A., & Murray, C. (2002). The epidemiologic transition revisited: Compositional models for causes of death by age and sex. *Population and Development Review*, *28*, 205–228.

Schoen, R. (1988). *Modeling multigroup populations.* New York: Plenum Press.

Sullivan, D. F. (1971). A single index of mortality and morbidity. *Health Services and Mental Health Administration Health Reports*, *86*, 347–354.

Likelihood Ratio

The likelihood ratio is used to compare different data models, perform hypothesis tests, and construct confidence intervals. The likelihood of a model is a measure of how well the model fits the observed data. Model likelihood is most often used when fitting a parametric model (e.g., a normal distribution) to data to find the parameters that best describe the data. The general formula for the likelihood function is

$$L(\theta|\mathrm{D}) = P(\mathbf{D}|\theta),$$

where

$\boldsymbol{\theta}$ is the model parameterization,

$\mathbf{D}$ is the observed data,

$L(\boldsymbol{\theta}|\mathbf{D})$ is the likelihood of the model parameterization given the observed data $\mathbf{D}$, and

$P(\mathbf{D}|\boldsymbol{\theta})$ is the probability density function of the data given the model parameterization $\boldsymbol{\theta}$.

The basis for this formulation of the likelihood function is best understood by using Bayes's theorem to calculate the probability of the model given the observed data:

$$P(\theta|\mathbf{D}) = \frac{P(\mathbf{D}|\theta)P(\theta)}{P(\mathbf{D})}.$$

If we assume that all models are equally likely and note that the probability of the observed data is fixed, this formula is proportional to the likelihood formula given above. That is, for any two model parameterizations if the likelihood of one parameterization is greater than that of the other, then the probability of that parameterization is also greater by the above formula.

When comparing two models, we use their likelihood ratio. The formula for the likelihood ratio is

$$\lambda(\mathbf{D}) = \frac{L(\theta_0|\mathbf{D})}{L(\theta_1|\mathbf{D})}.$$

In hypothesis testing, the top model parameterization ($\boldsymbol{\theta}_0$) represents the null hypothesis and the bottom parameterization ($\boldsymbol{\theta}_1$) represents the alternative hypothesis. We will reject the null hypothesis in favor of the alternative hypothesis if $\lambda(\mathbf{D}) < c$, where c is some preselected critical value. Confidence intervals, sometimes referred to as supported intervals when using this approach, can be calculated in a similar manner. In this approach, the value of $\boldsymbol{\theta}$ that maximizes the likelihood (the maximum likelihood estimate) is used to determine the denominator ($\boldsymbol{\theta}_1$) and alternative parameterizations are used in the numerator ($\boldsymbol{\theta}_0$). The supported region consists of those values of ($\boldsymbol{\theta}_0$) where $\lambda(\mathbf{D}) < c$, where c is some critical value. Some statisticians and epidemiologists argue that these approaches are superior to traditional approaches because they incorporate the probability of the observed data given the alternative models into the calculation, not only the probability under the null hypothesis.

Another common use of the likelihood ratio occurs when performing a likelihood ratio test. A likelihood ratio test is used when comparing a model with fewer parameters with a more complex one (i.e., one with more parameters). In general, the more complex model will fit the observed data better. To determine if this increase in goodness of fit is enough to justify the increase in model complexity, we first calculate the likelihood ratio test statistic:

$$LR = -2\frac{\ln L_0}{\ln L_1},$$

where

L_0 is the likelihood of the simpler model using the maximum likelihood parameter estimate and

L_1 is the likelihood of the more complex model using the maximum likelihood parameter estimate.

This statistic is then used as the test statistic in a chi-squared test with n degrees of freedom, where n is the number of additional parameters in the more complex model. If the chi-squared test is greater than the $1-\alpha$ region of a chi-squared distribution, the more complex model is accepted as valid; otherwise the simpler model is used.

—*Justin Lessler*

See also Bayes's Theorem; Chi-Square Test; Degrees of Freedom; Likelihood Ratio

Further Readings

Goodman, S. N. (1993). P-values, hypothesis tests, and likelihood: Implications for epidemiology of a neglected historical debate. *American Journal of Epidemiology*, *137*, 485–495.

Hills, M. (1996). *Statistical models in epidemiology*. Oxford, UK: Oxford University Press.

Likert Scale

Likert scales are rating scales used in questionnaires that measure people's attitudes, opinions, or perceptions. Subjects choose from a range of possible responses to a specific question or statement, such as "strongly agree," "agree," "neutral," "disagree," "strongly disagree." Often, the categories of response are coded numerically, in which case the numerical values must be defined for that specific study, such as $1 = \textit{strongly agree}$, $2 = \textit{agree}$, and so on. Likert scales are named for Rennis Likert, who devised them in 1932.

Likert scales are widely used in social and educational research. Epidemiologists may employ Likert scales in surveying topics such as attitudes toward health or toward specific behaviors that affect health (e.g., smoking); opinions about the relative importance, efficacy, or practicality of different treatment options; and public perceptions about health care, the role of health professionals, or risk factors for specific diseases. When using Likert scales, the researcher must consider issues such as categories of response (values in the scale), size of the scale,

direction of the scale, the ordinal nature of Likert-derived data, and appropriate statistical analysis of such data.

Categories of Response

Generally, a Likert scale presents the respondent with a statement and asks the respondent to rate the extent to which he or she agrees with it. Variations include presenting the subject with a question rather than a statement. The categories of response should be mutually exclusive and should cover the full range of opinion. Some researchers include a ''don't know'' option, to distinguish between respondents who do not feel sufficiently informed to give an opinion and those who are ''neutral'' on the topic.

Size of Likert Scales

The size of Likert scales may vary. Traditionally, researchers have employed a 5-point scale (e.g., strongly agree, agree, neutral, disagree, strongly disagree). A larger scale (e.g., seven categories) could offer more choice to respondents, but it has been suggested that people tend not to select the extreme categories in large rating scales, perhaps not wanting to appear extreme in their view. Moreover, it may not be easy for subjects to discriminate between categories that are only subtly different. On the other hand, rating scales with just three categories (e.g., poor, satisfactory, good) may not afford sufficient discrimination. A current trend is to use an even number of categories, to force respondents to come down broadly ''for'' or ''against'' a statement. Thus, 4-point or 6-point Likert scales are increasingly common.

Directionality of Likert Scales

A feature of Likert scales is their directionality: The categories of response may be increasingly positive or increasingly negative. While interpretation of a category may vary among respondents (e.g., one person's ''agree'' is another's ''strongly agree''), all respondents should nevertheless understand that ''strongly agree'' is a more positive opinion than ''agree.'' One important consideration in the design of questionnaires is the use of reverse scoring on some items. Imagine a questionnaire with positive statements about the benefits of public health education programs (e.g., ''TV campaigns are a good way to persuade people to stop smoking in the presence of children''). A subject who strongly agreed with all such statements would be presumed to have a very positive view about the benefits of this method of health education. However, perhaps the subject was not participating wholeheartedly and simply checked the same response category for each item. To ensure that respondents are reading and evaluating statements carefully, it is good practice to include a few negative statements (e.g., ''Money spent on public health education programs would be better spent on research into new therapies''). If a respondent answers positively to positive statements and negatively to negative statements, the researcher may have increased confidence in the data. Thus, it is good practice to employ reverse scoring of some items.

Ordinal Measures and the Use of Descriptive and Inferential Statistics

Likert scales fall within the ordinal level of measurement: The categories of response have directionality, but the intervals between them cannot be presumed equal. Thus, for a scale where $1 = \textit{strongly agree}$, $2 = \textit{agree}$, $3 = \textit{neutral}$, $4 = \textit{disagree}$, and $5 = \textit{strongly disagree}$, we can say 4 is more negative than 3, 2, or 1 (directionality); but we cannot infer that a response of 4 is twice as negative as a response of 2.

Deciding which descriptive and inferential statistics may legitimately be used to describe and analyze the data obtained from a Likert scale is a controversial issue. Treating Likert-derived data as ordinal, one should employ the median or mode as the measure of central tendency. In addition, one may state the frequency/percentage frequency of responses in each category. The appropriate inferential statistics for ordinal data are those employing nonparametric tests, such as chi-square, Spearman's rho, or the Mann-Whitney U test.

However, many researchers treat Likert-derived data as if it were at the interval level (where numbers on the scale not only have directionality but also are an equal distance apart). They use the mean and standard deviation to describe their data and analyze it with ''powerful'' parametric tests, such as ANOVA or Pearson's product-moment correlation, arguing this is legitimate provided one states the assumption that the data are interval level. Calculating the mean, standard deviation, and parametric statistics requires arithmetic manipulation of data (e.g., addition, multiplication).

Since numerical values in Likert scales represent verbal statements, one might question whether it makes sense to perform such manipulations. Moreover, Likert-derived data may fail to meet other assumptions for parametric tests (e.g., a normal distribution). Thus, careful consideration must also be given to the appropriate descriptive and inferential statistics, and the researcher must be explicit about any assumptions made.

—*Susan Jamieson*

See also Chi-Square Test; Measures of Central Tendency; Nonparametric Statistics; Normal Distribution; Questionnaire Design

Further Readings

Burns, R. B. (2000). Attitude surveys. In R. B. Burns (Ed.), *Introduction to research methods* (4th ed., pp. 555–565). London, UK: Sage.

Clegg, F. (1998). *Simple statistics.* Cambridge, UK: Cambridge University Press.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London, UK: Routledge Falmer.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*, 1217–1218.

Lind, James
(1716–1794)

James Lind, the founder of British naval hygiene, was a Scottish physician who discovered the cause of scurvy, a dietary deficiency due to lack of vitamin C. His legacy as one of the first modern clinical investigators reflected his desire to improve the health of soldiers and sailors.

Lind was born in Edinburgh, Scotland, the son of Margaret Smelum and James Lind, a merchant. In 1731, Lind registered as an apprentice to George Langlands, an Edinburgh physician. Lind began his naval career in 1739 as a surgeon's mate and was promoted to surgeon in 1747.

While serving on the *H.M.S. Salisbury* in 1747, Lind carried out experiments on scurvy. He selected 12 men from the ship, all suffering from symptoms of scurvy and divided them into six pairs. He then gave each group different supplements to their basic diet. Two men were given an unspecified elixir three times a day; two were treated with seawater; two were fed with a combination of garlic, mustard, and horseradish; two were given spoonfuls of vinegar; two received a quart of cider a day; and the last two were given two oranges and one lemon every day. Lind recorded no improvement with the first four diets, slight improvement in the men given cider, and significant improvement in those fed citrus fruit.

In 1748, Lind left the navy and obtained his M.D. from the University of Edinburgh later that year. In 1750, Lind became fellow of the Royal College of Physicians of Edinburgh and later served as treasurer from December 1756 to 1758. He published *A Treatise of the Scurvy* in 1753. Although the *Treatise* garnered little acclaim at the time, it attracted the attention of Lord Anson, then first Lord of the Admiralty, and to whom it was dedicated. Lord Anson was influential in securing Lind's appointment to the Royal Naval Hospital at Haslar in 1758. During the 1760s, Lind published several treatises on preventive and tropical medicine, proposing a simple method of supplying ships with fresh water distillation and providing advice on the prevention of tropical fevers. When Lind retired from the Naval Hospital in 1783, his son John succeeded him as chief physician.

Lind was never elected a fellow of the Royal Society, nor were his dietary recommendations immediately realized. Forty years passed before an official Admiralty Order on the supply of lemon juice to ships was issued in 1795, a year after his death. When this order was implemented, scurvy disappeared from the Fleets and Naval hospitals.

—*Todd M. Olszewski*

See also Nutritional Epidemiology; Vitamin Deficiency Diseases

Further Readings

Carpenter, K. J. (1986). *The history of scurvy and vitamin C.* Cambridge, UK: Cambridge University Press.

Lind, J. (1753). *A treatise of the scurvy.* Edinburgh, UK: Millar.

Lind, J. (1757). *An essay on the most effectual means of preserving the health of seamen in the Royal Navy*. London: D. Wilson.

Lind, J. (1768). *An essay on diseases incidental to Europeans in hot climates, with the method of preventing their fatal consequences*. London: T. Becket & P. A. de Hondt.

Linear Regression

See Regression

Linkage Analysis

Linkage analysis is used to pinpoint the location of disease genes within the genome. Given genetic data from a family with a strong history of a disease, it is possible to trace the inheritance of the disease through the family, and thereby localize the genetic region (or locus) responsible for disease in that family. In practice, linkage analysis combines the biological rules of inheritance with statistical inference to identify a linked locus.

Linkage analysis is based on research conducted by Gregor Mendel in the 1860s. Mendel's second law of inheritance states that genetic loci are inherited independently of one another. This independent assortment occurs because genes are located on individual chromosomes, which are redistributed randomly every time an egg or sperm cell is produced. This is why siblings have similar, but not identical, traits; each sibling arose from a unique assortment of parental chromosomes. When two loci are located on separate chromosomes, they will be inherited independently, meaning that the loci will be distributed to the same offspring about 50% of the time. However, loci on the same chromosome are inherited together more frequently. Because portions of each chromosome are rearranged during meiosis, two loci that are near each other are inherited together more frequently than loci distant from each other on the same chromosome. Thus, it is possible to determine the relative position of loci by examining inheritance patterns in a family.

Linkage analysis traces inheritance patterns, and therefore must use genetic and phenotypic data collected from groups of related individuals, some of which have the trait of interest. In most linkage studies, genetic data are collected by genotyping family members for several hundred genetic markers distributed at known locations throughout the genome. Commonly used markers are microsatellite markers and single nucleotide polymorphisms. Researchers then use statistical software to determine which markers are likely to be near the trait locus, based on marker inheritance patterns in affected and unaffected family members.

Linkage analysis requires making several assumptions. First, it is assumed that the trait of interest is genetic. This is usually established by performing a familial aggregation analysis, which can determine if relatives of an affected individual are more likely to have a trait than individuals in the general population. However, familial aggregation of a trait can also be caused by nongenetic factors, including shared environmental exposures and shared behaviors. Second, linkage analysis assumes that the trait follows Mendelian rules of inheritance and that the mode of inheritance is known. This assumption can be tested via segregation analysis that uses statistical tests to identify the most likely mode of inheritance for a trait.

Several phenomena can complicate linkage analyses. Genetic heterogeneity, a situation in which several distinct genes cause the same phenotype via different pathways, can make it nearly impossible to establish linkage. Epistasis, a state in which two or more genes interact to cause a phenotype, will also obscure linkage. While linkage analysis can provide clues about the location of a gene associated with a phenotype, it does not identify a causal allele or mutation. Linkage analyses in several families may identify the same locus linked to the same disease, but each family's disease may be caused by a unique genetic variant. For this reason, linkage analyses are often followed by a genetic association study, which can identify causal alleles in the candidate gene identified by linkage analysis.

In 1990, Hall and colleagues published results of a linkage analysis that linked a region on Chromosome 17 to early-onset familial breast cancer. This analysis used data from 329 individuals within 23 families with a history of early-onset breast cancer. This gene was later identified as a tumor suppressor and named BRCA1. Today, many women from high-risk families choose to be tested for BRCA1 mutations to predict their personal risk for early-onset breast and ovarian cancer.

—Megan Dann Fesinmeyer

See also Association, Genetic; Family Studies in Genetics; Genetic Epidemiology; Genetic Markers

Further Readings

Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H., & Wjst, M. (2001). Genomewide scans of complex human diseases: True linkage is hard to find. *American Journal of Human Genetics*, *69*(5), 936–950.

Elston, R. C. (1998). Methods of linkage analysis and the assumptions underlying them. *American Journal of Human Genetics*, *63*(4), 931–934.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., et al. (1990). Linkage of early-onset familial breast cancer to Chromosome 17q21. *Science*, *250*(4988), 1684–1689.

Khoury, M. J, Beaty, T. H., & Cohen, B. H. (1993). *Fundamentals of genetic epidemiology.* New York: Oxford University Press.

Lister, Joseph
(1827–1912)

Joseph Lister was a British surgeon best remembered today for pioneering antiseptic techniques to reduce infection rates, and thus morbidity and mortality, following surgical procedures. Lister was born to a Quaker family in Upton, Essex, England; his father was the physicist Joseph Jackson Lister, who invented the achromatic microscope. Joseph Lister studied medicine at University College, London, and graduated in 1852. He served as a resident in University Hospital, London, and was then appointed as an assistant surgeon at the Edinburgh Royal Infirmary in 1856, where he worked under James Syme. In 1859, Lister was appointed to the Regius Professorship of Surgery at Glasgow University, and in 1861, he became surgeon of the Glasgow Royal Infirmary.

When Lister began his career, the mortality rate for surgery was nearly 50%, and a common complication of surgery was infection. It was commonly believed that wound infections or sepsis were caused by oxidation that could be prevented by not allowing exposure to the air. However, following an idea of Louis Pasteur's, Lister believed that wound infection was caused by microorganisms and if the organisms could be prevented from entering the wound site, the infection could be prevented.

Lister began using carbolic acid (previously demonstrated to kill parasites in sewage) as an antiseptic solution in the wards he supervised at the Glasgow Royal Infirmary; it was used to clean and dress wounds and to disinfect surgical instruments, the air in the operating theatre was sprayed with a mist of carbolic acid, and surgeons were required to wash their hands with a carbolic acid solution before operating. There were some complications, primarily skin irritation, from the carbolic acid, but infection rates plummeted with the adoption of these practices. Lister reported, at a British Medical Association meeting in 1867, that the wards he supervised at the Glasgow Royal Infirmary had remained free of sepsis for 9 months. Despite these obvious successes, Lister's ideas were slow to gain support in England, where many surgeons resented having to perform extra procedures they felt were unnecessary, and who rejected the implication that they might be a cause of infection. His ideas were adopted more readily in Germany, and antiseptic procedures saved many lives during the Franco-Prussian War (1870–1871). Lister's theory that wound infections were caused by microorganisms was further reinforced by the demonstration in 1878 by the German physician Robert Koch that heat sterilization of surgical instruments and dressings dramatically reduced infection rates.

Lister became Chair of Surgery in King's College, London, in 1877, and was knighted by Queen Victoria in 1883. He converted many skeptics to the merits of antiseptic surgery following a successful operation to repair a fractured patella (kneecap), an operation considered inordinately risky at the time. He was also one of the first British surgeons to operate on a brain tumor and developed improved techniques for mastectomy and for the repair of the patella.

—Sarah Boslaugh

See also Observational Studies; Public Health, History of

Further Readings

Fisher, R. B. (1977). *Joseph Lister, 1827–1912*. London: Macdonald & James.

Goldman, M. (1987). *Lister Ward.* Bristol, UK: Adam Hilger.

Lister, J. (2004). *On the antiseptic principle of the practice of surgery*. (Originally published 1867). Retrieved October 20, 2006, from http://etext.library.adelaide.edu.au/l/lister/joseph/antiseptic/index.html.

Locus of Control

The construct *locus of control*, also referred to as perceived control, is one of the most studied and cited dispositional constructs in psychology and the social behavioral sciences, and it plays an important role in public health research and health behavior interventions. Locus of control may be either internal

or external. Rotter (1990) explains this distinction in terms of the degree to which people assume that the results of their behavior depend on their own actions or personal characteristics rather than on chance, luck, or the influence of powerful others. The popularity of the locus of control construct in health research is demonstrated by the existence of almost 2,500 publications in the *PsycInfo* database for the years 1967 to 2006, which are indexed by the combination of keywords *locus of control* and *health.*

Understanding Locus of Control

The locus of control construct is rooted in social learning theory, and the foundation for this work comes out of research on human and animals. For example, Herbert Lefcourt notes that when rats were able to exercise control over an aversive stimulus, they exhibited less fear of that stimulus than if they could not exercise such control. Similar results have been found in human studies. For example, when participants believed that their behavior could reduce electric shock duration, they gave lower ratings of the painfulness and aftereffects of shocks, compared with when they thought they did not have control. Anecdotal clinical observations support the importance of the perceived control construct in behavior change; for instance, Lefcourt noted that some clients learned from psychological therapy and other new experiences and subsequently changed their behavior, but other clients did not change their behavior as a result of these experiences. Often, the latter would attribute their lack of change to the belief that it was really other people, not themselves, who controlled relevant outcomes for them. In social learning theory terms, the construct of perceived control is a generalized expectancy of external or internal control of reinforcement, for either positive or negative events. It is an abstraction derived from many expectancy-behavior-outcome cycles in which people viewed the causes of their success or failure as being under internal or external control. A person's actions are a function of the situation, expectations, and values. More specifically, the probability that behavior *B* by person *P* will occur in situation *S*, given reinforcement *R*, is a function of *P*'s expectation that reinforcement *R* will occur after *P*'s behavior *B* in situation *S*, and of the value *V* to *P* of the reinforcement *R* in situation *S*.

For example, while in college, Pat has tried to lose weight through diet and exercise many times in the past, and he has always been unsuccessful. Therefore, he has developed a low generalized expectancy of success resulting from his memory of and reflection on years of specific expectancy-behavior-outcome sequences. Based on past experience, Pat has a fairly stable estimate of the probability that certain behaviors will lead to the goal of losing weight. In addition, Pat has developed some beliefs as to why his weight loss efforts have been unsuccessful for so long. Perceived locus of control, then, is Pat's abstraction of why weight loss has been unsuccessful all those times—a generalized expectancy of internal (e.g., "I have no willpower when it comes to food") or external (e.g., "My busy class and work schedules prevent me from losing weight"). The above example used Pat's failure as an illustration, but Pat's success could also be used as an example. It is important to note that both successes and failures may be related to either internal or external loci of control. So even with successful weight loss efforts, Pat may have external perceived control ("My family's support, my doctor's instructions, and a gym on campus will make it very easy for me to lose weight"). Therefore, perceived locus of control focuses on how the individual perceives self in relationship to things that happen to him or her and the meaning that the self makes of those experiences.

In many ways, this is very similar to constructs in attribution theory found in social psychology, which has also been applied to interventions to change health behaviors. Attribution theory provides a framework to explain the process that people use information to make inferences about the causes of behaviors or events. For example, we might ask ourselves why Matt can't stop smoking. Our reasons or attributions may be, similar to locus of control theory, internal (Matt does not have much willpower) or external (All Matt's friends smoke, so it is hard for him to quit with them around). There are other dimensions of attributions as well (e.g., temporary vs. permanent aspects).

Measuring Locus of Control

Lefcourt strongly suggests that researchers and practitioners tailor or target the locus of control measure to their populations and their specific domains of interest rather than depending on the more global measures that are quite probably irrelevant to the people and behavior of interest. This recommendation is very similar to findings from the attitude-behavior consistency literature that suggests that if

we want to predict a particular behavior (e.g., jogging behavior) from attitudes, then we should assess specific attitudes toward that specific behavior (attitudes toward jogging) instead of more global attitudes (e.g., attitudes toward health).

The first measures of perceived control were created as part of dissertations in the 1950s at Ohio State University by Phares and James. Phares used a Likert-type scale and found that participants with more external attitudes behaved in a fashion similar to participants who received "chance" instructions for a task. James used a longer scale, referred to as the James-Phares Scale, based on what seemed to be the most useful items from Phares's scale. James also found modest correlations between his measure of locus of control and his participants' responses to failure and success. The items from these scales provided the basis for the well-known and often used Rotter Internal-External Control Scale developed by Rotter. The Rotter scale used forced-choice items (e.g., participants had to choose the internal response or the external response) and included items from many different domains (e.g., war, personal respect, school grades and examination performance, leadership, being liked and respected, personality, getting a good job), as well as filler items. It became clear that expectancies of internal-external control were assessable with paper-and-pencil measures, although Lefcourt, who developed such a measure, warns against strict interpretation of locus of control as a trait or typology based on scales.

Consistent with the early writings of Rotter and Lefcourt, many social science researchers have attempted to develop and tailor perceived control scales to their populations and domains of interest, including many for health-related domains. Interestingly, many of the measures are multidimensional. Factor analyses often suggest several dimensions that are not highly intercorrelated. There seems to be an increasing acceptance that locus of control is a multidimensional construct and that internal and external are not just opposite ends of a single bipolar dimension. These results suggest that perceived control is more complex than early researchers thought, since scholars have found that respondents may score high on both internal and external dimensions (e.g., highly religious or spiritual persons may believe that weight loss will be due both to God and their own exercise and dieting efforts), high on one and low on the other, or even low on both internal and external dimensions (perhaps indicating that the domain is not relevant to that individual). Furthermore, there is also some recognition that there may be multiple internal and multiple external dimensions. Even early scholars realized, for example, that luck, chance, God, and powerful others may reflect different dimensions of external perceived control. For example, someone with a high external score in the domain of weight loss may attribute his or her successes or failures mostly to God, and less to luck, chance, and powerful others (e.g., their physician), while those with high internal scores on weight loss may attribute their successes or failures to ability, motivation, or effort. These findings suggest that other dimensions must be considered along with perceived control. For example, stability over time may be an important part of perceived control, with ability reflecting internal perceived control and fixed stability, and effort reflecting internal control and variable stability (e.g., Pat may put more effort into losing weight in January, right after overeating during the holidays, than in September when classes are starting). Similarly, external control and fixed stability might characterize task difficulty, and external control and variable stability may characterize luck. Some of the scales developed to measure locus of control include the Crandall, Katkovsky, and Crandall Intellectual Achievement Responsibility Questionnaire, which assesses children's beliefs about their control and responsibility for failure and success in intellectual achievement; the Nowicki-Strickland Internal-External Control Scale for Children, which assesses generalized expectancies among children, with items from a variety of domains, such as catching a cold, getting good grades, being punished, being good at sports, choosing friends; the Lefcourt, von Bayer, Ware, and Cox Multidimensional-Multiattributional Causality Scale, which assesses perceived control in areas of affiliation and achievement for older/university students; the Miller, Lefcourt, and Ware Marital Locus of Control Scale, which assesses perceived control of marital satisfaction; and the Campis, Lyman, and Prentice-Dunn Parenting Locus of Control Scale, which assesses perceived control beliefs regarding parents' perspective of child rearing successes and failures.

Measuring Health-Related Perceived Control

Health-related control scales are increasing in popularity. Among the scales that measure control in health-related areas are the Wallston, Wallston, and

DeVellis Multidimensional Health Locus of Control Scale, which assesses control beliefs relevant to health; the Hill and Bale Mental Health Locus of Control Scale, which assesses perceived control of therapeutic changes by patient/internal or therapist/external; the Keyson and Janda Drinking Locus of Control Scale, which assesses control expectancies for drinking-related behaviors; the Saltzer Weight Locus of Control Scale, which assesses internal and external control beliefs regarding determinants of weight; and the Catania, McDermott, and Wood Dyadic Sexual Regulation Scale, which assesses control beliefs relevant to sexual activity. Recently, Holt and colleagues have developed and revised their Spiritual Health Locus of Control Scale, which assesses African Americans' perceived control beliefs regarding health, including several dimensions regarding the influence of God, as well as internal control beliefs.

Lefcourt has suggested that higher perceived control is associated with better access to opportunities, so that those who are able to more readily reach valued outcomes that allow a person to feel satisfaction are more likely to hold internal control expectancies (i.e., an internal locus of control). He suggests that, compared with whites in the United States, members of ethnic minority groups who do not have such access are more likely to hold external control and fatalistic beliefs. However, Banks, Ward, McQuater, and DeBritto (1991) have warned against making such a sweeping generalization about ethnic groups, in particular African Americans. They have reviewed some of the research and noted some of the methodological and theoretical weaknesses of the original construct and method. Some researchers have also suggested that the domains studied are usually those in which individuals have at least some control; rarely, if ever, are domains studied in which the individual has no actual control. These domains may be ones in which it is not beneficial to have internal perceived control. Not inconsistently with these ideas, both Lefcourt and Rotter have suggested that perceived control should have a curvilinear relationship with measures of psychological adjustment.

Another measurement issue is one of construct validity. Researchers sometimes confuse constructs in developing measures of perceived locus of control, self-efficacy, perceived behavioral control, and comparable variables. For example, Albert Bandura, who was instrumental in the development of social learning theory, has argued strongly that the construct of perceived locus of control is very different from self-efficacy, even though both originate from social learning theory. However, it is not clear that all measures of perceived control (or efficacy, for that matter) reflect these theoretical differences. More care is needed in developing and validating locus of control scales.

Lefcourt recommends that researchers place more emphasis on a neglected aspect of perceived control: the value or importance of the domain and the reinforcement. According to social learning theory, the value of the reinforcement, not just the expectancy, is an important part of the process, yet few researchers assess this variable or incorporate it into their interpretation of their results. Some researchers have suggested that locus of control best predicts performance in areas that are highly valued, and both Lefcourt and Rotter emphasize the importance of embedding perceived control into the larger social learning theory when designing research and interpreting research findings.

Interventions: Changing Perceived Control

Lefcourt suggests that perceived locus of control can be changed, at least as assessed by current scales, and at least for short periods of time. He notes that there is less evidence for long-term change of perceived control. In fact, change in locus of control may be an important goal of physicians, physical therapists, and other health and mental health professionals. Lefcourt suggests that people can change their attributions if they have experiences that change the perception of contingencies between their behavior and the perceived outcomes. However, another approach is not to attempt to change someone's locus of control but to develop health education materials and other interventions that are tailored to the individual's current locus of control and thus are perceived as more relevant by the participants. Such efforts should take into account the fact that perceived control may change if the individual is more likely to use the health information. (For example, if a woman with an external locus of control regarding her health is induced by health education materials tailored to obtain a mammogram, this may influence her to change to a more internal locus of control, at least for that particular health behavior.) Research that has tested the possibility of changing an individual's locus of control in this

manner has been performed, but with somewhat related construct of self-efficacy. Like many dispositional variables and behaviors, accomplishing long-term change in locus of control will probably require more than a one-time intervention and will need to include follow-up sessions aimed at maintaining the change over time and preventing relapse.

Locus of Control and Health Behaviors

Catherine Sanderson suggests that locus of control and other personality/dispositional variables may influence health behaviors but that this influence may depend on other mediating variables. For example, locus of control may influence the perceptions of a stressor, the coping strategies used, how well the individual gathers social support, and the amount of physiological reaction the individual has to the situation. These variables, in turn, then influence the health behavior of interest. The study of possible mediators is providing new information in answering the how and why questions regarding the role of perceived control in related health behavior.

—Eddie M. Clark

See also Health Behavior; Health Communication; Self-Efficacy

Further Readings

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

Banks, W. C., Ward, W. E., McQuater, G. V., & DeBritto, A. M. (1991). Are blacks external: On the status of locus of control in Black populations. In R. L. Jones (Ed.), *Black psychology* (3rd ed., pp. 181–192). Berkeley, CA: Cobb & Henry.

Holt, C. L., Clark, E. M., & Klem, P. R. (2007). Expansion and validation of the spiritual locus of control scale: Factor analysis and predictive validity. *Journal of Health Psychology*, *12*(4), 597–612.

Holt, C. L., Clark, E. M., Kreuter, M. W., & Rubio, D. M. (2003). Spiritual health locus of control and breast cancer beliefs among urban African American women. *Health Psychology*, *22*, 294–299.

Lefcourt, H. M. (Ed.). (1981). *Research with the locus of control construct: Vol. 1. Assessment methods*. New York: Academic Press.

Lefcourt, H. M. (1982). *Locus of control: Current trends in theory and research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Lefcourt, H. M. (1991). Locus of control. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 413–499). San Diego, CA: Academic Press.

Lefcourt, H. H., von Bayer, C. L., Ware, E. E., & Cox, D. V. (1979). The multidimensional-multiattributional causality scale: The development of a goal specific locus of control. *Canadian Journal of Behavioural Science*, *11*, 286–304.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, *80*(1, Whole No. 609).

Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist*, *45*, 489–493.

Saltzer, E. B. (1982). The weight locus of control (WLOC) scale: A specific measure for obesity research. *Journal of Personality Assessment*, *46*, 620–628.

Sanderson, C. A. (2004). *Health psychology.* Hoboken, NJ: Wiley.

Wallston, K. A., Wallston, B. S., & DeVellis, R. (1978). Development of the multidimensional health locus of control scales. *Health Education Monographs*, *6*, 161–170.

Log-Binomial Regression

See Regression

Logistic Regression

Logistic regression is a statistical technique for analyzing the relationship of an outcome or dependent variable to one or more predictors or independent variables when the dependent variable is (1) *dichotomous*, having only two categories, for example, the presence or absence of symptoms, or the use or nonuse of tobacco; (2) *unordered polytomous*, a nominal scale variable with three or more categories, for example, type of contraception (none, pill, condom, intrauterine device) used in response to services provided by a family planning clinic; or (3) *ordered polytomous*, an ordinal scale variable with three or more categories, for example, whether a patient's condition deteriorates, remains the same, or improves in response to a cancer treatment. Here, the basic logistic regression model for dichotomous outcomes is examined, noting its extension to polytomous outcomes and its conceptual roots in both log-linear analysis and the general linear model. Next, consideration is given to methods for assessing the goodness of fit and predictive utility of the overall model, and calculation

and interpretation of logistic regression coefficients and associated inferential statistics to evaluate the importance of individual predictors in the model. Throughout, the discussion assumes an interest in prediction, regardless of whether causality is implied; hence the language of "outcomes" and "predictors" is preferred to the language of "dependent" and "independent" variables.

The equation for the logistic regression model with a dichotomous outcome is $\text{logit}(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$, where Y is the dichotomous outcome; logit(Y) is the natural logarithm of the odds of Y, a transformation of Y to be discussed in more detail momentarily; and there are $k = 1, 2, \ldots K$ predictors X_k with associated coefficients β_k, plus a constant or intercept, α, which represents the value of logit(Y) when all the X_k are equal to 0. If the two categories of the outcome are coded 1 and 0, respectively, and P_1 is the probability of being in the category coded as 1, and P_0 is the probability of being in the category coded as 0, then the odds of being in Category 1 is $P_1/P_0 = P_1/(1 - P_1)$ (since the probability of being in one category is one minus the probability of being in the other category). Logit(Y) is the *natural logarithm of the odds*, $\ln[P_1/(1 - P_1)]$, where ln represents the natural logarithm transformation.

Polytomous Logistic Regression Models

When the outcome is polytomous, logistic regression can be implemented by splitting the outcome into a set of dichotomous variables. This is done by means of contrasts that identify a reference category (or set of categories) with which to compare each of the other categories (or sets of categories). For a nominal outcome, the most commonly used model is called the *baseline category logit model.* In this model, the outcome is divided into a set of dummy variables, each representing one of the categories of the outcome, with one of the categories designated as the reference category, in the same way that dummy coding is used for nominal predictors in linear regression. If there are M categories in the outcome, then $\text{logit}(Y_m) = \ln(P_m/P_0) = \alpha_m + \beta_1 m X_1 + \beta_2 m X_2 + \cdots + \beta_K m X_K$, where P_0 is the probability of being in the reference category and P_m is the probability of being in Category $m = 1,2, \ldots, M - 1$, given that the case is either in Category m or in the reference category. A total of $(M - 1)$ equations or logit functions is thus estimated, each with its own intercept α_m and logistic regression coefficients $\beta_{k,m}$, representing the relationship of the predictors to logit(Y_m).

For ordinal outcomes, the situation is more complex, and there are several different contrasts that may be used. In the *adjacent category logit* model, for example, each category is contrasted only with the single category preceding it. In the *cumulative logit* model, for the first logit function, the first category is contrasted with all the categories following it; then, for the second logit function, the first *two* categories are contrasted with all the categories following them, and so forth; until for the last $(M - 1)$ logit function, all the categories preceding the last are contrasted with the last category. Other contrasts are also possible. The cumulative logit model is the model most commonly used in logistic regression analysis for an ordinal outcome, and has the advantage over other contrasts that splitting or combining categories (representing more precise or cruder ordinal measurement) should not affect estimates for categories other than the categories that are actually split or combined, a property not characteristic of other ordinal contrasts. It is commonly assumed in ordinal logistic regression that only the intercepts (or *thresholds*, which are similar to intercepts) differ across the logit functions. The ordinal logistic regression equation can be written (here in the format using intercepts instead of thresholds) as $\text{logit}(Y_m) = \alpha_m + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$, where $\alpha_m = \alpha_1, \alpha_2, \ldots, \alpha_{M-1}$ are the intercepts associated with the $(M - 1)$ logit functions, but $\beta_1, \beta_2, \ldots, \beta_K$ are assumed to be identical for the $(M - 1)$ logit functions, an assumption that can be tested and, if necessary, modified.

Logistic Regression, Log-linear Analysis, and the General Linear Model

Logistic regression can be derived from two different sources, the general linear model for linear regression and the logit model in log-linear analysis. Linear regression is used to analyze the relationship of an outcome to one or more predictors when the outcome is a continuous interval or ratio scale variable. Linear regression is extensively used in the analysis of outcomes with a *natural metric*, such as kilograms, dollars, or numbers of people, where the unit of measurement is such that it makes sense to talk about larger or smaller differences between cases

(the difference between the populations of France and Germany is smaller than the difference between the populations of France and China), and (usually) it also makes sense to talk about one value being some number of times larger than another ($10,000 is twice as much as $5,000), comparisons that are not applicable to the categorical outcome variables for which logistic regression is used. The equation for linear regression is $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$, and the only difference from the logistic regression equation is that the outcome in linear regression is Y instead of logit(Y). The coefficients β_k and intercept α in linear regression are most commonly estimated using ordinary least squares (OLS) estimation, although other methods of estimation are possible.

For OLS estimation and for statistical inferences about the coefficients, certain assumptions are required, and if the outcome is a dichotomy (or a polytomous variable represented as a set of dichotomies) instead of a continuous interval/ratio variable, several of these assumptions are violated. For a dichotomous outcome, the predicted values may lie outside the range of possible values (suggesting probabilities >1 or <0), especially when there are continuous interval or ratio scale predictors in the model; and inferential statistics are typically incorrect because of *heteroscedasticity* (unequal residual variances for different values of the predictors) and nonnormal distribution of the residuals. It is also assumed that the relationship between the outcome and the predictors is linear; however, in the *general linear model*, it is often possible to linearize a nonlinear relationship by using an appropriate nonlinear transformation. For example, in research on income (measured in dollars), it is commonplace to use the natural logarithm of income as an outcome, because the relationship of income to its predictors tends to be nonlinear (specifically, logarithmic). In this context, the logit transformation is just one of many possible linearizing transformations.

An alternative to the use of linear regression to analyze dichotomous and polytomous categorical outcomes is logit analysis, a special case of log-linear analysis. In log-linear analysis, it is assumed that the variables are categorical, and can be represented by a *contingency table* with as many dimensions as there are variables, with each case located in one cell of the table, corresponding to the combination of values it has on all the variables. In log-linear analysis, no distinction is made between outcomes and predictors, but in logit analysis, one variable is designated as the outcome and the other variables are treated as predictors, and each unique combination of values of the predictors represents a covariate pattern. Logit model equations are typically presented in a format different from that used in linear regression and logistic regression, and log-linear and logit models are commonly estimated using *iterative maximum likelihood (ML) estimation*, in which one begins with a set of initial values for the coefficients in the model, examines the differences between observed and predicted values produced by the model (or some similar criterion), and uses an algorithm to adjust the estimates to improve the model. This process of estimation and adjustment of estimates is repeated in a series of steps (*iterations*) that end when, to some predetermined degree of precision, there is no change in the fit of the model, the coefficients in the model, or some similar criterion.

Logistic regression can be seen either as a special case of the general linear model involving the logit transformation of the outcome or as an extension of the logit model to incorporate continuous as well as categorical predictors. The basic form of the logistic regression equation is the same as for the linear regression equation, but the outcome, logit(Y), has the same form as the outcome in logit analysis. The use of the logit transformation ensures that predicted values cannot exceed observed values (for an individual case, the logit of Y is either positive or negative infinity, $+\infty$ or $-\infty$), but it also makes it impossible to estimate the coefficients in the logistic regression equation using OLS. Estimation for logistic regression, as for logit analysis, requires an iterative technique, most often ML, but other possibilities include iteratively reweighted least squares, with roots in the general linear model, or some form of quasi-likelihood or partial likelihood estimation, which may be employed when data are clustered or nonindependent. Common instances of nonindependent data include multilevel analysis, complex sampling designs (e.g., multistage cluster sampling), and designs involving repeated measurement of the same subjects or cases, as in longitudinal research. Conditional logistic regression is a technique for analyzing related samples, for example, in matched case-control studies, in which, with some minor adjustments, the model can be estimated using ML.

Assumptions of Logistic Regression

Logistic regression assumes that the functional form of the equation is correct, and hence the predictors X_k are linearly and additively related to logit(Y), but variables can be transformed to adjust for nonadditivity and nonlinearity (e.g., nonlinearly transformed predictors or interaction terms). It also assumes that each case is independent of all the other cases in the sample, or when cases are not independent, adjustments can be made in either the estimation procedure or the calculation of standard errors (or both) to adjust for the nonindependence. Like linear regression, logistic regression assumes that the variables are measured without error, that all relevant predictors are included in the analysis (otherwise the logistic regression coefficients may be biased), and that no irrelevant predictors are included in the analysis (otherwise standard errors of the logistic regression coefficients may be inflated). Also, as in linear regression, no predictor may be perfectly collinear with one or more of the other predictors in the model. Perfect collinearity means that a predictor is completely determined by or predictable from one or more other predictors, and when perfect collinearity exists, there exist an infinite number of solutions that maximize the likelihood in ML estimation, or minimize errors of prediction more generally. Logistic regression also assumes that the errors in prediction have a binomial distribution, but when the number of cases is large, the binomial distribution approximates the normal distribution. Various diagnostic statistics have been developed and are readily available in existing software to detect violations of assumptions and other problems (e.g., outliers and influential cases) in logistic regression.

Goodness of Fit and Accuracy of Prediction

In logistic regression using ML (currently the most commonly used method of estimation), in place of the sum of squares statistics used in linear regression, there are log-likelihood statistics, calculated based on observed and predicted probabilities of being in the respective categories of the outcome variable. When multiplied by -2, the difference between two log-likelihood statistics has an approximate chi-square distribution for sufficiently large samples involving independent observations. One can construct -2 log-likelihood statistics (here and elsewhere designated as D) for (1) a model with no predictors, D_0, and (2) the tested model, the model for which the coefficients are actually estimated, D_M. D_M, sometimes called the *deviance statistic*, has been used as a goodness-of-fit statistic, but has somewhat fallen out of favor because of concerns with alternative possible definitions for the saturated model (depending on whether individual cases or covariate patterns are treated as the units of analysis), and the concern that, for data in which there are few cases per covariate pattern, D_M does not really have a chi-square distribution. The Hosmer-Lemeshow goodness-of-fit index is constructed by grouping the data, typically into deciles, based on predicted values of the outcome, a technique applicable even with few cases per covariate pattern. There appears to be a trend away from concern with goodness of fit, however, to focus instead on the model chi-square statistic, $G_M = D_0 - D_M$, which compares the tested model with the model with no predictors. G_M generally does follow a chi-square distribution in large samples, and is analogous to the multivariate F statistic in linear regression and analysis of variance. G_M provides a test of the statistical significance of the overall model in predicting the outcome. An alternative to G_M for models not estimated using ML is the multivariate Wald statistic.

There is a substantial literature on coefficients of determination for logistic regression in which the goal is to find a measure analogous to R^2 in linear regression. When the concern is with how close the predicted probabilities of category membership are to observed category membership (quantitative prediction), two promising options are the likelihood ratio R^2 statistic, $R^2_L = G_M/D_0$ applicable specifically when ML estimation is used, and the OLS R^2 statistic itself, calculated by squaring the correlation between observed values (coded 0 and 1) and the predicted probabilities of being in Category 1. Advantages of R^2_L are that it is based on the quantity actually being maximized in ML estimation, it appears to be uncorrelated with the *base rate* (the percentage of cases in Category 1), and it can be calculated for polytomous as well as dichotomous outcomes. Other R^2 analogues have been proposed, but they have various problems that include correlation with the base rate (to the extent that the base rate itself appears to determine the calculated accuracy of prediction), having no reasonable value for perfect prediction or for perfectly incorrect prediction, or being limited to dichotomous outcomes.

Alternatively, instead of being concerned with predicted probabilities, one may be concerned with how accurately cases are qualitatively classified into the categories of the outcome by the predictors (qualitative prediction). For this purpose, there is a family of indices of predictive efficiency, designated lambda-*p*, tau-*p*, and phi-*p*, that are specifically applicable to qualitative prediction, classification, and selection tables (regardless of whether they were generated by logistic regression or some other technique), as opposed to contingency tables more generally. Finally, none of the aforementioned indices of predictive efficiency (or R^2 analogues) takes into account the ordering in an ordered polytomous outcome, for which one would naturally consider ordinal measures of association. Kendall's tau-*b* is an ordinal measure of association which, when squared (τ_b^2), has a proportional reduction in error interpretation and seems most promising for use with ordinal outcomes in logistic regression. Tests of statistical significance can be computed for all these coefficients of determination.

Unstandardized and Standardized Logistic Regression Coefficients

Interpretation of unstandardized logistic regression coefficients (b_k, the estimated value of β_k) is straightforward and parallel to the interpretation of unstandardized coefficients in linear regression: a one-unit increase in X_k is associated with a b_k increase in logit(Y) (not in Y itself). If we raise the base of the natural logarithm, $e = 2.718\ldots$, to the power b_k, we obtain the odds ratio, here designated ω_k, which is sometimes presented in place of or in addition to b_k, and can be interpreted as indicating that a one-unit increase in X_k multiplies the odds of being in Category 1 by ω_k. Both b_k and ω_k convey exactly the same information, just in a different form. There are several possible tests of statistical significance for unstandardized logistic regression coefficients. The univariate Wald statistic can be calculated either as the ratio of the logistic regression coefficient to its standard error, $b_k/SE,(b_k)$, which has an approximate normal distribution, or $[b_k/SE(b_k)]^2$, which has an approximate chi-square distribution. The Wald statistic, however, tends to be inflated for large b_k, tending to fail to reject the null hypothesis when the null hypothesis is false (Type II error), but it may still be the best available option when ML is not used to estimate the model. Alternatives include the Score statistic and the likelihood ratio statistic (the latter being the difference in D_M with and without X_k in the equation). When ML estimation is used, the likelihood ratio statistic, which has a chi-square distribution and applies to both b_k and ω_k, is generally the preferred test of statistical significance for b_k and ω_k.

Unless all predictors are measured in exactly the same units, neither b_k nor ω_k clearly indicates whether one variable has a stronger impact on the outcome than another. Likewise, the statistical significance of b_k or ω_k tells us only how sure we are that a relationship exists, not how strong the relationship is. In linear regression, to compare the *substantive significance* (strength of relationship, which does not necessarily correspond to statistical significance) of predictors measured in different units, we often rely on standardized regression coefficients. In logistic regression, there are several alternatives for obtaining something like a standardized coefficient. A relatively quick and easy option is simply to standardize the predictors (standardizing the outcome does not matter, since it is the probability of being in a particular category of Y, not the actual value of Y, that is predicted in logistic regression). A slightly more complicated approach is to calculate $b_k^* = (b_k)(s_x)(R)/s_{logit(Y)}$, where b_k^* is the fully standardized logistic regression coefficient, b_k is the unstandardized logistic regression coefficient, s_x is the standard deviation of the predictor X_k, R is the correlation between the observed value of Y and the predicted probability of being in Category 1 of Y, $s_{logit(Y)}$ is the standard deviation of the predicted values of logit(Y), and the quantity $s_{logit(Y)}/R$ represents the estimated standard deviation in the observed values of logit(Y) (which must be estimated, since the observed values are positive or negative infinity for any single case). The advantage of this fully standardized logistic regression coefficient is that it behaves more like the standardized coefficient in linear regression, including showing promise for use in path analysis with logistic regression, a technique under development as this is being written.

Logistic Regression and Its Alternatives

Alternatives to logistic regression include probit analysis, discriminant analysis, and models practically identical to the logistic regression model but with different distributional assumptions (e.g., complementary

log-log or extreme value instead of logit). Logistic regression, however, has increasingly become the method most often used in empirical research. Its broad applicability to different types of categorical outcomes and the ease with which it can be implemented in statistical software algorithms, plus its apparent consistency with realistic assumptions about real-world empirical data, have led to the widespread use of logistic regression in the biomedical, behavioral, and social sciences.

—Scott Menard

See also Chi-Square Test; Collinearity; Discriminant Analysis; Likelihood Ratio; Regression

Further Readings

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: MIT Press.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, *54*, 17–24.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Menard, S. (2004). Six approaches to calculating standardized logistic regression coefficients. *The American Statistician*, *58*, 218–223.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, *9*, 705–724.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Simonoff, J. S. (1998). Logistic regression, categorical predictors, and goodness of fit: It depends on who you ask. *The American Statistician*, *52*, 10–14.

Log-Rank Test

The log-rank test is a statistical method to compare two survival distributions—that is, to determine whether two samples may have arisen from two identical survivor functions. The log-rank test is easy to compute and has a simple heuristic justification and is therefore often advocated for use to nonstatisticians. The log-rank test can also be thought as a censored data rank test.

Suppose we obtain two samples from two populations and we are interested in the null hypothesis

$$H_0 : S_1 = S_2$$

that the survival distribution from Sample 1 is identical to the survival distribution in Sample 2.

The idea behind the log-rank test is to compare the observed number of deaths at each failure time with the expected number of deaths under the null hypothesis (i.e., assuming that the null hypothesis is true).

To do this, we consider the ordered failure times for the combined samples $t_1 \leq t_2 \leq \cdots t_j \cdots \leq t_k$. We then divide the observation period into small intervals ($I_1 = [t_0, t_1)$ $(M-1)$ $I_2 = [t_1, t_2), \ldots, I_j = [t_j - 1, t_j), \ldots, I_K = [t_K - 1, t_K)$), each one corresponding to the survival time of the noncensored individuals. For each interval I_j $(j = 1, K)$,

d_j is the number of individuals who die at t_j and

r_j is the number of individuals who are alive and at risk just before t_j.

For each table, the quantities d_{1j}/r_{1j} and d_{2j}/r_{2j} are hazard estimates.

To perform the log-rank test, we construct a 2×2 table at each of the failure times t_j.

	Dead	*Alive*	*Total*
Sample 1	d_{1j}	$r_{1j} - d_{1j}$	r_{1j}
Sample 2	d_{2j}	$r_{2j} - d_{2j}$	r_{1j}
Total	d_{j}	$r_{\mathrm{j}} - d_{\mathrm{j}}$	r_{j}

From this table, define

$$O_j = d_{1j},$$
$$E_j = \tfrac{d_{1j} r_{1j}}{r_j}, \text{ and}$$
$$V_j = \mathrm{Var}(E_j) = \frac{d_j (r_j - d_j) r_{1j} r_{2j}}{r_j^2 (r_j - 1)},$$

and calculate

$O_\bullet = \sum_j O_j =$ Total number of deaths in Sample 1,

$E_\bullet = \sum_j E_j =$ Total number of expected deaths in Sample 1 under the null hypothesis of no difference between the two survival distributions, and

$V_\bullet = \sum_j V_j =$ Variance term for the failures in Sample 1.

The log-rank test is given by

$$T_{L-R} = \frac{(O_\bullet - E_\bullet)^2}{V_\bullet}.$$

This test statistic follows a chi-square distribution with 1 degree of freedom under the null hypothesis (although the tables are not really independent, the distributional result still holds).

Large values of the test statistic indicate that the observed distribution of deaths in Sample 1 diverges from the expected number of deaths if the two survival distributions were identical. Although different censoring patterns do not invalidate the log-rank test, the test can be sensitive to extreme observations in the right tail of the distribution.

The log-rank test is particularly recommended when the ratio of hazard functions in the population being compared is approximately constant. For small data set, the log-rank test can be easily calculated by hand. Most statistical software contains routines for the calculation of the log-rank test.

—Emilia Bagiella

See also Hypothesis Testing; Kaplan-Meier Method; Survival Analysis

Further Readings

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, *50*, 163–170.

Savage, I. R. (1956). Contribution to the theory of rank order statistics: The two sample case. *Annals of Mathematical Statistics*, *27*, 590–615.

Longitudinal Research Design

In epidemiology, a longitudinal study refers to the collection of data from the same unit (e.g., the same person) at two or more different points in time. The great advantage of longitudinal studies is that each subject serves as his or her own control in the study of change across time. This reduces between-subject variability and requires a smaller number of subjects compared with independent subject designs, and it allows the researcher to eliminate a number of competing explanations for effects observed—most important, the cohort effect. The main disadvantages of longitudinal designs are that they are expensive and time-consuming relative to cross-sectional designs, and that they are subject to difficulties with retention, that is, subjects may drop out of the study. In addition, special statistical techniques are needed to account for the fact that repeated measurements taken on the same person or unit will be more similar than the same number of measurements taken on different people.

Designing a longitudinal study is a complex task that involves a number of decisions. The primary decision is whether the data should be recorded prospectively (from the starting point of the study into the future) or retrospectively (collecting data on events that have already occurred). The investigator must also determine how to select a sample of subjects that will represent the target population and how large a sample is needed to have adequate power. Finally, the investigator must choose the variables that will represent the phenomenon under investigation and the frequency at which these variables should be measured.

Methods for the analysis of data in longitudinal studies depend primarily on whether time is considered as a covariate or as an outcome. When the time is viewed as a covariate, regression techniques that account for within-subject association in the data can be used to study the change across time. For time-to-event data, survival analysis that takes into account the potential censoring of data, that is, the unavailability of end points, is required. This entry is concerned with studies in which time is considered to be a covariate. Survival analysis and related methods, which consider time-to-event as the outcome, are treated in a separate entry.

Types of Longitudinal Studies

Longitudinal studies allow the separation of the cohort effect (e.g., the effect of being born in 1956 vs. being born in 1966) from the time effect (e.g., the change in risk behavior for someone at age 20 vs. age 30). They are more difficult and time-consuming to perform than cross-sectional studies, but they allow the investigator to make more convincing conclusions about cause and effect. In addition, longitudinal designs generally need fewer subjects than cross-sectional designs, and the fact that the same subjects are studied repeatedly reduces

the variability due to subjects and increases the study's power.

There are several variations of longitudinal studies, but the most common ones are prospective, retrospective, and nested case-control designs. In a prospective study, the investigator plans a study ahead of time, deciding what data to gather, and then records pertinent information on the exposures, outcomes, and potential confounders. The main advantage of this design is that the researcher can collect data that are needed to answer a particular research question, as opposed to simply gleaning whatever existing data are available from other sources. In a retrospective study, the events that are being studied occurred in the past, and the researcher is studying data gleaned from existing sources, such as hospital records. This type of study is feasible only if adequate data about the risk factors, potential confounders, and main outcomes are available on a cohort that has been assembled for other reasons. The main advantage of a retrospective design is that it is possible to gather data in a relatively short period, and these designs are particularly useful in studying rare diseases.

The third type of longitudinal study is the nested case-control design. As the name suggests, nested case-control designs have a case-control study nested within a prospective or retrospective study. They are most useful for predictor variables that are costly to collect, such as the analysis of human specimens. One good example of a nested case-control study is the Baltimore Longitudinal Study on Aging, an ongoing study that started in 1958. The primary objective of this prospective study is to study the process of normal human aging. Participants in the study are volunteers who return approximately every 2 years for 3 days of biomedical and psychological examinations. As reported by Geert Verbeke and Geert Molenberghs, this study is a unique resource for rapidly evaluating longitudinal hypotheses because of the availability of data from repeated examinations and a bank of frozen blood samples from the same participants above 35 years of follow-up. For example, to study the natural history of prostate cancer, subjects with the disease (cases) are matched with their peers who are disease free (controls) and compared according to their previously recorded prostate-specific antigen profiles across time. The disadvantage of this type of study is that development of prostate cancer may have been influenced by many covariates that were not recorded in the study and that, therefore, cannot be examined as risk factors; the advantage is that the research begins with existing cases of prostate cancer rather than beginning with a large number of healthy subjects and observing them over many years to see which will become ill.

Observation Time Points, Duration, Sample Size, and Power of the Study

The type of design the investigator chooses for the study and the nature of the outcome have a major impact on the observation time points (time points at which data will be collected) and the time lag between these time points, sample size, and power of the study. All power calculations involve considerations of effect size, effect variability, and sample size; however, in a longitudinal design, two other factors are involved—the number of time points at which data will be collected and the time lag between these time points. Typically, these observation time points as well as the time lags between them are preselected by the researcher based on the etiology of the phenomenon under investigation. Generally, one of the factors is computed, usually through simulations, assuming that the remaining factors are prespecified by the investigator. As an example, to compute the sample size necessary for a given statistical power, the investigator has to specify the smallest worthwhile effect size representing the smallest effect that would make a difference to the interpretation of the research question, the number of observation time points, and the time lags between these points and the variance. The smallest worthwhile effect size and the variance are typically estimated from a pilot study, or previously conducted studies with similar objectives. Alternatively, the sample size, variance, effect size, number of design time points, and time lags can be fixed for acceptable power to be calculated.

Statistical Analysis in Longitudinal Studies

When the time variable is viewed as a covariate, regression models known as growth curve models are typically used to summarize longitudinal profiles of the dependent variable under investigation. These models include the popular linear mixed models, generalized linear mixed models, and generalized estimating equation (GEE) models.

Modeling Continuous Outcomes

Linear mixed-effects models are a useful tool to analyze normal continuous repeated measurements recorded on the same subject. They are likelihood-based models for which the conditional expectations (given random effects) are made of two components, a fixed-effects term and a random-effects term. The fixed-effects term represents the average effects of time-dependent covariates, such as the time itself and the effects of time-independent covariates, that is, those whose values may not change during the course of the study, such as baseline covariates. The random effects represent a deviation of a subject's profile from the average profile, and they account for the within-subject correlation across time. These random effects adjust for the heterogeneity of subjects, which can be viewed as unmeasured predispositions or characteristics. These models have an appealing feature in that the fixed-effects parameters have both a subject-specific interpretation and a population-averaged one. Specifically, the effects of time-dependent covariates are interpreted using the conditional expectations given random effects, whereas that of time-independent covariates are conducted using the marginal mean (unconditional of random effects). Before the advent of linear mixed models, longitudinal data were analyzed using techniques, such as repeated measures analysis of variance (ANOVA). This approach has a number of disadvantages and has generally been superseded by linear mixed modeling, which is now available in commonly used statistical packages. For example, repeated measures ANOVA models require a balanced design in that measurements should be recorded at the same time points for all subjects, a condition not required by linear mixed models.

Modeling Discrete and Categorical Outcomes

Although there are a variety of standard likelihood-based models available to analyze data when the outcome is approximately normal, models for discrete outcomes (such as binary outcomes) generally require a different methodology. Kung-Yee Liang and Scott Zeger have proposed the so-called generalized estimating equations model, which is an extension of generalized linear model to correlated data. The basic idea of this family of models is to specify a function that links the linear predictor to the mean response and use the so-called sandwich estimator to adjust the standard errors for association in the data. As a result, the within-subjects association is not modeled explicitly, but treated as a nuisance parameter. GEE regression parameter estimates have a population-averaged interpretation analogous to those obtained from a cross-sectional data analysis. A well-known alternative to this class of models is the generalized linear mixed models, which explicitly model the association in the data using random effects. These models are also likelihood based and are typically formulated as hierarchical models. At the first stage, the conditional distribution of the data given random effects is specified, usually assumed to be a member of the exponential family. At the second stage, a prior distribution is imposed on the random effects. One of the drawbacks of these models is that the fixed-effects parameters, with the exception of few link functions, have a subject-specific interpretation in that they give the magnitude of change occurring within an individual profile. To assess changes between subjects, the investigator is then required to integrate out the random effects from the quantities of interest. There exist other likelihood methods for analyzing discrete data for which parameters have a population-averaged interpretation. One good example is the multivariate Probit model for binary and ordered data that uses the Pearson's correlation coefficient to capture the association between time point responses. Another alternative is the multivariate Plackett distribution that uses odds ratios for association. One drawback of these marginal models is that they require the time points to be the same for all subjects.

Missing Data

The problem of missing data is common to all studies in epidemiological research. In the context of longitudinal studies, missing data take the form of dropouts, intermittent missing data, or both. A dropout occurs when a subject begins the study but fails to complete it, and intermittent missingness refers to the situation where a subject misses at least one visit but ultimately completes the study. When the missing data process is not properly investigated by the investigator, inferences may be misleading. Any attempt to accommodate missing data in the modeling process depends primarily on the missing data process—that is, the underlying mechanism by which the data are missing. Little and Rubin (1987) have developed a terminology that is helpful in categorizing and understanding different types of missing data. Data are classified as missing completely at random (MCAR),

missing at random (MAR), and missing not at random (MNAR), depending on whether the fact of missingness is related to (1) none of the outcomes, (2) the observed outcomes only, or (3) both observed and unobserved outcomes, respectively. Under an MCAR mechanism, subjects with complete data may be considered a random sample of all subjects, and a simple remedy to the missing data problems is to delete all subjects with incomplete records. Although the analysis of the resulting complete data set is reasonably straightforward to perform using standard commercial software, this approach may result in substantial loss of subjects, particularly in studies with a large amount of incomplete data. In addition, the MCAR assumption seldom applies to dropouts in longitudinal studies: More typically, those who drop out differ from those who do not drop out with respect to the outcome under investigation. Under an MAR mechanism, imputation techniques that are available with many standard software packages can be used to fill in the missing holes. These techniques range from simple to multiple imputation methods. Such methods, particularly simple imputation, must be used with care because they can introduce new types of bias into the data.

From a modeling standpoint, likelihood-based models produce valid inferences under the MAR mechanism. GEE-based models only produce valid inferences under an MCAR process, but when properly weighted, these models have been shown to be also valid under an MAR mechanism. When the MNAR mechanism applies, that is, when the missingness mechanism depends on the unobserved outcomes, both likelihood and GEE-based models are known to produce biased inferences. The missing data process then needs to be modeled explicitly. Several authors have proposed a family of models that incorporates both the information from the response process and the missing data process into a unified estimating function. This has provoked a large debate about the identifiability of such models, which is possible only on the basis of strong and untestable assumptions. One then has to impose restrictions to recover identifiability. Such restrictions are typically carried out by considering a minimal set of parameters, conditional on which the others are estimable. This, therefore, produces a range of models that form the basis of sensitivity analysis, the only meaningful analysis when the missing data process is likely to be informative.

Software Packages

Longitudinal data analysis has been greatly facilitated by the development of routines for longitudinal analysis within standard statistical packages. SAS, STATA, and SPSS are the most common statistical analysis packages used to analyze data from longitudinal studies. SAS has great data management capabilities and is suitable for standard statistical problems, but may be difficult to master for the practicing epidemiologist. STATA is becoming increasingly popular among epidemiologists because of its interactive nature. Most important, STATA has many procedures tailored to sophisticated biomedical analysis. SPSS is easier to learn and very popular among socioepidemiologists. These statistical software packages come with manuals that explain the syntax of the routines as well as the underlying theories behind the statistical techniques. Despite these advances, it should be noted that there are some complex data that may not be analyzed using standard statistical software with appropriate methods. The investigator is then required to generate his or her own computer codes using matrix-oriented programming languages to answer a specific research question. This then necessitates a good collaboration between the epidemiologist and the statistician or the computer programmer.

—David Todem

See also Cohort Effects; Missing Data Methods; Study Design; Survival Analysis

Further Readings

Diggle, P. D., & Kenward, M. G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, *43*, 49–93.

Hulley, S. B., & Cummings, S. R. (1988). *Designing clinical research*. Baltimore: Williams & Wilkin.

Liang, K.-Y., & Scott, L. Z. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data.* New York: Springer.

Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, *60*, 516–522.

Singer, J. B., & Willett, J. B. (2003). *Applied longitudinal data analysis.* New York: Oxford University Press.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer-Verlag.

Love Canal

In the late 1970s, Love Canal—first a toxic waste site, then a neighborhood in southeastern Niagara Falls, New York—ignited national concerns on hazardous waste disposal and its possible health effects. Following closure of the waste facility in 1953, the land surrounding Love Canal was developed into a blue-collar neighborhood. From the time of its development, residents complained of contamination and health problems. In 1978, high groundwater levels surfaced toxic waste, leading President Jimmy Carter to declare the first man-made federal emergency. Two years later, the crisis at Love Canal provided the impetus for the creation of the Comprehensive Environmental Response, Compensation, and Liability Act, or Superfund.

Background

By the end of the 19th century, Niagara Falls, New York, was a heavily industrialized city. To provide hydroelectricity for local industries, in 1894, Love began construction of a canal connecting the Niagara River and Lake Ontario. A few years later following the discovery of alternating current, financial support for Love's canal bottomed out and construction ceased. Then, in 1942, Hooker Chemicals and Plastics Corporation purchased the incomplete canal and its surrounding land for use as a toxic waste dump. A decade later, the canal filled to capacity with almost 22 tons of mixed chemical waste, the site was closed and covered with dirt. During its 10 years as a toxic waste dump, hundreds of chemicals, including halogenated organics, pesticides, chlorobenzenes, and dioxin, were disposed of at Love Canal. In 1953, Hooker Chemicals sold the property to the Niagara Falls Board of Education for $1. Within the year, a school (the 99th Street School) and residences were built around the former landfill.

From the late 1950s through the early 1970s, Love Canal residents complained of chemical odors, surfacing chemicals, and minor explosions and fires. But it was not until 1976 and 1977 that high groundwater levels due to heavy rains revealed widespread contamination. According to firsthand reports, corroding waste-disposal drums surfaced, vegetation began dying off, and pools of noxious chemicals formed in yards and basements. Testing for toxic chemicals in soil, air, and water by health agencies, prompted by unremitting reporting by the *Niagara Falls Gazette*, confirmed the presence of contamination.

Government Response

In 1978, the New York State Department of Health Commissioner, Robert Whalen, responded to the crisis by declaring a health emergency. Immediately, the 99th Street School was closed, and pregnant women and children below 2 years of age residing closest to the site were evacuated. Over the next 2 years, President Jimmy Carter declared the site a federal state of emergency twice, and about 950 families were relocated from within a 10-mile radius of the site. The use of federal disaster assistance at Love Canal marks the first time federal emergency funds were granted for a nonnatural disaster. In 1980, the Carter Administration passed the Comprehensive Environmental Response, Compensation, and Liability Act, or Superfund, largely in response to the Love Canal crisis.

Health Consequences

Early epidemiological studies of the potential health effects experienced by Love Canal residents have had inconclusive or conflicting results. These studies were limited by a number of factors, including the lack of precise exposure data, small sample size and selection bias, recall bias, and lack of control for confounders. A 2006 study by the New York State Department of Health investigated mortality, cancer incidence, and reproductive outcomes of Love Canal residents. Although this study had limitations similar to those of the earlier research, it suggested increased rates of congenital malformations and proportions of female births. The study also revealed an increased number of adverse reproductive outcomes for women exposed as a child or whose mothers resided in Love Canal during pregnancy.

—Michelle Kirian

See also Birth Defects; Environmental and Occupational Epidemiology; Pollution

Further Readings

Beck, E. (1979). *History: The Love Canal tragedy*. *The United States environmental protection agency*. Retrieved December 11, 2006, from http://www.epa.gov/history/topics/lovecanal/01.htm.

Center for Environmental Health, New York State Department of Health. (2006). *Love Canal follow-up health study-project report to ATSDR (Public Comment Draft)*. Love Canal: Author. Retrieved December 12, 2006, from http://www.health.state.ny.us/nysdoh/lcanal/docs/report_public_comment_draft.pdf.

Ecumenical Task Force of the Niagara Frontier. (1998). *Love Canal collection*. Buffalo: University Archives, University Libraries, State University of New York at Buffalo. Retrieved December 13, 2006, from http://ublib.buffalo.edu/libraries/projects/lovecanal/.

M

Malaria

Malaria is a parasitic disease that causes between 1 million and 3 million deaths each year, mainly African children. Three billion persons—close to 50% of the world's population—live in 107 countries and territories in which malaria is endemic. Most mortality is due to *Plasmodium falciparum*, a protozoan parasite transmitted by the Anopheles mosquito, which is responsible for more than 515 million cases of disease annually: In addition, almost 5 billion febrile episodes resembling malaria, but which cannot be definitively identified as such, occur in endemic areas annually. The medical, epidemiologic, and economic burdens due to malaria have greatly impeded development in endemic countries, particularly in Africa.

Cause of Malaria and Natural Cycle

The four species of the genus *Plasmodium* that cause malarial infections in humans are *P. falciparum*, *P. vivax*, *P. ovale*, and *P. malariae.* Human infection begins when the malaria vector, a female anopheline mosquito, inoculates infectious plasmodial sporozoites from its salivary gland into humans during a bloodmeal. The sporozoites mature in the liver and are released into the bloodstream as merozoites. These invade red blood cells, causing malaria fevers. Some forms of the parasites (gametocytes) are ingested by anopheline mosquitoes during feeding and develop into sporozoites, restarting the cycle.

Manifestations

The complex interrelationships of the malaria parasite, the female Anopheles mosquito vector, and the human target, along with environmental factors and control measures, determine the expression of disease manifestations and epidemiology.

The first symptoms of malaria are nonspecific: Patients are unwell and have headache, fatigue, abdominal discomfort, and muscle aches followed by fever—all similar to a minor viral illness. Later, fever spikes, chills, and rigors occur. Anemia, hypoglycemia, cerebral manifestations, and low birthweight newborns result frequently from malaria as do neurocognitive sequelae after severe illness. Those who have been exposed to malaria develop partial immunity, but not protection from infection, such that they have parasitemia but not illness: This condition is called premunition and is the reason that adults living in malarious areas have much less illness, despite being bitten by infected mosquitoes. In addition, many persons may have comorbidity—malaria parasitemia or illness at the same time that they have other diseases, complicating the diagnosis.

Case Fatality Rates and Sequelae

Correctly and promptly treated, uncomplicated falciparum malaria has a mortality rate of approximately 0.1%. Once vital organ dysfunction occurs or the proportion of erythrocytes infected increases to more than 3%, mortality rises steeply. Coma is a characteristic and ominous feature of falciparum malaria and,

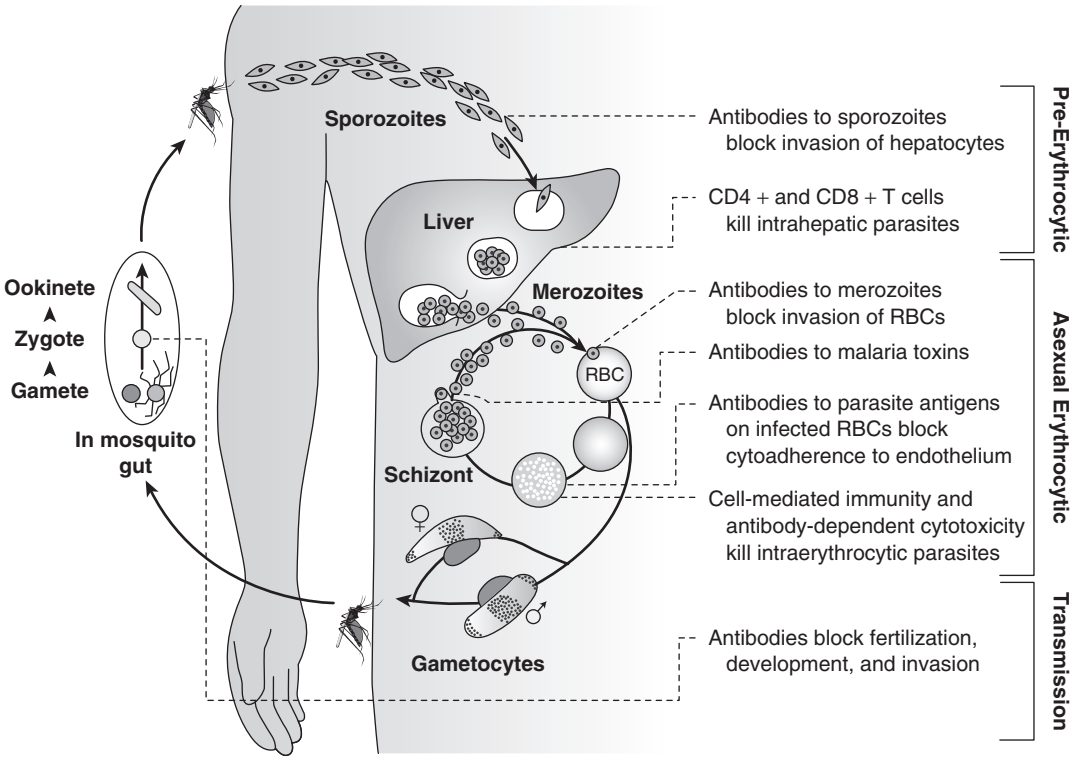


Figure 1 Natural Cycle of Malaria

Source: White and Breman (2005). Copyright © 2006 by The McGraw-Hill Companies, Inc. Reproduced with permission of The McGraw-Hill Companies.

despite treatment, is associated with death rates of some 20% among adults and 15% among children. Convulsions, usually generalized and often repeated, occur in up to 50% of children with cerebral malaria (CM). Whereas less than 3% of adults suffer neurological sequelae, roughly 10% to 15% of children surviving CM—especially those with hypoglycemia, severe malarial anemia (SMA), repeated seizures, and deep coma—have some residual neurological deficit when they regain consciousness. Protein-calorie undernutrition and micronutrient deficiencies, particularly zinc and vitamin A, contribute substantially to the malaria burden.

Where, When, and Why Malaria Occurs

P. falciparum predominates in Haiti, Papua New Guinea, and sub-Saharan Africa. *P. vivax* is more common in Central America and the Indian subcontinent and causes more than 80 million clinical episodes of illness yearly. The prevalence of these two species is approximately equal in the Indian subcontinent, eastern Asia, Oceania, and South America. *P. malariae* is found in most endemic areas, especially throughout sub-Saharan Africa, but is much less common than the other species. *P. ovale* is unusual outside Africa, and where it is found accounts for less than 1% of isolates.

While more than 40 anophelines can transmit malaria, the most effective are those such as *Anopheles gambiae*, which are long-lived, occur in high densities in tropical climates (particularly sub-Saharan Africa), breed readily, and bite humans in preference to other animals. Females require blood for nourishing their eggs; therefore, they bite animals. The entomological inoculation rate (EIR)—that is, the number of sporozoite-positive mosquito bites per person per year—is the most useful measure of malarial transmission and varies from less than 1 in some parts of Latin America and Southeast Asia to more than 300 in parts of tropical Africa. Also important is the basic reproduction rate (Ro), the number of infected persons

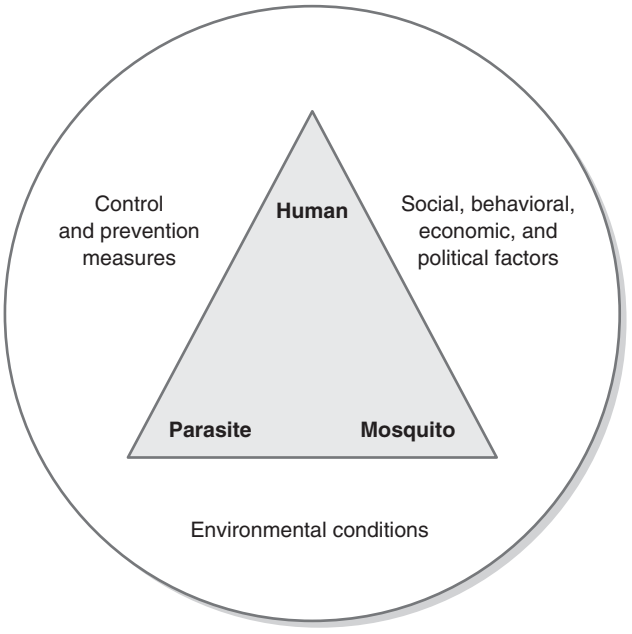


Figure 2 Malaria Ecology and Burden: Intrinsic and Extrinsic Factors

deriving from a single infected person. Ro for malaria may range from 1 to $> 1{,}000$.

Geographic and climate-driven (mainly rainfall) models of suitability for malaria transmission characterize the diversity of malaria transmission. The African continent illustrates four distinct areas of transmission that also exist in Latin America and Asia:

- Class 1, no transmission (northern and parts of southern Africa)
- Class 2, marginal risk (mainly in some areas of southern Africa and in high-altitude [$> 1{,}500$ m] settings)
- Class 3, seasonal transmission with epidemic potential (along the Sahara fringe and in highlands)
- Class 4, stable and unstable malarious areas (most areas south of the Sahara to southern Africa and below an altitude of around 1,500 m)

The epidemiology of malaria may vary considerably within relatively small geographic areas. In tropical Africa or coastal Papua New Guinea, with *P. falciparum* transmission, more than one human bite per infected mosquito can occur per day and people are infected repeatedly throughout their lives. In such areas, morbidity and mortality during early childhood are considerable. For survivors, some immunity against disease develops in these areas, and by adulthood, most malarial infections are asymptomatic. This situation, with frequent, intense, year-round transmission and high EIRs is termed *stable malaria.* In areas where transmission is low, erratic, or focal, full protective immunity is not acquired and symptomatic disease may occur at all ages. This situation is termed *unstable malaria.* An epidemic or complex emergency can develop when changes in environmental, economic, or social conditions occur, such as heavy rains following drought or migrations of refugees or workers from a nonmalarious to an endemic region. A breakdown in malaria control and prevention services intensifies epidemic conditions. Epidemics occur most often in areas with unstable malaria, such as Ethiopia, northern India, Madagascar, Sri Lanka, and southern Africa. Many other African countries situated in the Sahelian and sub-Saharan areas are susceptible to epidemics. Public health specialists have only recently begun to appreciate the considerable contribution of urban malaria, with up to 28% of the burden in Africa occurring in rapidly growing urban centers.

Burden

In 2001, the World Health Organization (WHO) ranked malaria as the eighth highest contributor to the global disease burden, as reflected in disability-adjusted life years (DALYs), and the second highest in Africa after HIV/AIDS. Malaria is the biggest killer of African children below the age of 5 years (more than 1 million deaths per year), followed by pneumonia and diarrhea. Alarmingly, the burden of malaria in children in Africa has been growing since 1990, whereas overall childhood mortality is dropping.

The DALYs attributable to malaria were estimated largely from the effects of *P. falciparum* infection as a direct cause of death and the much smaller contributions of short-duration, self-limiting, or treated mild febrile events, including malaria-specific mild anemia and neurological disability following CM. The estimate assumes that each illness event or death can be attributed only to a single cause that can be measured reliably. Table 1 shows deaths and DALYs from deaths attributable to malaria and to all causes by WHO region. It does not include the considerable toll caused by the burden of malaria-related moderate and severe anemia, low birthweight, and comorbid events. Sub-Saharan African children below 4 years represent

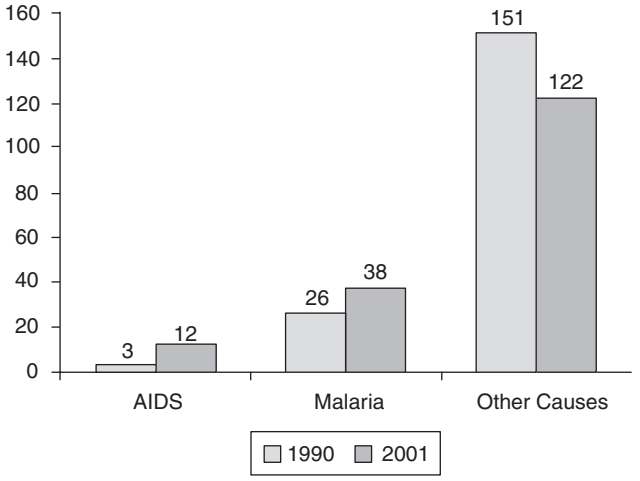


Figure 3 Under-5 Deaths per 1,000 Births

Source: Jamison (2006).

82% of all malaria-related deaths and DALYs. Malaria accounts for 2.0% of global deaths and 2.9% of global DALYs. In the African region of WHO, 9.0% of deaths and 10.1% of DALYs are attributable to malaria.

While the malaria incidence globally is 236 episodes per 1,000 persons per year in all endemic areas, it ranges from about 400 to 2,000 (median 830) episodes per 1,000 persons per year in areas with intense, stable transmission; these areas represent 38% of all falciparum endemic areas.

The vast majority of deaths and illness in developing countries occur outside the formal health service, and in Africa, most government civil registration systems are incomplete. Newer demographic and disease-tracking systems are being used globally and should help rectify the woefully inadequate vital statistics available for malaria and other diseases. The current woeful health system coverage and registration of health events is analogous to "the ears of a hippopotamus".

Health personnel usually attribute causes of death during demographic surveillance system surveys through a verbal autopsy interview with relatives of the deceased about the symptoms and signs associated with the terminal illness. Both the specificity and the sensitivity of verbal autopsy vary considerably depending on the background spectrum of other common diseases, such as acute respiratory infection, gastroenteritis, and meningitis, which share common clinical features with malaria.

In most of the countries where malaria is endemic, laboratory confirmation occurs infrequently. Clinical impressions and treatments are based on febrile illness, which may occur 5 to 10 times a year in young children and may have causes other than malaria. In malarious Africa, some 30% to 60% of outpatients with fever may have parasitemia. Monthly surveillance of households will detect a quarter of the medical events that are detected through weekly surveillance, and weekly contacts with cohorts identify approximately 75% of events detected through daily surveillance. Given the predominance of fevers, malaria case management in Africa and other endemic areas usually centers on presumptive diagnosis.

Estimates of the frequency of fever among children suggest one episode every 40 days. If we assume that the perceived frequency of fever in Africa is similar across all transmission areas (and possibly all ages), African countries would witness approximately 4.9 billion febrile events each year. Estimates indicate that in areas of stable malaria risk, a minimum of 2.7 billion exposures to antimalarial treatment will occur each year for parasitemic persons, or 4.93 per person per year. While these diagnostic, patient management, and drug delivery assumptions are debatable, they indicate the magnitude of the challenges that malaria presents.

Studies of neurological sequelae after severe malaria indicated that 3% to 28% of survivors suffered from such sequelae, including prolonged coma and seizures. CM is associated with hemiparesis, quadriparesis, hearing and visual impairments, speech and language difficulties, behavioral problems, epilepsy, and other problems. The incidence of neurocognitive sequelae following severe malaria is only a fraction of the true residual burden, and the impact of milder illness is unknown.

Indirect and Comorbid Risks

The DALY model of malaria does not sufficiently take it into account as an indirect cause of broader morbid risks. Some consider anemia to be caused indirectly unless linked to acute, high-density parasitemia. Similarly, low birthweight may also be indirectly attributable to malaria, and a child's later undernutrition and growth retardation linked to malaria infection enhances the severity of other concomitant

Table 1 Deaths and DALYs From Deaths Attributable to All Causes and to Malaria by WHO Region, 2000

		Deaths, 2000					*DALYs From Deaths, 2000*				
		All Causes		*Malaria*			*All Causes*		*Malaria*		
Region	*Population*	*Thousands*	*Percentage*	*Thousands*	*Percentage*	*Malaria Deaths as a Percentage of All Deaths*	*Thousands*	*Percentage*	*Thousands*	*Percentage*	*Malaria DALYs as a Percentage of All DALYs*
World	6,122,211	56,554	100.0	1,124	100.0	2.00	1,467,257	100.0	42,279	100.0	2.90
Africa	655,476	10,681	18.9	963	85.7	9.00	357,884	24.4	36,012	85.2	10.10
Americas	837,967	5,911	10.5	1		0.02	145,217	9.9	108	0.2	0.07
Eastern Mediterranean	493,091	4,156	7.3	55	4.9	1.30	136,221	9.3	2,050	4.8	1.50
Europe	874,178	9,703	17.2	< 1	< 0.1	< 0.010	151,223	10.3	20	0.04	0.01
Southeast Asia	1,559,810	14,467	25.6	95	8.5	0.70	418,844	28.5	3,680	8.7	0.90
Western Pacific	1,701,689	11,636	20.6	10	0.9		257,868	17.6	409	1.0	0.20

Source: Breman, Mills, and Snow (2006).

Note: Percentages may not add up to 100 because of rounding.

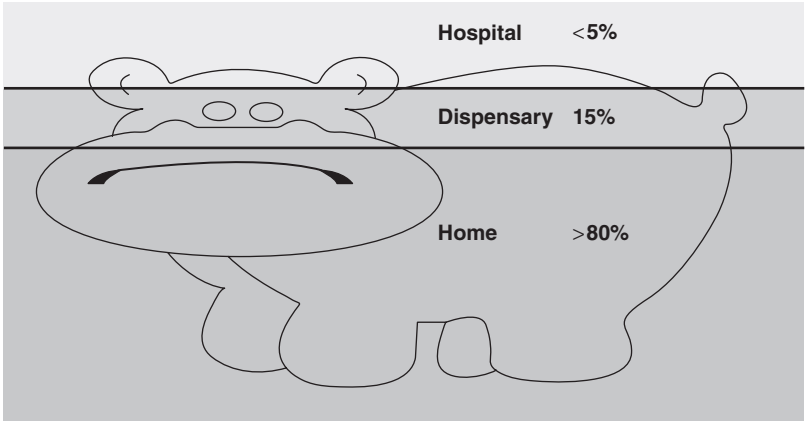


Figure 4 The Ears of the Hippopotamus: Where Malaria Patients Are Managed … and Die

Source: Breman, Egan, and Keusch (2001, p. 6). Reprinted with permission.

or comorbid infectious diseases through immune suppression. Thus, malaria infection contributes to broad causes of mortality beyond the direct fatal consequences of infection and is probably underestimated.

In Africa, pregnant women experience few malaria-specific fever episodes but have an increased risk of anemia and placental sequestration of the parasite. Maternal clinical manifestations are more apparent in areas with less intense transmission, particularly in Asia. Estimates indicate that in sub-Saharan Africa, malaria-associated anemia is responsible for 3.7% of maternal mortality, or approximately 5,300 maternal deaths annually. Prematurity and intrauterine growth retardation resulting in low birthweight associated with maternal malaria account for 3% to 8% of infant mortality in Africa. Assuming an infant mortality rate of 105 per 1,000 live births in 2000, 71,000 to 190,000 infant deaths were attributable to malaria in pregnancy. Other studies indicate that malaria-associated low birthweight accounted for 62,000 to 363,000 infant deaths.

Anemia among African children is caused by a combination of nutritional deficiencies and iron loss through helminth infection, red cell destruction, decreased red cell production as a result of infectious diseases, and genetically determined hemoglobinopathies. Chronic or repeated infections, often associated with parasite resistance to drugs, are more likely to involve bone marrow suppression.

It is estimated that 190,000 to 974,000 deaths per year in sub-Saharan Africa are attributable to SMA. Children residing in areas where the prevalence of *P. falciparum* was more than 25% had a 75% prevalence of anemia. By modeling the relationship between anemia and parasite prevalence, it was found that mild anemia rose 6% for every 10% increase in the prevalence of infection. Reducing the incidence of new infections through insecticide-treated nets (ITNs) or the prevalence of blood-stage infections through chemoprophylaxis or intermittent preventive treatment (IPT) for children halved the risk of anemia.

Iron, zinc, and protein-calorie deficits are responsible for a considerable amount of malaria-related mortality and morbidity and indicate that 57.3% of deaths of underweight children below 5 years are attributable to nutritional deficiencies. One striking feature of the global distribution of anthropometric markers of undernutrition is its congruence with the distribution of endemic malaria.

Early during the HIV epidemic, it was demonstrated that malaria-associated anemia treated with unscreened blood transfusions contributed to HIV transmission. At the same time, two longitudinal cohort studies in Kenya and Uganda and one hospital-based case-control study in Uganda demonstrated that HIV infection approximately doubles the risk of malaria parasitemia and clinical malaria in nonpregnant adults and that increased HIV immunosuppression is associated with higher-density parasitemias. In pregnant women, the presence of HIV increases the rate and intensity of parasitemia and frequency of anemia.

Malaria accounts for 13% to 15% of medical reasons for absenteeism from school, but little information is available on the performance of parasitized schoolchildren. A randomized placebo control study of chloroquine prophylaxis in Sri Lankan schoolchildren demonstrated an improvement in mathematics and language scores by those who received chloroquine but found no difference in absenteeism. As noted earlier, malaria may result in low birthweight, and low birthweight can lead to a range of persistent impaired outcomes, predominantly behavioral difficulties, cerebral palsy, mental retardation, blindness, and deafness. The recently launched studies of intermittent preventive treatments during infancy (IPTi) should provide a more precise means of examining the benefits of IPTi and consequences on learning and performance of infection early in life.

Interventions and Their Effectiveness

Malaria will be conquered only by full coverage, access to, and use of antimalarial services by priority groups; prompt and effective patient management (rapid, accurate diagnosis, treatment, counseling and education, referral); judicious use of insecticides to kill and repel the mosquito vector, including the use of ITNs; and control of epidemics. All interventions must be applied in a cost-effective manner. Eliminating malaria from most endemic areas remains a huge, but surmountable challenge because of the widespread *Anopheles* breeding sites, the large number of infected people, the use of ineffective antimalarial drugs, and the inadequacies of resources, infrastructure, and control programs. WHO Global Malaria Programme and the Roll Back Malaria Partnership, which began in 1998, aim to halve the burden of malaria by 2010; they have developed strategies and targets for 2005 (Table 2).

While ambitious, the initiative is making substantial progress by means of effective and efficient deployment of currently available interventions. Indeed, Brazil, Eritrea, India, Vietnam, and other countries are reporting recent successes in reducing the malaria burden. Despite the enormous investment in developing a malaria vaccine administered by means of a simple schedule and recent promising results in the laboratory and in field trials in Africa, no effective, long-lasting vaccine is likely to be available for general use in the

Table 2 Targets Established at the Abuja Malaria Summit, April 2000

The goal of Roll Back Malaria (RBM) is to halve the burden of malaria by 2010. The following targets for specific intervention strategies were established at the Abuja Malaria Summit, April 2000.

RBM Strategy	*Abuja Target (by 2005)*
• Prompt access to effective treatment • Insecticide-treated nets (ITNs) • Prevention and control of malaria in pregnant women • Malaria epidemic and emergency response	• 60% of those suffering with malaria should have access to and be able to use correct, affordable, and appropriate treatment within 24 hr of the onset of symptoms • 60% of those at risk for malaria, particularly children below 5 years of age and pregnant women, will benefit from a suitable combination of personal and community protective measures, such as ITN • 60% of pregnant women at risk of malaria will be covered with suitable combinations of personal and community protective measures, such as ITN • 60% of pregnant women at risk of malaria will have access to intermittent preventive treatment[a] • 60% of epidemics are detected within 2 wk of onset • 60% of epidemics are responded to within 2 wk of detection

Source: Breman, Mills, and Snow (2006).

a. The original Abuja declaration included the recommendation for chemoprophylaxis as well, but present WHO and RBM policy strongly recommends intermittent preventive treatment, and not chemoprophylaxis, for prevention of malaria during pregnancy.

near future. Yet the search for new and improved interventions continues, and discovery of better drugs, vaccines, diagnostics, and vector control inventions will someday lead to the conquest of this disease.

—Joel Breman

See also Epidemiology in Developing Countries; Insect-Borne Disease; Parasitic Diseases

Further Readings

Breman, J. G., Alilio, M. S., & Mills, A. (2004). Conquering the intolerable burden of malaria: What's new, what's needed: A summary. *American Journal of Tropical Medicine and Hygiene*, *71*(Suppl. 2), 1–15.

Breman, J. G., Egan, A., & Keusch, G. (2001). The ears of the hippopotamus: Manifestations, determinants, and estimates of the malaria burden. *American Journal of Tropical Medicine and Hygiene*, *64*(Suppl. 1/2), 1–11.

Breman, J. G., Mills, A., & Snow, R. (2006). Conquering malaria. In D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, et al. (Eds.), *Disease control priorities in developing countries* (2nd ed., pp. 413–431). New York: Oxford University Press.

Craig, M. H., Snow, R. W., & le Sueur, D. (1999). A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today*, *15*, 105–111.

Greenberg, A. E., Ntumbanzondo, M., Ntula, N., Mawa, L., Howell, J., & Davachi, F. (1989). Hospital-based surveillance of malaria-related paediatric morbidity and mortality in Kinshasa, Zaire. *Bulletin of the World Health Organization*, *67*(2), 189–196.

Holding, P. A., & Kitsao-Wekulo, P. K. (2004). Describing the burden of malaria on child development: What should we be measuring and how should we be measuring it? *American Journal of Tropical Medicine and Hygiene*, *71*(Suppl. 2), 71–79.

Jamison, D. T. (2006). Investing in health. In D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, et al. (Eds.), *Disease control priorities in developing countries* (2nd ed., pp. 3–34). New York: Oxford University Press.

Mendis, K., Sina, B. J., Marchesini, P., & Carter, R. (2001). The neglected burden of *Plasmodium vivax* malaria. *American Journal of Tropical Medicine and Hygiene*, *64*(Suppl. 1), 97–105.

Mung'Ala-Odera, V., Snow, R. W., & Newton, C. R. J. C. (2004). The burden of the neurocognitive impairment associated with falciparum malaria in sub-Saharan Africa. *American Journal of Tropical Medicine and Hygiene*, *71*(Suppl. 2), 64–70.

Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., & Hay, S. I. (2005). The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*, *434*, 214–217.

Steketee, R. W., Wirima, J. J., Hightower, A. W., Slutsker, L., Heymann, D. L., & Breman, J. G. (1996). The effect of malaria and malaria prevention in pregnancy on offspring birthweight, prematurity, and intrauterine growth retardation in rural Malawi. *American Journal of Tropical Medicine and Hygiene*, *55*(Suppl.) 33–41.

White, N. J., & Breman, J. G. (2005). Malaria and babesiosis: Diseases caused by red blood cell parasites. In D. Kasper, E. Braunwald, A. S. Fauci, S. L. Hauser, D. L. Longo, & J. L. Jameson (Eds.), *Harrison's principles of internal medicine* (16th ed., pp. 1218–1233). New York: McGraw-Hill.

World Health Organization. (2005). *The World Malaria Report 2005: Roll Back Malaria.* Geneva, Switzerland: Author.

Worrall, E., Rietveld, A., & Delacollette, C. (2004). The burden of malaria epidemics and cost-effectiveness of interventions in epidemic situations in Africa. *American Journal of Tropical Medicine and Hygiene*, *71*(Suppl. 2), 136–140.

Malnutrition, Measurement of

Malnutrition is a serious global issue, affecting more than 2 billion people worldwide. The problem has two principal constituents—protein-energy malnutrition and deficiencies in micronutrients—and affects women and young children in particular. Malnutrition is the most important risk factor for illness and death in developing countries. Of the many factors that may cause malnutrition, most are related to poor intake of food or to severe or frequent infection, especially in underprivileged populations. Because malnutrition and social factors are closely linked, the nutritional status of a population is a good indicator of the quality of life in a community.

Assessment of Nutritional Status

Nutritional status can be measured at the individual or population level. Population-based assessments are typically performed to measure the extent of malnutrition in a community, identify high-risk groups, and estimate the number of people requiring interventions such as supplementary and therapeutic feeding. Estimates of the burden of malnutrition are important at the national and local levels to define strategies for improving the health of the population.

Methods to assess malnutrition include anthropometry, biochemical indicators (e.g., decrease in serum albumin concentration), and clinical signs of malnutrition (e.g., edema, changes in the hair and skin, visible thinness). Anthropometry is the preferred method to assess malnutrition in both individual people and surveyed populations because body measurements are sensitive over the full spectrum of malnutrition, while biochemical and clinical indicators are useful only when malnutrition is advanced. The purpose of the assessment should guide the choice of measurement methods.

Common anthropometric indicators of malnutrition in childhood include combinations of body measurements (e.g., either length or height combined with weight) according to age and sex. Anthropometric measurements of children below the age of 5 years are used to draw conclusions about the nutritional well-being of the population in which they live, because children are more vulnerable to adverse environments and respond more rapidly than adults to dietary changes.

To interpret anthropometric data and determine an individual child's level of malnutrition, the child's height and weight are compared with reference curves of height-for-age, weight-for-age, and weight-for-height. The internationally accepted references were developed by the Centers for Disease Control and Prevention (CDC) using data collected from a population of healthy children in the United States. More recently, the World Health Organization (WHO) released international standards for child growth that were based on a pooled sample from six countries. Anthropometric data can be plotted using software available from the CDC (www.cdc.gov/epiinfo) or WHO (www.who.int/childgrowth/software). The internationally recommended indicators used to characterize the different types of malnutrition in childhood include the following:

- Low height-for-age: Assess stunting or shortness
- Low weight-for-height: Assess wasting or thinness
- Low weight-for-age: Assess underweight

Stunting reflects a failure to reach one's linear growth potential (maximum height) because of chronic malnutrition due to either inadequate intake of food or recurrent illness; wasting, in contrast, indicates recent loss of weight, usually as a consequence of famine or severe disease. Underweight reflects both wasting and stunting, and thus in many cases, it reflects a synthesis of undesirable body proportions and reduced linear growth. The choice of anthropometric indicator depends on the purpose of the assessment. For example, in emergency situations, weight-for-height is the index most often used, because wasting has the greatest potential for causing mortality or widespread morbidity. Other anthropometric indices, such as mid-upper-arm circumference and triceps skinfold thickness, are sometimes used but are less reliable.

Three different classification systems can be used to distinguish "normal" from "not normal" growth in childhood: z scores, percentiles, and percentage of median. Although percentiles are typically used in the United States, WHO recommends the use of the z score (a z score of 1 represents 1 standard deviation from the reference median). Malnutrition is defined as a z score of less than -2 for weight-for-height, height-for-age, or weight-for-age. A cutoff point of a z score of -3 is used to identify severely malnourished children. Abnormal findings in an individual child may reflect normal variation in growth (e.g., low height-for-age because both parents are short).

A simple clinical examination can help detect certain causes of malnutrition. For example, examining a sample of children for pretibial or pedal edema (swelling of the foot/ankle) can determine community rates of kwashiorkor (malnutrition with edema). Children with edema should always be classified as having severe acute malnutrition regardless of their weight-for-height or weight-for-age z scores. To detect malnutrition in adults, particularly nonpregnant women of childbearing age, it is recommended that underweight be used as a proxy for malnutrition. The body mass index (BMI) can be used to measure the prevalence of underweight. BMI is determined by dividing the weight in kilograms by the square of the height in meters. Levels of malnutrition can be classified as follows using the BMI:

- Mild: 17 to < 18.5
- Moderate: 16 to < 17
- Severe: < 16

Conducting a Population-Based Assessment of Nutritional Status

In most population-based assessments of nutritional status, the children are between the ages of 6 and 59 months. Children in this age group are highly

vulnerable to increased morbidity and mortality during a nutrition crisis and will often be the first to exhibit signs of malnutrition.

Before any data are collected, it is important to define what an appropriate population-based sample would be and determine a sample size that gives results of sufficient precision and power. Criteria for obtaining data of good quality include using the right equipment to collect the data and employing standard measuring techniques. Children should be weighed in their underwear without shoes, and the scale should read to the nearest 0.1 kg. For height, children below the age of 2 years should lie on a suitable board to have their length measured; children above 2 years should stand up to have their height determined. Accurate measurements of weight, height, and age are required for the identification of malnutrition to be made. To maintain a high level of data quality, fieldworkers should be trained to collect data in a standardized format and should be supervised.

Once a comparison has been made for every child in the sample between his or her nutritional status and the reference population, the prevalence of malnutrition in the population can be calculated. For example, if the intention is to calculate the prevalence of wasting, one can count all the children in the sample with a weight-for-height z score less than -2 and divide the final count by the total number of children in the sample. A prevalence of wasting $>10\%$ is cause for concern and indicates a need for rapid intervention. A prevalence of stunting $>30\%$ and a prevalence of underweight $>20\%$ are also indicators of severe malnutrition in a community.

Data from a nutritional assessment should be presented in a standard format. Items to report include the general characteristics of the population, the design of the survey used to collect the data, methods of measurement, summary statistics, and the prevalence of malnutrition (stunting, wasting, and underweight) at z scores of less than -2 and less than -3. As needed, prevalence might be presented by age group or sex. The ability to produce meaningful estimates by subgroup will depend on having sufficiently large samples for these classifications.

—Parminder S. Suchdev

See also Body Mass Index (BMI); Centers for Disease Control and Prevention; Nutritional Epidemiology; Sampling Techniques; Vitamin Deficiency Diseases; World Health Organization

Further Readings

Blossner, M., & de Onis, M. (2005). *Malnutrition: Quantifying the health impact at national and local levels*. Geneva, Switzerland: World Health Organization.

Kleinmann, R. (Ed.). (2004). *Pediatric nutrition handbook* (5th ed.). Elk Grove Village, IL: American Academy of Pediatrics.

Muller, O., & Krawinkel, M. (2005). Malnutrition and health in developing countries. *Canadian Medical Association Journal*, *173*, 279–286.

World Food Programme. (2005). *A manual: Measuring and interpreting malnutrition and mortality.* Rome: Author.

Web Sites

Centers for Disease Control and Prevention: http://www.cdc.gov.

World Health Organization: http://www.who.int.

Managed Care

Many Americans hold strong opinions about managed care. Audiences in movie houses across the nation famously broke into cheers when actress Helen Hunt's character cursed it in the 1997 movie *As Good as It Gets.* Managed care is an important element of the health care delivery system in the United States today, and familiarity with it is a prerequisite to understanding the contemporary health care context. However, managed care is not neatly defined. Instead, it is an umbrella term covering a historically changing collection of administrative practices, organizational forms, and business strategies intended to make provision of health care more efficient.

Defining efficiency in this context has posed challenges. Some efficiencies focus on information, including electronic medical records, and profiling provider quality to facilitate informed choice by consumers; however, advocates of managed care like to stress that providing better medical care can itself save money. Examples include preventive care and monitoring of chronic conditions, both of which can avert more expensive care. More problematic are instances when there are real or perceived conflicts between reducing costs and providing good care. These conflicts usually focus on restrictions of various kinds, since for every procedure denied, there is a provider or a patient unhappy to be told no.

Managed Care Practices

It is useful to describe common practices that may be bundled together in various organizational forms. *Gatekeeping* is a role assigned to primary care physicians that aims to limit use of high cost specialty services by requiring patients to get referrals from their primary care physician, who provides it only when judged warranted. *Utilization review* is an administrative practice that aims to limit use of unwarranted treatments by hiring third parties to evaluate them. *Capitation* is a payment practice in which the physician receives a set payment based on the number of patients under care, and in turn assumes the financial risk of providing for those patients' care, a strategy that aims to make physicians cost conscious. This strategy aims to reverse an incentive system that rewards physicians for multiple procedures, some of which might be unnecessary or inefficient. Instead, the healthier a patient remains, the greater the savings for the physician. In theory, the physician could retain the entire payment for the patient who needs no care, although in fact some of the capitation payment may be used to pay for insurance that physicians in capitated plans generally hold to limit their liability for patients with very high medical expenses. *Case management* is a practice in which a role is assigned to someone—typically a nonphysician, managed care employee—whose job it is to limit losses due to fragmentation of care and inadequate services.

Background and Development

Despite experiments outside the United States, the managed care story is overwhelmingly an American one, reflecting the distinctive challenges posed by U.S. employer-based insurance, extensive development and use of high cost technologies, and the relative absence of direct government influence on costs. Early elements of managed care can be found in the 20th century in prepaid group plans, such as the one established for workers in Kaiser Industries in 1945, or the Health Insurance Plan of New York, initially created for city workers in New York City in the 1940s.

The term *health maintenance organization* (HMO), coined in 1970 by physician Paul Ellwood, refers to an organized system of care that, in return for a fixed premium, provides (or contracts for) health care for members. HMOs can be classified as a *staff model* (that hires physicians who treat members, and often also owns hospitals, labs, and other services), a *group model* (that contracts with a physician group that is responsible for paying physicians and contracting with hospitals), or a *network model* (that contracts with multiple physician groups for services).

A second phase in the development of managed care can be dated from the HMO Act of 1973, which was a federal legislation passed as part of a new health policy initiative of the Nixon administration. The bill aimed to expand the proportion of the population with HMO coverage, removed legal barriers to HMO development, tried to seed new HMOs by providing federal dollars to ease the initial financing burden, and spurred entry into existing insurance markets by requiring employers with 25 or more employees to accompany traditional plans with an opportunity to participate in an HMO that met certain federal guidelines. The legislation spurred institutional development, but the overall growth of HMOs fell short of what was anticipated. Only when cost control gained new urgency in the 1980s did managed care gain momentum. Importantly, when rapid growth in managed care organizations (MCOs) was demanded by employers, it was supplied less often by HMOs that directly provided care than by for-profit, largely administrative structures with lower start-up costs. Independent practice associations (IPAs), for example, were formed by individual medical practices that contracted as an organization with an HMO for a per capita fee rate paid for patients.

Context of Growth

The flagging U.S. economy of the late 1980s strengthened the appeal of HMOs to purchasers of managed care. Real per capita health care costs had been increasing for years, but the rate of increase accelerated in the 1980s. Employers had to pay an even greater proportion of labor costs to provide health insurance for their employees, making it harder to compete with foreign counterparts unburdened by employer-based insurance. Health economist Uwe Reinhardt has credited the impact of job insecurity with providing employers with the leverage needed to move employees into health coverage that limited access and benefits to reduce costs. Put simply, Reinhardt believes that employees decided that accepting restricted insurance looked better than being unemployed.

The federal government was a major payer for health care through entitlement-based programs, such

as Medicaid and Medicare (both created in 1965), and so also felt pressure from rising health care costs. Bill Clinton was elected president in 1992 with a vague mandate to reform health care, and his administration saw control of health care costs as a key to restoring economic vitality and combined this economic argument with a political case for equity. Clinton created the Task Force on National Health Care Reform, with the mandate to develop a plan to provide universal health care to all Americans. The Clinton logic appealed to some large employers, but once the Clinton plan was defeated in 1994, employers had little reason to hope for cost relief by government intervention. Many shifted their strategies and began contracting with for-profit MCOs.

Doctors' private practices had long functioned like small businesses, providing care to patients who chose their services. But so long as third parties paid the bills, neither doctor nor patient had strong financial incentives to pay attention to costs. Profit-conscious MCOs are intended to introduce price sensitivity by giving physicians and patients an incentive to shop for the best price for medical care and to negotiate tough deals with providers. Many MCOs initially focused on increasing market share, sometimes sacrificing some short-term profits, to gain negotiating leverage and achieve economies of scale. As more and more insured patients moved into restricted plans, MCOs were able to stimulate price competition among hospitals and providers. Physicians often found themselves in the position of subcontractors, vulnerable to being excluded from health plan panels, that is, not having their services covered by health plans, if their practices did not conform to various plan expectations regarding price and practice style.

Power and Backlash

The tools of managed care proved successful in containing costs for a certain time period. Cost increases were constrained through the mid-1990s, and many benefited from low premiums. But most commentators agree that, sometime in the late 1990s, a widespread backlash against managed care occurred. Public distrust grew. Critics pushed for regulation. Hospitals and providers organized to increase negotiating power. Several explanations have been suggested for this shift in mood. Sheer visibility probably played some role because, although mistakes and inequities have always existed, managed care gave patients someone to blame. Certainly, value conflicts played some role. The shift to managed care ran headlong into the value Americans place on consumer choice of provider and their distaste for limits that connote rationing. In fact, any health care system requires resource allocation choices because it is not possible to provide unlimited health care with limited resources, but the MCOs made these choices explicit. Even if these dissatisfactions had long been present, the upturn in the economy and lower unemployment put employees in a stronger position to gain a hearing for their unhappiness. Once the rise in real per capita health care spending again turned upward after 1997, purchasers could no longer feel confident that economic benefits were being provided in return for the extensive administrative-control-granted MCOs.

MCOs responded with products that allow more choices, but require the patient to pay more for certain services, for instance, to see a physician outside the plan. Examples include reliance on preferred provider organizations (PPOs), point of service (POS) plans, and most recently, a variety of "tiered" plans. In a PPO, physicians contract to offer services to plan participants for a favorable rate. Patients are encouraged to use them by lower co-payments and deductibles, but they can go elsewhere. Facing a similar set of incentives, the patient in a POS plan can go to an individual plan physician for the prepaid fee or opt to go to an outside physician for a higher co-payment. However, these concessions to consumer desire for choice may have loosened the MCO grip on costs. Some see evidence of a modest return to stricter controls, such as utilization review. Also, some firms are building "tiered networks," created by classifying hospitals and providers based on their cost of care, then using cost-sharing incentives to encourage patients to rely on lower-cost sources.

The long-term significance of managed care for U.S. health care is far from clear. The most straightforward narrative stresses its slow development, followed by its rapid rise, the backlash against it, and its subsequent downfall. Yet an alternative version might stress, not the downfall of managed care, but its protean adaptability and the survival of its component mechanisms, as these have been shuffled, reshuffled, and reassembled. Ultimately, any persuasive assessment should acknowledge, first, that it helped end an era of unprecedented professional and patient autonomy, largely financed by third-party payers with remarkably limited attention to costs; second, that the expansion of market

strategies to increase competition was accompanied by widespread consolidation in the health care sector; and, third, that despite the constraints it introduced, there is little reason to believe that it can on its own provide a satisfactory solution to rising costs.

—James Walkup

See also Formulary, Drug; Governmental Role in Public Health; Health Care Delivery; Health Care Services Utilization

Further Readings

Mechanic, D. (2004). The rise and fall of managed care. *Journal of Health and Social Behavior*, *45*(Suppl. 1), 76–86.

Reinhardt, U. (2001). *On the meaning of a rational health system: The American Health System—past, current, future*. Retrieved December 1, 2006, from http://conferences.mc.duke.edu/privatesector/dpsc2001/da.htm.

Schlesinger, M. (2002). On values and democratic policy making: The deceptively fragile consensus around market-oriented medical care. *Journal of Health Politics, Policy, and Law*, *27*, 889–925.

Matching

Matching is the process of selecting a comparison group so that it is equivalent in terms of certain characteristics (e.g., age or gender) to the group to which it will be compared. Matching is most often used for selection of controls in case-control studies; however, it may be applied in cohort studies as well. This entry describes benefits and drawbacks of matching, as well as the analysis methods applied to matched data. Unless otherwise stated, the discussion refers to matching in case-control studies.

Matching can be performed in different ways. In *individual matching*, one or several controls are selected for each case so that they are equivalent to the case for their values of the variables being matched on. For example, if a case was a female nonsmoker, one or more female nonsmokers would be selected as controls. To match on a continuous variable, such as age, controls can be matched to cases within defined categories (age 20 to 29, 30 to 39, etc.) or within a given increment of the case's value (e.g., ± 3 years). The latter strategy is termed *caliper* matching.

Frequency matching involves selecting controls so that their distribution matches that of the cases for the variables of interest. Using the example above, where matching was on sex and smoking status, if 30% of cases were female nonsmokers, then 30% of controls would also be selected with these characteristics. Frequency matching will tend to require that all cases are identified before control selection to determine the required proportions, whereas individual matching is more conducive to concurrent identification of cases and controls.

Advantages of Matching

There are benefits to matching in addition to the intuitive appeal of comparing groups that appear similar to one another. Matching can facilitate the selection of a referent group without requiring identification of the entire base population. For example, it may be fairly easy to select as a matched control the "next" patient at a hospital or clinic where cases are identified. On the other hand, it may be much more difficult to enumerate and then enroll a random sample of all patients from the hospital or all potential patients from the surrounding geographic area. Matching can also be an efficient way to identify controls when controlling for factors such as neighborhood or sibship is of importance. Because there would be very few existing appropriate control subjects (people from the same neighborhood or sibship as cases) in the overall base population, choosing a random sample of this population is unlikely to yield a suitable control group. Finally, matching may result in a gain in precision of the estimate of association. This will be most apparent when the matching variable is a strong confounder.

Disadvantages of Matching

Matching also has disadvantages that should be carefully considered. Matching on many variables may make it difficult to locate matched controls, and information on the matching factors needs to be collected for "extra" controls that will not actually end up matching to cases. These factors may decrease cost efficiency of the study. Also, matching variables cannot be considered as independent risk factors themselves. This is because they have been set by design to be distributed similarly in cases and controls. (It *is* still possible to assess effect modification by the matching variables.) Finally, overmatching may result

in reduced statistical precision or a biased estimate of association. Matching on strong correlates of exposure, variables associated with exposure but not disease, or factors that are affected by the outcome or exposure of interest should be avoided.

Analysis of Matched Data

Matched data require special consideration in the analysis. This is because the process of matching induces selection bias, so an unmatched analysis will generally lead to a biased estimate of the odds ratio. For individually matched pairs, a crude odds ratio can be calculated using a 2×2 table, set up as shown in Figure 1.

In this table, pairs, rather than individuals, contribute to the cell counts. The matched pairs odds ratio (OR) is calculated as $OR = b/c$ based on the discordant pairs only. To calculate an odds ratio adjusted for multiple covariates, conditional logistic regression is used. For frequency matched data, the methods used are similar to those for unmatched data, with the matching variables included as covariates in all analyses. For individually matched data, it may be possible to conduct a frequency matched analysis if the pairs or sets can be condensed into strata where all sets have equal values for the matching variables.

Matching in Other Study Types

The concept of matching is implicit in certain specific study designs. A case-crossover study is a special instance of matching in which individuals serve as their own controls, while in a nested case-control study with incidence density sampling, controls are matched to cases according to follow-up time. Matching of unexposed to exposed subjects in cohort studies can also be done, and may increase precision. However, this benefit is not guaranteed and the practice is not very common. Other strategies for matching that fall somewhere in between random and fully matched sampling have been described using terms such as partial, marginal, and flexible matching. The term *countermatching* refers to a strategy of choosing controls in a nested study based on their exposure or a proxy of exposure, rather than a confounder. Instead of making members of matched case-control sets similar to one another, as accomplished by the typical matching strategy, countermatching aims to maximize variability within these sets.

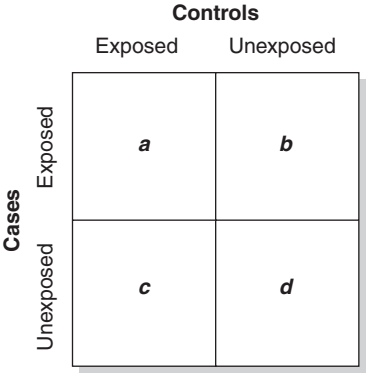


Figure 1 Calculation of the Odds Ratio Using Individually Matched Data.

—Keely Cheslack-Postava

See also Control Group; Overmatching; Stratified Methods; Study Design

Further Readings

Kelsey, J. L., Whittemore, A. S., Evans, A. S., & Thompson, W. D. (1996). Case-control studies: II. Further design considerations and statistical analysis. In *Methods in observational epidemiology* (2nd ed., pp. 214–243). New York: Oxford University Press.

Langholz, B., & Clayton, D. (1994). Sampling strategies in nested case-control studies. *Environmental Health Perspectives*, *102*(Suppl. 8), 47–51.

Rothman, K. J., & Greenland, S. (1998). Matching. In *Modern epidemiology* (2nd ed., pp. 147–161). Philadelphia: Lippincott Williams & Wilkins.

Szklo, M., & Nieto, F. J. (2000). Basic study designs in analytical epidemiology. In *Epidemiology beyond the basics* (pp. 3–51). Gaithersburg, MD: Aspen.

Maternal and Child Health Epidemiology

Maternal and child health epidemiology is concerned with determinants of health in populations of women, infants, children, and families, with a particular focus on women's health during pregnancy and after giving birth and on neonatal and early childhood health. This entry describes frequently used indicators of maternal and child health from different parts of the world. It also discusses problems related to obtaining valid data

about child and maternal health and puts these issues into a public health perspective.

The Mother-Child Relationship

Mother and child form the most basic partnership in evolution, and the health of the child is to a large extent determined by the health of the mother. In the very early phase of life, they also compete for nutrition. An example of this competition is that the average birthweight is less than the optimal birthweight (the birthweight with the lowest mortality). The mother needs to reserve a certain amount of her supply of nutrition to take care of the child and to be able to reproduce in the future. The fetus needs to grow as large as possible to survive the risks it meets outside the uterus.

The unborn child has a remarkable developmental plasticity. The fetus must adapt to the uterine environment and prepare for extrauterine life. When these adaptations are appropriate to the reality of the child's life outside the womb, they are beneficial; but in other cases, they may be inappropriate, and the health consequences may be serious. For example, insulin resistance may be a fetal response to a temporary insufficient food supply to slow down growth in order to preserve the limited energy available for brain development and to prepare for a life with an expected shortage of food. Insulin resistance may be an advantage in such an environment, since it facilitates storing fat in time periods when food is plentiful. On the other hand, if food becomes unduly plentiful, insulin resistance may predispose an individual to obesity and diabetes because glucose transport to cells is impaired and insulin production may not be able to keep up with demands. Lack of food, stress, infections, and environmental exposures not only affect the health of the pregnant mother but may also have lifelong implications for the unborn child.

Indicators of Maternal and Child Health

A number of specific indicators have been developed to facilitate surveillance and research on maternal and child health issues. Data on the following indicators based on the definitions of the World Health Organization (WHO) are available in many countries. Except for the fertility rate, they deal with causes of mortality.

- *Perinatal Mortality Rate.* The risk of fetal or infant death during late pregnancy (at 22 completed weeks of gestation and more), during childbirth, and up to 7 completed days of life, expressed per 1,000 total births.
- *Neonatal Mortality Rate.* The risk of dying between birth and within the first 28 days of life, expressed per 1,000 live births.
- *Infant Mortality Rate.* The risk of dying between birth and exactly 1 year of life, expressed per 1,000 live births.
- *Child Mortality Rate, Under 5.* The risk of dying before age 5, expressed per 1,000 live births.
- *Maternal Death.* The pregnancy-related death of a woman, either during pregnancy or within 42 days of termination of pregnancy.
- *Maternal Mortality Ratio.* The number of maternal deaths per 100,000 live births.
- *Total Fertility Rate.* Estimated number of live births that a woman will have from her 15th year through her childbearing years.

Child and Maternal Mortality

One of the main contributors to the improvements in life expectancy in the 20th century was the reduction in child mortality; in 2003, the global mortality rate in children under 5 reached a low of 79 per 1,000. However, the most progress has been made in high-income countries. There is an enormous gap in child mortality between the richest and the poorest parts of the world.

During the past 30 to 40 years, many national and international organizations have tried to reduce inequalities in health by improving vaccination coverage and breastfeeding practices, and by reducing malnutrition and deaths from diarrhea. The global mortality rate for children below 5 years of age was halved from 1960 to 1990. The WHO aims at reaching a two-thirds reduction from 1990 to 2015, but progress has been modest in recent years. Communicable diseases such as pneumonia, diarrhea, measles, malaria, and tuberculosis are still the main killers in poor countries, and the HIV/AIDS epidemic in sub-Saharan Africa has even erased the survival and health gains in these countries. About 11 million children below the age of 5 still die annually from preventable causes.

The global neonatal mortality rate in 2000 was estimated at 30 per 1,000, but with large geographical variations. Moreover, better survival has been achieved mainly in children who have survived the first month of life, while neonatal mortality, especially in the

Table 1 Neonatal and Maternal Mortality in Countries Where the Decline in Child Mortality Has Stagnated or Reversed

Decline in Child Mortality (1990–2003)	*Number of Countries*	*Percentage of Live Births (2000–2005)*	*Under-5 Mortality Rate (1990)*	*Under-5 Mortality Rate (2003)*	*Neonatal Mortality Rate (2000)*	*Maternal Mortality Rate (2000)*
On track	30 (OECD)	11%	22	13	7	29
	63 (non-OECD)	23%	78	39	19	216
Slow progress	51	44%	92	72	35	364
In reversal	14	6%	111	139	41	789
Stagnating	29	16%	207	188	47	959

Source: Adapted from the World Health Organization, "The World Health Report 2005. Make every mother and child count."

first critical days of life, has undergone more modest reductions. Almost 30% of all child deaths in 2000 happened in the first week of life and were due mainly to infections, birth asphyxia, and prematurity. To reach the WHO goal for child mortality, comprehensive health care programs during pregnancy, during childbirth, and in the postnatal period are needed. However, these interventions may be more expensive to implement than the interventions needed later in infancy.

In 2000, 529,000 women died as a result of pregnancy or childbirth, which is equivalent to a global maternal mortality ratio of 400 per 100,000. Only 1% of these deaths occurred in high-income countries. Most causes of maternal deaths are preventable complications to pregnancy and childbirth, such as hemorrhage, obstructed labor, sepsis, eclampsia, unsafe abortion, and anemia. In Africa, HIV/AIDS, malaria, and the tradition of genital mutilation contribute to and worsen these complications. Better health care during pregnancy and childbirth not only improves neonatal survival but also reduces maternal mortality and the numbers of stillbirths. At present, it is estimated that 3.3 million babies are stillborn—a number that approaches that of the 4 million infants who die within the first 28 days of life. The provision of skilled perinatal care is crucial to further reduce child and maternal mortality.

Other factors related to maternal mortality are birth spacing, use of contraceptives, and use of safe methods of abortion. It is estimated that 19% of married women in low-income countries have unmet contraceptive needs.

Other Aspects of Child and Maternal Health

In most high-income countries, maternal mortality is now low, but other aspects of maternal health are reasons for concern. The frequency of preterm deliveries (birth before 37 weeks of pregnancy) is often high (more than 10% in some countries) and may be increasing.

Global fertility rates (the total number of children a woman has over her lifetime) fell from six to three after the introduction of contraception in the last part of the 20th century, and in some countries, the rate is now only slightly above one. It is not known whether the increasing use of infertility treatments in high-income countries reflects increasing infertility problems. However, part of the decrease in fertility rates is an artifact related to delayed reproduction. The measure depends on a steady state situation where age-specific fertility rates do not change over calendar time.

The substantial increase in the rate of caesarean sections cannot be explained by purely medical indications or related improvements in neonatal outcome. The rates are as high as 30% to 50% in some countries of South America. Although caesarean section in a well-functioning health care system is considered a safe mode of delivery, the consequences for future pregnancies are still uncertain.

Children are more vulnerable to many environmental stressors because of their rapid growth, and during intrauterine life, the unborn child is not well protected from external exposures that cross the placenta barrier. The fetus may be unable to metabolize some of

these toxic compounds, and the brain may have vulnerable time periods where lesions could have long-lasting effects. Western lifestyle factors that may interfere with child and maternal health are in great contrast to some of the problems observed in the poorest parts of the world. The abundance of energy-rich foods combined with a lack of physical activity among both parents and children is related to obesity problems in many parts of the world, with an obesity prevalence of 10% to 25% among women of childbearing age in affluent countries and growing obesity problems in childhood. The total burden of diseases related to childhood obesity is still to be discovered, but it is expected that the occurrence of diabetes and cardiovascular diseases will increase and have an earlier onset, accompanied with reduced life quality and physical impairment.

Congenital malformations may be serious or trivial and are difficult to count because some are invisible or are simply deviations from normal structures. Estimates of the frequency of malformations at birth, therefore, ranges from 1% to 7% among all births, depending on how thoroughly the newborns are examined. Some causes of congenital malformations are known, such as radiation, some infections during pregnancy, or specific types of medicine, but in most situations the causes of a specific malformation are unknown. Whether this frequency is increasing or decreasing is unknown.

Asthma and atopic diseases are frequent (about 20% in some countries). The increasing prevalence and the worldwide variation indicate that environmental factors play an important role in the etiology of these diseases, but the underlying mechanisms are poorly understood.

Behavioral problems such as attention-deficit hyperactivity disorder (ADHD) have a frequency of 3% to 15% in affluent societies, and an increasing trend has been suggested in these countries. It should, however, be considered that the condition and therefore also its frequency are defined by manmade cutoff levels in a continuous distribution of behavioral problems.

Measurement Issues in Maternal and Child Health Epidemiology

Health is a broad concept, and data sources that allow international comparisons and comparisons over time cover only part of this concept while those that are available are often far from perfect. Even data on mortality, especially cause-specific mortality, are difficult to get from many countries. For instance, although death is well-defined, death rates may be difficult to calculate because accurate information on the total population may not be known in populations with poor demographic statistics. In addition, the validity of data on causes of death depends on the availability of diagnostic facilities for all who die in the population. In some cases, the causes of death rely on a "verbal autopsy," which is a retrospective interview of close relatives done by people with limited clinical training.

Some of the concepts used in maternal and child health epidemiology are also difficult to operationalize, such as "a pregnancy-related death" used to estimate maternal mortality. If a woman dies of eclamptic seizures or during labor, the death is clearly related to the pregnancy, but if she dies from a stroke or commits suicide, either of which may be related to the pregnancy, the death may not be counted as "pregnancy-related."

All mortality or disease rates are expressed as events per unit of time, and in reproduction the counting time starts at either the time of conception or at the time of birth. Deaths that happen during fetal life would normally be seen as a function of time since conception, but this period is only estimated, and the estimates may lack precision because the exact time of conception is unknown. Stillbirths are therefore routinely calculated not as a function of the number of fetuses in the population under study at the time of death but as a proportion of all births because data on births may be available.

The change from fetal time to extrauterine time normally starts at 266 days after conception. When variation from this time exceeds certain limits (< 37 weeks or > 42 weeks) the terms *preterm birth* or *postterm birth* are used, but these terms are purely descriptive and based on the estimate of gestational age. Estimating gestational age is key and unfortunately not easy. Traditionally, the estimates have been based on Naegel's rule, which states that the due date of birth can be calculated by adding 7 days and then subtracting 3 months to the first day of the last menstrual period. This works well only if menstrual periods are regular, which often is not the case. In affluent societies, estimates are now based on measures of growth by using ultrasound techniques. This principle rests on the assumption that growth in the early phase of life follows the same velocity. This assumption is good enough for making clinical predictions but have some limitations when used in epidemiologic studies.

Since the pregnant woman may carry more than one child, there may even be ambiguity in whether preterm births or preterm confinements have been counted. Preterm births would, for example, increase over time in countries that increasingly use infertility treatments leading to more twins.

Congenital anomalies are by definition present at birth but they may not be diagnosed at birth. For example, some malformations may not manifest themselves until much later in life, although their onset occurs during fetal life, usually in the 2nd and 3rd months of pregnancy. Also, many—perhaps most—of these malformations lead to spontaneous abortions. The number of anomalies (the prevalence) present at birth therefore reflects only some of the cases that occur during fetal life. Abortions that are induced as part of prenatal screening will also affect the prevalence of congenital malformations.

Most data on maternal and childhood health come from ad hoc data collections or review of medical records. Data on some conditions that do not always lead to hospitalization, such as obesity, asthma, pregnancy nausea, or even early spontaneous abortions, depend on results generated in specific studies. For example, it is expected that about 30% of all pregnancies end in spontaneous abortions, but only part of these will be known because they happen before the pregnancy is clinically recognized. These abortions will be detected only if it is possible to follow women trying to become pregnant over time and measure sensitive biological markers of a pregnancy in urine or blood.

Other studies would have to rely on asking questions on nausea, asthma symptoms, or behavioral problems during childhood. Most of these measures will come with some measurement errors. Measures of behavioral problems during childhood will be heavily influenced by actual and present problems when filling in the questionnaires. Such studies will also be biased by people who are invited to take part in the study but refuse or drop out during follow-up.

Public Health Implications

Maternal and child health is not only a medical concern. Social and economic factors play major roles (in particular because women and children are often a low priority in poor countries), and, in turn, poor maternal and child health lead to undesirable social consequences. Poverty and inequity, unstable and unjust political systems, and lack of education are important determinants of maternal and child health. These factors contribute to a vicious cycle that may require political actions to be broken. A child who grows up in poverty and with a shortage of food will not reach his or her optimal growth potential and may suffer reduced mental capacity. A girl with only a limited education will be at high risk of getting pregnant at a young age, and she will be less prepared to care for her own and her baby's health.

During pregnancy, she may have a too low weight gain and be more susceptible to infections, which will impair her chances of surviving the challenges of childbirth. Her short stature due to restricted growth during childhood will furthermore place her in higher risk of preterm birth, prolonged or obstructed labor, and giving birth to a low birthweight baby, and both of them may be at high risk of death or severe impairment. Furthermore, limited energy supplies and other hazards during intrauterine life may alter organ functions and the child may be more susceptible to diseases in later life. Unfortunately, most research takes place in countries where the health problems are smallest, and it is not certain that research results generated in one region can be applied in a different region with different resources and risk factors. Information on child and maternal health in low-income countries often stems from rather crude data of poor quality but is sufficient to demonstrate that the world's poorest countries carry the largest burden of diseases. They have serious health problems that impair long-term health, and they have the smallest capacity to cope with these problems.

—Ellen Aagaard Nohr and Jorn Olsen

See also Child and Adolescent Health; Fertility, Measures of; Fetal Death, Measures of; Health, Definitions of; Reproductive Epidemiology

Further Readings

Bellamy, C. (2002). Child health. In R. Detels, J. McEwen, R. Beaglehole, & H. Tanaka (Eds.), *Oxford textbook of public health* (pp. 1603–1622). Oxford, UK: Oxford University Press.

Gluckman, P. D. (2004). Living with the past. Evolution, development, and patterns of disease. *Science*, *305*, 1733–1736.

Janssen, I., Katzmarzyk, P. T., Boyce, W. F., Vereecken, C., Mulvihill, C., Roberts, C., et al. (2005). Comparison of

overweight and obesity prevalence in school-aged youth from 34 countries and their relationships with physical activity and dietary patterns. *Obesity Reviews*, *6*, 123–132.

Krickeberg, K., Kar, A., & Chakraborty, A. K. (2005). Epidemiology in developing countries. In W. Ahrens & I. Pigeot (Eds.), *Handbook of epidemiology* (pp. 545–1590). Berlin: Springer-Verlag.

Olsen, J., & Basso, O. (2005). Reproductive epidemiology. In W. Ahrens & I. Pigeot (Eds.), *Handbook of epidemiology* (pp. 1043–1110). Berlin: Springer-Verlag.

Pearce, N., & Douwes, J. (2006). The global epidemiology of asthma in children. *The International Journal of Tuberculosis and Lung Disease*, *10*, 125–132.

Wang, S., An, L., & Cochran, S. D. (2002). Women. In R. Detels, J. McEwen, R. Beaglehole, & H. Tanaka (Eds.), *Oxford textbook of public health* (pp. 1587–1602). Oxford, UK: Oxford University Press.

WHO 2005 Global Survey on Maternal and Perinatal Health Research Group. (2006). Caesarean delivery rates and pregnancy outcomes: The 2005 WHO global survey on maternal and perinatal health in Latin America. *Lancet*, *367*, 1819–1829.

World Health Organization. (2003). Diet, nutrition and chronic diseases in context. In *Diet, Nutrition and the Prevention of Chronic Diseases. Report of a Joint WHO/FAO Expert Consultation*. WHO Technical Report Series, No. 916. Geneva, Switzerland: Author. Retrieved July 26, 2006, from http://whqlibdoc.who.int/trs/WHO_TRS_916.pdf.

World Health Organization. (2005). Make every mother and child count. *The World Health Report.* Geneva, Switzerland: Author. Retrieved July 26, 2006, from http://www.who.int/whr/2005/whr2005_en.pdf.

Measles

Measles is a highly contagious viral infection, which prior to the introduction of effective vaccines was a common experience of childhood, sometimes with fatal consequences. Unfortunately, even today not all children receive the vaccine despite its efficacy and availability. In May 2003, the World Health Assembly endorsed resolution WHA56.20 urging Member countries to achieve a goal to reduce global measles deaths by half by end of 2005 compared with the 1999 estimates. Based on results from surveillance data and a natural history model, overall, global measles mortality decreased 48% from an estimated 871,000 deaths in 1999 to an estimated 454,000 deaths in 2004. Many of the recommended World Health Organization (WHO) measles control strategies now in place had been developed and first used during the early 1990s in the Americas, when the countries of the Caribbean and Latin America adopted a multi-tiered vaccination approach combining routine vaccination and mass vaccination campaigns.

Among the WHO regions, the Region of the Americas has had the most success in controlling measles. Starting in 1999, countries throughout the Region of the Americas embarked on accelerated measles elimination activities, using strategies building on the accomplishments of the polio elimination program. Implementing a measles elimination program was clearly an ambitious task, requiring the collaboration of ministries of health, the private sector, nongovernmental organizations, and multilateral and bilateral international partners. The last occurrence of widespread measles virus transmission in the Americas dates to November 2002. Sporadic cases and outbreaks have continued to occur, although 51% of the 370 measles cases reported in the Americas between January 2003 and April 2006 were positively linked to an importation.

Infectious Agent and Transmission

Measles virus is a member of the genus *Morbillivirus* of the Paramyxoviridae family. The virus appears to be antigenically stable—there is no evidence that the viral antigens have significantly changed over time. The virus is sensitive to ultraviolet light, heat, and drying.

Measles virus is transmitted primarily by respiratory droplets or airborne spray to mucous membranes in the upper respiratory tract or the conjunctiva. Man is the only natural host of measles virus. Although monkeys may become infected, transmission in the wild does not appear to be an important mechanism by which the virus persists in nature.

Measles is highly contagious and is most communicable 1 to 3 days before the onset of fever and cough. Communicability decreases rapidly after rash onset. Secondary attack rates among susceptible household contacts have been reported to be more than 80%. Due to the high transmission efficiency of measles, outbreaks have been reported in populations where only 3% to 7% of the individuals were susceptible.

Prior to the development of effective vaccines, measles occurred worldwide. Presently, in countries that have not embarked on eradication or elimination campaigns or achieved a very high level of sustained

measles immunization coverage, the disease still exists. In temperate climates, outbreaks generally occur in late winter and early spring. In tropical climates, transmission appears to increase after the rainy season. In developing countries with low vaccination coverage, epidemics often occur every 2 to 3 years and usually last between 2 and 3 months. Even countries with relatively high vaccination coverage levels may experience outbreaks when the number of susceptible children becomes large enough to sustain widespread transmission.

Epidemiology

Since the introduction of effective measles vaccines, the epidemiology of measles has changed in both developed and developing countries. As vaccine coverage has increased, there has been a marked reduction in measles incidence; and, with decreased measles virus circulation, the average age at which infection occurs has increased. Even in areas where coverage rates are high, outbreaks may still occur. Periods of low incidence may be followed by a pattern of periodic measles outbreaks, with increasing number of years between epidemics. Outbreaks are generally due to the accumulation of susceptibles, including both unvaccinated children and vaccine failures. Approximately 15% of children vaccinated at 9 months and 5% to 10% of those vaccinated at 12 months of age are not protected after vaccination. Outbreaks among older children also occur and usually involve those children who have not been vaccinated and have previously escaped natural measles infection because of the relatively low measles incidence. Since measles vaccine is less than 100% effective, some vaccinated children may also contract measles, especially during periods of intense transmission.

In large urban areas, even where measles vaccine coverage is high, the number of susceptible infants and children may still be sufficient to sustain transmission. Conditions such as high birth rates, overcrowding, and immigration of susceptible children from rural areas can facilitate transmission. Measles remains endemic in such areas, and a large proportion of cases occurs in infants before their first birthday. In endemic areas, only a brief period exists between the waning of maternal antibody and children's exposure to circulating measles virus. The highest age-specific measles case-fatality rates occur in children below 1 year of age.

Clinical Features

The incubation period is approximately 10 days (with a range of 8 to 13 days) from the time of exposure to the onset of fever and about 14 days from exposure to the appearance of the rash. Measles infection presents with a 2- to 3-day prodrome of fever, malaise, cough, and a runny nose. Conjunctivitis and bronchitis are commonly present, and the patient is highly contagious. A harsh, nonproductive cough is present throughout the febrile period, persists for 1 to 2 weeks in uncomplicated cases, and is often the last symptom to disappear. Generalized lymphadenopathy commonly occurs in young children. Older children may complain of photophobia (light sensitivity) and, occasionally, of arthralgias (joint pains). Koplik's spots, slightly raised white dots 2 to 3 mm in diameter on an erythematous base, may be seen shortly before rash onset in 80% of the cases. Initially, there are usually one to five of these lesions, but as rash onset approaches there may be as many as several hundred.

Within 2 to 4 days after the prodromal symptoms begin, a characteristic rash made up of large, blotchy red areas initially appears behind the ears and on the face. At the same time, a high fever develops. The rash peaks in 2 to 3 days and becomes most concentrated on the trunk and upper extremities. The density of the rash can vary. It may be less evident in children with dark skin. The rash typically lasts from 3 to 7 days and may be followed by a fine desquamation (shedding of the outer layers of skin). Some children develop severe exfoliation, especially if they are malnourished.

Complications

Complications from measles include otitis media, pneumonia, diarrhea, blindness, and encephalitis. It is estimated that otitis media plus pneumonia occurs in 10% to 30% of infants and young children with measles. Respiratory infections are the most common cause of significant morbidity and mortality in infants and children with measles. Pneumonia may be due to the measles virus alone or to secondary infection with other viral agents or bacterial organisms. Diarrhea is one of the major factors contributing to the adverse impact of measles on the nutritional status in children in developing countries. Measles infection is more severe among children who are already malnourished.

Neurological complications occur in 1 to 4 of every 1,000 infected children. The most common

manifestation is febrile convulsions. Encephalitis or postinfectious encephalopathy occurs in approximately 1 of every 1,000 infected children. Subacute sclerosing panencephalitis (SSPE; an infection of the nervous system) with an incidence of approximately 1 per 100,000 measles cases and may develop several years after a measles infection.

In developed countries, the case-fatality rate for measles tends to be low (between 0.1 and 1.0 per 1,000 cases). In developing countries, the overall case-fatality rate has been estimated at between 3% and 6%; the highest case-fatality rate occurs in infants 6 to 11 months of age, with malnourished infants at greatest risk. These rates may be an underestimate because of incomplete reporting of outcomes of severe measles illnesses. In certain high-risk populations, case-fatality rates have been reported to be as high as 20% or 30% in infants below 1 year of age.

Other than supportive therapies, there is currently no specific treatment for measles infection. Administration of vitamin A to children at the time of measles diagnosis has been shown to decrease both the severity of disease and the case-fatality rate. Accordingly, the WHO has recommended that vitamin A be administered to all children diagnosed with measles infection.

Immunity and Vaccination

Prior to the availability of measles vaccine, measles infection was virtually universal by 10 years of age. Infants are generally protected until 5 to 9 months of age by passively acquired maternal measles antibody. Some infants who are immunized before they are 9 months old may not develop detectable immunity because of interference by maternal measles antibody. Immunity following natural infection is believed to be lifelong, and vaccination with measles vaccine has been shown to be protective for at least 20 years.

Serologic studies have demonstrated that measles vaccines induce seroconversion in about 95% of children 12 months of age and older. Immunity conferred by a single dose vaccination against measles has been shown to persist for at least 20 years and is generally thought to be lifelong for most individuals. Studies indicate that antibody responses to the measles component when given as multiple antigens is equivalent to receiving the measles vaccine separately.

The likelihood of detecting immunoglobulin M (IgM) antibodies decreases with time. Aspirates, throat swabs, or nasopharyngeal swabs are the preferred sample for viral detection/isolation for measles viruses, but urine samples are an acceptable alternative. Data on viral genotypes are critical for tracking transmission pathways, investigating suspected vaccine-related cases, documenting the elimination of endemic strains, and supporting the hypothesis of importations from other regions.

Vaccine Schedule

Routine immunization schedules usually recommend that the first dose of measles vaccine be administered to children aged ≥ 12 months. However, if measles is present in a community, consideration may be given to lowering the age of measles vaccination to 6 months (with an additional dose at 12 months of age.) All children should have a second opportunity to receive a measles-containing vaccine. This may be provided either as a second dose in the routine immunization schedule or through periodic mass vaccination campaigns.

Vaccine Safety

The measles vaccines are generally extremely safe. Adverse events range from pain and swelling at the injection site to rare systemic reactions such as anaphylaxis. They tend to occur among people who have never been vaccinated and are very rare after revaccination. There are only two major contraindications to measles vaccination; those who have experienced an anaphylactic or severe hypersensitivity reaction to a previous dose of measles vaccine or to neomycin. In addition, pregnant women or those who have severe immunosuppressive diseases should not be vaccinated.

Vaccination Strategies

Vaccination of each successive birth cohort with a single dose of measles vaccine delivered through routine health services was a strategy originally used in many countries to control measles. Nevertheless, while vaccine coverage increased, measles outbreaks continued to occur. Since measles vaccine is less than 100% effective and coverage is rarely universal via routine health services, an accumulation of nonimmune children results over time. With each successive birth cohort, the number of children susceptible to measles increases (including both children who were never vaccinated and those who are vaccine failures). This

buildup of susceptible children over time in a population is the most serious obstacle to measles elimination.

To improve measles control, a number of countries have adopted a vaccination schedule that recommends two doses of a measles vaccine. The first dose is usually given at or after 12 months of age; the second dose is often given when children start school. For those countries with sufficient resources, a well-developed health services delivery system, and school attendance by the majority of children, this schedule reduces the number of susceptible children and ultimately interrupts measles transmission. However, the routine addition of a second dose is not an appropriate strategy for measles elimination in those countries where large segments of the population do not have access to routine health services and/or where many children do not attend school. Unfortunately, children who never received the first routine dose of measles vaccine are also those who are unlikely to receive the scheduled second routine dose.

To rectify this shortcoming, the Pan American Health Organization (PAHO) developed a three-tiered vaccination strategy. Its implementation allowed significant interruption of transmission of the measles virus in the Region of the Americas. The three main components of the PAHO vaccination strategy are as follows:

- First, measles virus circulation in a community is rapidly interrupted by conducting a one-time-only "catch-up" measles vaccination campaign over a wide age cohort of infants, children, and adolescents.
- Second, to maintain the interruption of measles virus circulation, routine immunization programs (or "keep-up" vaccination) must provide measles vaccine to at least 95% of each new birth cohort of infants before the age of 2 years in every district of the country.
- Finally, to counter the inevitable buildup of children susceptible to measles, periodic "follow-up" vaccination campaigns among preschool-aged children are carried out every 4 years. In addition to these three components, special intensive efforts, known as "mop-up" vaccination, may be required to provide measles vaccine to children living in high-risk areas who missed routine vaccination and also escaped vaccination during the "catch-up" and "follow-up" campaigns.

Surveillance and Global Eradication

A sensitive surveillance system is essential to monitor progress toward and to sustain measles elimination. In the initial stages of measles elimination efforts, the primary purpose of measles surveillance is to detect in a timely manner all areas where the measles virus is circulating, not necessarily to investigate every suspected measles case. However, once endemic transmission has become rare or has been interrupted, the surveillance goal becomes to detect and investigate all suspected measles cases, including those imported, and to implement activities that prevent or limit secondary transmission. This goal requires rapid notification and investigation of all suspected measles patients.

Both the successful smallpox eradication program and the efforts to control polio suggest that achieving measles eradication depends on several factors: the biological characteristics of the organism, vaccine technology, surveillance and laboratory identification, effective delivery of vaccination programs, and international commitment. Clearly, experience in the Americas has shown that these factors favor a measles eradication effort. There is also growing international support for such an initiative both from governmental and donor agencies.

—Marc Strassburg

Note: The author worked as a consultant for the PAHO measles elimination program, and in that capacity assisted in writing a number of earlier versions of the measles elimination field guide. Sections from both previous and current field guides were liberally adapted for this article.

See also Child and Adolescent Health; Disease Eradication; Public Health Surveillance; Vaccination

Further Readings

Clements, J. C., Strassburg, M. A., Cutts, Ft., Milstien, J. B., & Torel, C. (1992). Global control of measles: New strategies for the 1990s. In E. Kurstak (Ed.), *Control of virus diseases* (2nd ed., pp. 179–210). New York: Marcel Dekker.

Clements, J. C., Strassburg, M. A., Cutts, Ft., Milstien, J. B., & Torel, C. (1993). Challenges for global control of measles in the 1990's. In E. Kurstak (Ed.), *Measles and poliomyelitis* (pp. 13–24). Vienna, Austria: Springer-Verlag.

Clements, J. G., Strassburg, M., Cutts, Ft., & Torel, C. (1992). The epidemiology of measles. *World Health Statistics Quarterly*, *45*, 285–291.

de Quadros, C., Olive, J. M., Hersh, B., Strassburg, M. A., Henderson, D. A., Brandling-Bennet, D., et al. (1996).

Measles elimination in the Americas: Evolving strategies. *Journal of the American Medical Association*, *275*, 224–229.

Drotman, D. P., & Strassburg, M. A. (2001). Sources of data. In J. C. Thomas & D. J. Weber (Eds.), *Epidemiologic methods for the study of infectious diseases* (pp. 119–137). New York: Oxford University Press.

Galindo, M. A., Santin, M., Resik, S., Ribas, M. L. A., Guzman, M., Lago, P. M., et al. (1998). The eradication of measles from Cuba. *Pan American Journal of Public Health*, *4*(3), 171–177.

Morgan, O. W. C. (2004). Following in the footsteps of smallpox: Can we achieve the global eradication of measles? [Electronic version]. *BMC International Health and Human Rights*, *4*(1), doi: 10.1186/1472-698X-4-1. Retrieved February 28, 2007, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=387835.

Pan American Health Organization. (2005). *Measles elimination: Field guide*. Washington, DC: Author. (Scientific and Technical Publication No. 605). Retrieved February 28, 2007, from http://www.paho.org/english/ad/fch/im/fieldguide_measles.pdf.

Pan American Health Organization. (2006). *Immunization Newsletter*, *28*(4) [Electronic version]. Retrieved February 28, 2007, from http://www.paho.org/English/AD/FCH/IM/sne2804.pdf.

Weekly Epidemiological Record. (2006, March 10). *81*(10), 89–96.

Measurement

Scientists from numerous disciplines frequently make sense of the world by using yardsticks that they hope will show how their study participants are performing, what they are thinking, and how they interact with others. Numbers are faithfully recorded, spun through various forms of software, and prepared for publication. All this is fine if the yardsticks themselves are true—all the time, in every single place they are used, regardless of who is doing the actual recording of the numbers, and regardless of the circumstances in which the numbers are obtained. But what if the yardsticks themselves are shaky?

In epidemiologic analyses based strictly on counting, a few units in dispute here or there may seem rather unlikely to significantly change the overall interpretation of the data set. However, even a single reclassification of a case from one cell to another, for instance, can force a confidence interval to bracket 1.0 where it otherwise might not do so, or a statistical test to just miss threshold. Although it might not appear as an issue when the data are highly differentiable, the need for quality measurement is fundamentally inescapable. Equally important, whole sections of the field of epidemiology have long ago been unbound from the simple exercise of counting, working instead in arenas in which measurements take the form of scores, scales, and other assessments. In such settings, the challenges to designing and conducting strong and reproducible studies are magnified.

The domain of psychometrics gives criteria concerning the quality of a measurement. Although "psychometrics" has been a label narrowly applied to a particular specialized branch of mathematical and statistical thinking within educational research, this entry uses the term in a broader sense. It explores a handful of concepts that are crucial to all measurements and considers recent examples in the epidemiologic literature that show the importance of such considerations.

Reliability

Psychometricians have been preaching for decades that the core considerations of good measurement must not be simply assumed whenever a set of assessments is made. Principal among these considerations is that the measurements be reliable and valid. Reliability is defined as the consistency of measurements made by a specific group of persons using the measurement instrument, working under the same conditions. In the most elementary sense, high reliability means that data will be consistent if the identical study is run again. Even in closely monitored laboratory conditions, however, there are numerous possible contaminants that can interfere with obtaining reliable data. To reduce error and improve reliability, laboratories make constant use of standardizing and correcting baseline values for all their measurement devices. Likewise, measurements made in the field need comparable standardizations: One often-used method for standardization is to be sure that different field workers show high levels of agreement when facing the same situations for data collection. A simple but informative analysis is to evaluate overall agreement between workers using varying tolerances: A tolerance of zero (equivalent to exact agreement) results in a overall percentage between 0 and 100, then (if not 100%) tolerances are widened step by step (i.e, liberalizing the definition of agreement) until 100% agreement is achieved. (Software to accomplish this task is

available in the R package.*) The climb toward full agreement as tolerances are made less restrictive is a direct reflection of the reliability of the sources of data.

Three other classical methods to assess reliability are test-retest, multiple forms, and split-half assessments. In test-retest, the specific test instrument is used by the same workers at different times. The correlation coefficients between the scores achieved at the differing times serve as coefficients of reliability. In multiple forms, the testing sequence is systematically varied, given to either the same workers twice or to two or more different groups of workers. Split-half reliability is estimated by analyzing half the results from the test instrument in comparison with the results of the overall analysis. Both Cronbach's alpha and the Kuder-Richardson coefficient called KR-20 provide readily interpretable statistics describing the amount of reliability in the measurements.

A measurement tool that does not yield reliable scores leads to the possibility that every subsequent interpretation will be suspect. If measurements are unreliable, there are few good ways of disentangling how much the variation within those measurements is explainable and how much is due simply to error. Even when excellent research designs and high-level statistical procedures are employed, in general whatever real effects are present will be underestimated.

An example of the importance of understanding test reliability is seen in Schiffman and Schatzkin's (1994) analysis of two earlier studies in molecular epidemiology conducted by their research team. In analyses of the relationship between human papillomavirus (HPV) infection and cervical neoplasia, the issue in brief was whether molecular assays produced consistent results across many clinical specimens collected over a period of months or years. Two different case-control studies had been conducted several years apart to investigate the presence of HPV and cervical intraepithelial neoplasia. While both used the same case definitions, HPV testing underwent significant transformation during the intervening years and the resulting assays differed. In the first study, comparisons of results between laboratories found poor agreement, but in the second study, data were far less often misclassified. The association between HPV and neoplasia was an order of magnitude greater in the second study. Indeed, the conclusions of the two studies differed dramatically—the first pointed to HPV infection as a risk factor but not the key etiologic agent, while in the second, HPV was found to have a central, causal role. The authors concluded that measurement error can be a common problem in studies that rely on highly technical assays and can lead directly to wrong conclusions.

Validity

The term *validity* refers to a number of different concerns about measurements. Internal validity is an indication of how measurements perform in settings or with cases that are similar to those for which the measurement was first developed. A close synonym for internal validity is reproducibility. External validity is an indication of how the measurements perform in new settings or in cases with characteristics different from the original. A close synonym for external validity is generalizability. Face validity is obtained by having experts vet the measurement in question to assure that the measurement appears to reflect ground truth. Content validity can be ascertained by study of the relationships between different measurement dimensions. Criterion validity can be assessed by benchmarking one measurement against a gold standard; additionally, criterion validity can be separated into the success with which the measurements estimate concurrent events and the success with which they predict future events. Construct validity can be appraised by how well the measurement tool matches the underlying theory or model of what is being measured; additionally, construct validity can be separated into the degree to which the elements within the measurement scales converge on the same result, and the degree to which the measurements succeed in discriminating between cases that diverge from one another by greater or lesser amounts.

An important consideration is that a given measurement tool can have high reliability but poor validity. This is not unlike achieving great success with a bow and arrow, hitting the same target repeatedly, but then discovering that the target is not the one we had been aiming at. On the other hand, a given

*R is an international collaborative open-source software product for data manipulation, calculation, and graphical display—available at no charge—that is highly extensible, integrating contributions from numerous statistical professionals by means of packages of code and documentation built to strict criteria. R is a product of the R Foundation for Statistical Computing, Vienna, Austria. Extensive information is available at http://www.r-project.org.

measurement tool cannot have high validity without also being reliable. That is, a broad sweep of arrow volleyed toward the vague area of the target will not lead to a well-focused series of strikes. Many authors on this topic have discussed threats to validity and reliability and the essential need to assure optimal research designs to standardize measurement and reduce error.

An example of a noteworthy analysis of validity is seen in Hukkelhoven et al.'s (2006) detailed comparison of prediction success using competing models of outcome after traumatic brain injury. The authors identified 10 competing prognostic models in the recent literature. Each had been developed from careful study of samples of brain-injured patients assessed by numerous measurement tools, with the final product in every instance being touted as the best combination of indicators of outcome. Hukkelhoven et al. systematically applied each of these models to validation populations composed of 4,238 brain-injured patients from published sources. The success or failure of each model was examined in terms of discrimination and calibration, two terms that are underpinned by strong statistical methods. Discrimination refers to a given model's capacity to distinguish between patients who have different outcomes and can be immediately determined from the receiver operating characteristic (ROC) curve (see R package "ROCR"). Calibration refers to the degree to which a model's estimates match reality by producing unbiased estimates and can be tested by goodness-of-fit statistics.

Hukkelhoven et al. (2006) found that the selected models demonstrated substantial variability in discrimination: The range was from 0.61 to 0.89 (where perfect discrimination would be 1.0 and no discrimination would be 0.50). Additionally, the same models varied in discrimination depending on which validation population was being considered. Calibrations for four of the six competing prognostic models were poor, with the direct implication that predicted mortality was too high compared with actual mortality. For example, one model suggested that one of the validation populations should have a mortality of 60% when in fact the observed mortality was merely 35%. Calibration curves were often nonlinear, signifying that some of the models might be relatively more correct for some cases but not for others. Reasons for the diminished success of prognostic models undergoing the process of external validation can include small original sample sizes that limit statistical power and precision, insufficient numbers of predictors, and differences in study populations and therapeutic approaches.

Item Response Theory

An elemental point of reference in psychometrics has long been that a given individual can be assigned a score that truly (or, perhaps, adequately) represents that person's condition, capacity, or characteristic. Entire generations of psychometric analysts were imbued with rules for their work that stemmed from enumerating pupil skills in settings involving educational or psychological testing. Classical test theory, however, has become increasingly supplanted by other techniques because the underlying assumptions were found to be either unrealistic or too restrictive for broader applications. A spectrum of analytic tools is now available for performing detailed psychometric statistics across many different professional domains. Item response theory (IRT) is one of the leading sources of these new rules for measurement.

IRT is a collection of statistical methods and models that rely on the probabilistic relationship between a person's response to test items and that person's position on an underlying construct on which the test is attempting to focus. Within IRT, two key assumptions are made. First is that whatever model is used is adequate to explain how changes in the level of the characteristic being measured related to changes in the probabilities of responding to items. Second is that the terms included in the model fully explain the interrelationships between the persons being tested and the items used for testing. From these two assumptions, it follows that there might be a proliferation of models. Indeed there are at least a hundred separate IRT models, all of which are mathematical expressions of how unidimensional or multidimensional data should fit together to reflect the underlying construct being measured. They differ primarily in how many mathematical terms are estimated and how many constraints are placed on the estimation process.

Embretson and Reise (2000) point out that IRT draws on analogies to clinical inference, asking how plausible a certain diagnosis might be in the face of selected behaviors and test responses. How likely is that diagnosis to explain the presenting behaviors? What level of the measured characteristic is most likely to explain the person's test responses? In IRT-based analyses, identifying that measurement level is

a matter of seeking the highest likelihood for the responses observed. To find the most likely level or score, the likelihood of the person's actual response pattern is evaluated within the mathematical model. The likelihood calculation allows evaluation of any point along a hypothetical line that represents the full range of the condition or behavior.

To explain response patterns in the simplest of terms, imagine a short test constructed out of only five items, which can only either be scored true or false, correct or wrong, present or absent. If the test items are lined up properly in terms that reflect their inherent difficulty—that is, they are tied to the construct being tested in an orderly manner, then only a limited series of response patterns from persons taking this test make sense. One acceptable pattern is that a given person misses every item (and the resulting vectors of scores reads "00000"). That pattern ought to signify that the person is at the bottom of the test construct. Equally acceptable is a vector that reads "11111," signifying a person at the top of the construct. In theory, we can easily make sense of "10000," "11000," "11100," and "11110" as well-ordered score vectors. If the difficulty steps between items are equal—the difference between each item reflects the same difference no matter which side-by-side pair is evaluated—then the individual performances shown by these vectors are themselves readily interpretable.

Matters get more interesting when score vectors are unexpected (such as "10001"—succeeding on both the easiest and most difficulty items but failing on the others) and the underlying interpretability of the person's performance on the test is thrown into some doubt. Probabilistically, such a score vector should be quite rare, if everything else about the test is well constructed. IRT allows explicit methods to sort out just how interpretable a given person's performance is, and where the test items themselves may be poorly functioning. Work on understanding the nature of disorderly response patterns harks back to the delightful phrase "higgledy-piggledy," which was used to label this phenomenon in the earliest attempts to formally describe systematic testing.

Formal evaluation of the score vector for every respondent is made through IRT software analyses that address the simultaneous computation of likelihoods for each response, each person, and each test item. Trait levels are developed by maximizing the likelihood of a person's response pattern in the context of the particular IRT model employed. Most common are models that invoke only a single parameter reflecting item difficulty, or an additional parameter keyed to how each item differs in its discrimination between low-performing and high-performing responses, or yet another parameter reflecting the role of success by guessing or chance. The one-parameter IRT approach is known as the Rasch model, after the Danish professor Georg Rasch who discovered its properties. Rasch models are directly able to estimate reliability and provide detailed information about individual person performance and test score error. Several IRT software products are available, including TESTFACT, BILOG, MULTILOG, RUMM, WINSTEPS, and the R package "ltm."

Structural Equation Modeling

A different approach to understanding the relationship between outcome measures and a set of indicators or constructs is contained in the analytic technique called structural equation modeling (SEM). Like IRT, the probability of a certain indicator being positive for a certain outcome measures is directly assessable in SEM. Unlike IRT, additional analysis can be made of relations among factors and between factors and other variables that may be plausible covariates. One way in which the SEM approach is used is the creation of uni- and bidirectional paths that describe regression relationships and correlations, respectively, among the variable sets.

An SEM approach to understanding how cerebral white matter abnormalities relate to cognitive functioning in elderly persons is shown in a recent study by Deary, Leaper, Murray, Staff, and Whalley (2003). Questions about this association included whether it was independent of mental ability during youth and whether it was related to general and/or specific mental abilities. Ratings were made of periventricular and subcortical and deep white matter abnormalities seen on magnetic resonance images taken of each participant. SEM techniques found that white matter abnormalities accounted for 14% of the variance found in cognitive function in old age.

SEM can be extended to encompass factor analyses, modeling of multiple indicators and multiple causes, and analyses involving complex longitudinal data. A variety of SEM software products are available, including LISREL, EQS, M-PLUS, and AMOS, as well as an R package called "sem."

It is entirely possible that some measurements can be demonstrated to be both reliable and valid yet will never conform to IRT or SEM specifications. Both IRT and SEM also risk having computational complexity grow to be enormous as sample sizes and test batteries are enlarged. However, both have distinctive mathematical underpinnings that can be used to examine features of research that are otherwise exceedingly difficult to analyze.

Robustness

Understanding how data sets fit models, and how model inferences can be affected by both misspecifications in the model and particular points within the data that have high influence, is the goal of robustness analyses. The principal question pursued in such analyses is how outliers are identified and how might such outliers affect the ultimate interpretation of a study. Two different directions have been actively pursued: The first is to use analytic methods that are themselves robust in the sense that underlying assumptions of normal distributions and common variances found in traditional statistics are entirely supplanted by far more powerful techniques. Wilcox (2005) explores the foundation for robust statistical analysis and provides systematic solutions (including R code) for estimating robust measures of location and scale, developing robust confidence intervals, and working with robust solutions to correlation and regression problems. The second direction is the development of tools for sensitivity analyses, which rely on changing various model assumptions and parameters and checking whether relatively large changes have negligible effects on calculated quantities. This approach also allows especially thorny analytic problems, such as the failure of estimates in logistic regression to converge, to be addressed systematically (see R package "accuracy").

Developments over the last several decades in psychometrics and statistics have been extraordinary in terms of the potency with which such improvements can be made. We have not sought to impugn any extant epidemiologic study for failure to address its quality of measurement. But, as the reader surely has sensed, we are making a case for assessing data quality (and analysis quality) on a regular basis. Anytime we allow imprecise measurements to be included in a study, we have reduced the quality of the science itself.

—David L. McArthur

See also Quantitative Methods in Epidemiology; Reliability; Validity

Further Readings

Altman, M., Gill, J., & McDonald, M. P. (2003). *Numerical issues in statistical computing for the social scientist*. Hoboken, NJ: Wiley-Interscience.

Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *American Journal of Medicine*, *119*(2), 166.e7–166.e16.

Deary, I. J., Leaper, S. A., Murray, A. D., Staff, R. T., & Whalley, L. J. (2003). Cerebral white matter abnormalities and lifetime cognitive change: A 67-year follow-up of the Scottish Mental Survey of 1932. *Psychology and Aging*, *18*, 140–148.

Embretson, S. W., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Hukkelhoven, C. W. P. M, Rampen, A. J. J, Maas, A. I. R., Farace, E., Habbema, J. D. F, Marmarou, A., et al. (2006). Some prognostic models for traumatic brain injury were not valid. *Journal of Clinical Epidemiology*, *59*, 132–143.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Schiffman, M. H., & Schatzkin, A. (1994). Test reliability is critically important to molecular epidemiology: An example from studies of human papillomavirus infection and cervical neoplasia. *Cancer Research*, *54*(Suppl.), 1944s–1947s.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed). Amsterdam: Elsevier.

Measures of Association

Measures of association encompass methods designed to identify relationships between two or more variables and statistics used to measure the relationship when it exists. Although the terms *correlation* and *association* are often used interchangeably, correlation in a stricter sense refers to linear correlation and association refers to any relationship between variables, including the relationship between two categorical variables.

Choosing the Correct Method

Choosing the correct method to measure association involves a determination of the data characteristics for each variable. Data may be measured on an interval/ratio scale, an ordinal/rank scale, or a nominal/categorical scale. These three characteristics can be thought of as continuous, integer, and qualitative categories.

Pearson's Correlation Coefficient

A typical example for measuring the association between two variables measured on an interval/ratio scale is the analysis of relationship between a person's height and weight. Each of these two characteristic variables is measured on a continuous scale. The appropriate measure of association for this situation is the Pearson's correlation coefficient.

The Pearson's correlation coefficient, ρ (rho), measures the strength of the linear relationship between the two variables measured on a continuous scale. The coefficient ρ takes on the values of -1 through $+1$. Values of -1 or $+1$ indicate a perfect linear relationship between the two variables, whereas a value of 0 indicates no linear relationship. Correlation coefficients that differ from 0 but are not $+1$ or -1 indicate a linear relationship, although not a perfect linear relationship. Negative values simply indicate the direction of the association: As one variable increases, the other decreases. In practice, ρ (the population correlation coefficient) is estimated by r, the correlation coefficient derived from sample data.

Although the Pearson's correlation coefficient is a measure of the *strength* of an association (specifically the linear relationship), it is not a measure of the *significance* of the association. The significance of the association is a separate analysis of the sample correlation coefficient, r, using a t test to measure the difference between the observed r and the expected r under the null hypothesis.

Spearman Rank-Order Correlation Coefficient

The Spearman rank-order correlation coefficient (Spearman rho) is designed to measure the strength of a monotonic (in a constant direction) association between two variables measured on an ordinal or ranked scale. Examples that indicate the Spearman rho should be used to include data obtained on preferences where the data result from ranking. It is also appropriate for data collected on a scale that is not truly interval in nature, such as data obtained from Likert-scale administration. Any interval data may be transformed to ranks and analyzed with the Spearman rho, although this results in a loss of information; for instance, this may be done if one variable of interest is measured on an interval scale and the other is measured on an ordinal scale. Like the Pearson's correlation coefficient, the Spearman rho may be tested for its significance. A similar measure of strength of association is the Kendall tau, which may also be applied to measure the strength of a monotonic association between two variables measured on an ordinal or rank scale.

As an example of when Spearman rho would be appropriate, consider the case where there are seven substantial health threats to a community. Health officials wish to determine a hierarchy of threats in order to most efficiently deploy their resources. They ask two credible epidemiologists to rank the seven threats from 1 to 7, where 1 is the most significant threat. The Spearman rho or Kendall tau may be calculated to measure the degree of association between the epidemiologists indicating the collective strength of the action plan. If there is a significant association between the two sets of ranks, health officials will feel more confident in their strategy than if a significant association is not evident.

Chi-Square Test

The chi-square test for association (contingency) is a standard measure for association between two categorical variables. The chi-square test, unlike the Pearson's correlation coefficient or the Spearman rho, is a measure of the significance of the association rather than a measure of the strength of the association.

A simple and generic example follows. If a scientist was studying the relationship between gender and political party, then he could count people from a random sample belonging to the various combinations: female-Democrat, female-Republican, male-Democrat, and male-Republican. He could then perform a chi-square test to determine whether there was a significant disproportionate membership among these groups indicating an association between gender and political party.

Relative Risk and Odds Ratio

Several other measures of association between categorical variables are used in epidemiology, including

the relative risk and odds ratio. The relative risk is appropriately applied to categorical data derived from an epidemiologic cohort study. The relative risk measures the strength of an association by considering the incidence of an event in an identifiable group (numerator) and comparing that with the incidence in a baseline group (denominator).

A relative risk of 1 indicates no association; a relative risk other than 1 indicates an association. For example, if 10 of 1,000 people exposed to *X* developed liver cancer, but only 2 of 1,000 people (who were never exposed to *X*) developed liver cancer, then we can say the relative risk is $(10/1000)/(2/1000) = 5$. The strength of the association is 5: People exposed to *X* are five times more likely to develop liver cancer than others. If the relative risk was < 1, perhaps 0.2, then the strength of the association is equally evident but with another explanation: Exposure to *X* reduces the likelihood of liver cancer fivefold—a protective effect. The categorical variables are exposure to *X* (yes or no) and the outcome of liver cancer (yes or no). Of course, this calculation of the relative risk does not test whether the relative risk $= 5$ is statistically significant or not. Questions of significance may be answered by calculation of a 95% confidence interval: If the confidence interval does not include 1, the relationship is considered significant.

Similarly, an odds ratio is an appropriate measure of strength of association for categorical data derived from a case-control study. The odds ratio is often interpreted the same way that a relative risk is interpreted when measuring the strength of the association, although this is somewhat controversial when the risk factor being studied is common.

Additional Methods

There are a number of other measures of association for a variety of circumstances. For example, if one variable is measured on an interval/ratio scale and the second variable is dichotomous, then the point-biserial correlation coefficient is appropriate. Other combinations of data types (or transformed data types) may require the use of more specialized methods to measure the association in strength and significance.

Other types of association describe the way data are related but are usually not investigated for their own interest. Serial correlation (also known as autocorrelation), for instance, describes how in a series of events occurring over a period of time, events that occur closely spaced in time tend to be more similar than those more widely spaced. The Durbin-Watson test is a procedure to test the significance of these correlations. If these correlations are evident, then we may conclude that these data violate the assumptions of independence, rendering many modeling procedures invalid. A classical example of this problem occurs when data are collected over time for one particular characteristic. For example, if an epidemiologist wanted to develop a simple linear regression for the number of infections by month, there would undoubtedly be serial correlation: Each month's observation would depend on the prior month's observation. This serial effect (serial correlation) would violate the assumption of independent observations for simple linear regression and accordingly render the parameter estimates for simple linear regression as not credible.

Inferring Causality

Perhaps the greatest danger with all measures of association is the temptation to infer causality. Whenever one variable causes changes in another variable, an association will exist. But whenever an association exists, it does not always follow that causation exists. The ability to infer causation from an association in epidemiology is often weak because many studies are observational and subject to various alternative explanations for their results. Even when randomization has been applied, as in clinical trials, inference of causation is often limited.

—Mark Gerard Haug

See also Causation and Causal Inference; Chi-Square Test; Hill's Considerations for Causal Inference; Pearson Correlation Coefficient

Further Readings

Freedman, D. A. (2005). *Statistical models: Theory and practice.* Cambridge, UK: Cambridge University Press.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). New York: McGraw-Hill.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied regression analysis: A research tool* (2nd ed.). New York: Springer.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: CRC Press.

Tietjen, G. L. (1986). *A topical dictionary of statistics.* Boca Raton, FL: Chapman & Hall.

Measures of Central Tendency

Measures of central tendency provide a single summary number that captures the general location of a set of data points. This measure should be a good representation of the set of data. There are three common measures of central tendency used: the mean, the median, and the mode. Depending on the characteristics of the data, one measure may be more appropriate to use than the others.

The mean and the median are most commonly used to summarize data that can take on many different values (i.e., continuous data); the mode is often used to summarize data that can only take on a finite number of specific values (i.e., categorical data). The construction of each measure is illustrated with the gender and height data in Table 1 collected from 22 subjects.

The Mean

The mean is the most commonly used measure of central tendency. It is often referred to as *x-bar*, $\overline{x}$, and is found by using the formula,

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where

x_i represents the individual observation from the ith subject;

$\sum$ is the summation sign that indicates that you sum over everything that follows it. The limits below and above the sign indicate where you start, and stop, the summation, respectively. As it is written above, it says you should begin summing with x_1 and stop with x_n; and

n represents the number of observations in your data set.

For illustrative purposes, consider the data in Table 1. To compute the mean height, we do the following:

1. Sum over all the observations (the x_i values).
2. Divide the quantity in Step 1 by the total number of observations.

Table 1 Heights of 22 Students in an Introductory Statistics Class

Observation	*Gender*	*Height in Inches*
x_1	Female	61
x_2	Female	62
x_3	Female	63
x_4	Female	63
x_5	Female	64
x_6	Female	64.5
x_7	Female	65
x_8	Female	65
x_9	Female	65
x_{10}	Female	65
x_{11}	Female	66
x_{12}	Female	66
x_{13}	Female	66
x_{14}	Male	67
x_{15}	Female	67
x_{16}	Male	67
x_{17}	Male	68
x_{18}	Female	68
x_{19}	Male	69
x_{20}	Female	69.5
x_{21}	Male	72
x_{22}	Male	74

For the data set above, the above formula gives

$$\overline{x} = \frac{1}{22}\sum_{i=1}^{22} x_i = \frac{1}{22}(1457) = 66.2 \text{ in.}$$

Additional Notes About the Mean

- The small n represents the sample size when computing the mean for a sample. When computing the mean for a population, a large N is generally used.
- The sample mean represents the population mean better than any other measure of central tendency.

- The mean can be thought of as being like a fulcrum that balances the weight of the data.
- The sum of deviations of each observation from the mean is 0.
- The mean is in the same units of measurement as the observations in your data set.
- The mean is very sensitive to outliers, that is, extreme values. An observation that lies far away from the others will pull the mean toward it. For example, if the last observation in Table 1 were 740 instead of 74, the mean would jump to 96.5 (an increase of more than 30 in.).
- The mean is generally the preferred measure of central tendency for data that are symmetric (evenly distributed about their center), but is not generally recommended to be used to describe data with outliers, or data that are not symmetric.

The Median

The median is used less often than the mean to describe the central tendency of a set of continuous data, but is still a commonly used measure. It is the midpoint of the data, and is often denoted with the letter M. To find the median, the following steps are taken:

1. Sort the observations from smallest to largest (sorting from largest to smallest is also valid).
2. Choose the correct step below depending on the number of observations in the data set:
   a. If you have an odd number of observations, observation number $(n+1)/2$ is the median.
   b. If you have an even number of observations, the average of observation number $n/2$ and observation number $(n/2)+1$ is the median.

Consider the data in Table 1. The observations have already been sorted in ascending order according to height. Since there is an even number of observations (22) in the data set, the median is the average of the 11th and 12th observations. Thus,

$$M = \frac{x_{11} + x_{12}}{2} = \frac{132}{2} = 66 \text{ in.}$$

Note that if the 22nd observation (x_{22}) were not in the data set, the median would be the 11th observation ($M = 66$ in.).

Additional Notes About the Median

- The median is often called the 50th percentile since it marks the midpoint of the data. That is, half the observations are less than the median and half the observations are greater than the median.
- The median is in the same units of measurement as the observations in your data set.
- The median is not sensitive to outliers unlike the mean. For example, if the last observation in the table above were 740 instead of 74, the median would still be 66 in.
- The median is the recommended measure of central tendency for data that have outliers or that are not symmetric.

The Mode

The mode is most commonly used for data that are categorical (data that can only take on one of a set of distinct values). It is defined as the most frequently occurring value. In Table 1, the mode of the gender variable is "female" since there are 16 females, but only 6 males. The mode can also be used to summarize continuous data, although this is less common. The height variable in Table 1 has a mode of 65 in. since it occurs more often than any other height. The mode is in the measurement units of the variable it is summarizing.

—Liam M. O'Brien

See also Box-and-Whisker Plot; Histogram; Measures of Variability; Percentiles

Further Readings

Rosner, B. (2006). *Fundamentals of biostatistics* (6th ed.). Belmont, CA: Duxbury Press.

Measures of Variability

Numerical summaries used to describe a set of data generally include a measure of central tendency. While this provides a single estimate that describes where the data are located, it does not describe how spread out the data are about this central point. There are several numerical summaries that describe the variability in a data set. Four of the most common are the variance,

the standard deviation, the interquartile range (IQR), and the range. They are illustrated in Table 1 using data collected on height from 22 subjects.

Table 1 Measures of Variability for the Heights of 22 Students in an Introductory Statistics Class

Observation	*Height in Inches*	*Deviation From Mean*	*Deviation From Mean Squared*
x_1	61.0	–5.2	27.04
x_2	62.0	–4.2	17.64
x_3	63.0	–3.2	10.24
x_4	63.0	–3.2	10.24
x_5	64.0	–2.2	4.84
x_6	64.5	–1.7	2.89
x_7	65.0	–1.2	1.44
x_8	65.0	–1.2	1.44
x_9	65.0	–1.2	1.44
x_{10}	65.0	–1.2	1.44
x_{11}	66.0	–0.2	0.04
x_{12}	66.0	–0.2	0.04
x_{13}	66.0	–0.2	0.04
x_{14}	67.0	0.8	0.64
x_{15}	67.0	0.8	0.64
x_{16}	67.0	0.8	0.64
x_{17}	68.0	1.8	3.24
x_{18}	68.0	1.8	3.24
x_{19}	69.0	2.8	7.84
x_{20}	69.5	3.3	10.89
x_{21}	72.0	5.8	33.64
x_{22}	74.0	7.8	60.84

The Variance

The variance is approximately equal to the average squared distance of each observation about the mean and is generally denoted by s^2. This is most easily seen in its formula, which is given by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

where

x_i represents the individual observation from the ith subject;

the mean of the data is given by $\overline{x}$;

$\sum$ is the summation sign, which indicates that you sum over everything that follows it. The limits below and above the sign indicate where you start, and stop, the summation, respectively. As it is written above, it says you should begin summing with the squared deviation of x_1 from the mean and stop with the squared deviation of x_n from the mean; and

n represents the number of observations in your data set.

For illustrative purposes, consider the data in Table 1 above. To compute the variance of the height measurements, do the following:

1. Calculate the mean height $\overline{x}$.
2. For each observation, calculate the deviation from the mean $x_i - \overline{x}$.
3. Square the deviation of each observation from the mean $(x_i - \overline{x})^2$.
4. Sum over all the squared deviations.
5. Divide this sum by $n - 1$, where n is the sample size.

Table 1 gives the quantities described in Steps 2 and 3 above for each observation. For this set of data the procedure above gives

$$\begin{aligned} s^2 &= \frac{1}{(22-1)}\sum_{i=1}^{22}(x_i - \overline{x})^2 \\ &= \frac{1}{21}\sum_{i=1}^{22}(x_i - 66.2)^2 = \frac{200.38}{21} \\ &= 9.54 \text{ in.}^2 \end{aligned}$$

Additional Notes About the Variance

- The small n represents the sample size when computing the mean for a sample. When computing the mean for a population, you divide the sum of the squared deviations from the population mean and divide that sum by the population size N. This is

done only if you have data from a census, that is, when you collected data from every member of a population.
- The units of the variance are the *square* of the measurement units of the original data.
- The variance is very sensitive to outliers. It is generally not the recommended measure of variability to use if there are outliers in the data or if the data are not symmetric.

The Standard Deviation

The standard deviation is the most commonly used measure of variability for data that follow a bell-shaped (or normal) distribution. It is generally denoted by s, and is simply the square root of the variance. Formally, it is given by the formula

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{22}(x_i - \overline{x})^2}$$

where the quantities in the formula are defined in the same way as described above for the variance. If we consider the data in Table 1, we can calculate the standard deviation easily using the value that we calculated for the variance:

$$s = \sqrt{s^2} = \sqrt{9.54 \text{ in.}^2} = 3.09 \text{ in.}$$

Additional Notes About the Standard Deviation

- The standard deviation has the same measurement units as the original data.
- The standard deviation is sensitive to outliers just like the variance.
- The standard deviation is not recommended as a measure of variability if there are outliers, or if the data are not symmetric.
- The standard deviation is the recommended measure of variability for data that are symmetric. It is generally used to describe the variability when the mean is used to describe central tendency.
- If the data are bell shaped, then the ''empirical rule'' states that
- approximately 67% of the data fall within 1 *SD* of the mean,
- approximately 95% of the data fall within 2 *SD* of the mean, and
- approximately 99.7% of the data fall within 3 *SD* of the mean.

The Interquartile Range

The IQR is a measure that describes the range of the middle half of the data. It is found by locating the points in the data set that mark the 25th and 75th percentiles. The IQR is then the 75th percentile minus the 25th percentile. This can be found by performing the following steps:

1. Sort your data from the smallest observation to the largest observation.
2. Divide your data into two equally sized halves. If you have an odd number of observations, remove the midpoint (i.e., the median) and divide the remaining data into halves.
3. The median of the lower half of the data marks the 25th percentile.
4. The median of the upper half of the data marks the 75th percentile.
5. Subtract the 25th percentile from the 75th percentile.

The resulting quantity is the IQR. For the data in Table 1, the lower half of the data consists of observations 1 through 11, and the upper half consists of observations 12 through 22. The IQR is found by

$$\begin{gathered}\text{25th percentile} = x_6 = 64.5 \text{ in.} \\ \text{75th percentile} = x_{17} = 68.0 \text{ in.} \\ \text{IQR} = 68.0 - 64.5 = 3.5 \text{ in.}\end{gathered}$$

Additional Notes About the IQR

- The IQR has the same measurement units as the original data.
- The IQR is a *single number* that describes the range of the middle half of the data.
- The IQR is not sensitive to outliers or data that are not symmetric.
- It is the preferred measure of variability for data that have outliers or that are not symmetric. It is used when the median is the preferred measure of central tendency.

The Range

The range is the simplest measure of variability to calculate. It is the largest observation minus the smallest observation. For the data in Table 1, the range is given by

$$x_{22} - x_1 = 74 - 61 = 13 \text{ in.}$$

Additional Notes About the Range

- The range is very sensitive to outliers. Due to its extreme sensitivity, it is the least commonly used of the four measures described here.
- The range is in the same measurement units as the original data.

—Liam M. O'Brien

See also Box-and-Whisker Plot; Histogram; Measures of Central Tendency; Percentiles

Further Readings

Rosner, B. (2006). *Fundamentals of biostatistics* (6th ed.). Belmont, CA: Duxbury Press.

Mediating Variable

A mediator is a variable that explains, totally or partially, the relationship between a predictor and an outcome. In other words, a mediating variable is a mechanism through which a predictor exerts its effect on an outcome variable. Mediation is important in epidemiology because health events are rarely due to direct causes. For instance, low socioeconomic condition increases the risk of low birthweight (LBW) through a complex mechanism mediated by food supply to the pregnant woman and weight gain during pregnancy. If the predictor is an intervention, identifying mediating variables is essential to understand how some actions produce certain outcomes.

Given any two correlated variables, X and Y, and no outside theoretical information, it is impossible to say whether changes in X cause changes in Y, changes in Y cause changes in X, some third variable, Z, produces changes in both X and Y, or any combination of these alternatives. This is, of course, an oversimplification of the analysis, because in practice the problem is never confined to just three variables.

Confounding, Mediation, and Effect Modification

Most often the influence of one variable on another is affected by the confounding, mediating, or modifying effect of a third one. It is important to clearly distinguish between these three concepts.

Given D (disease) and E (exposure), E may cause D, but it can also be related to D if both are caused by factor F. This case is illustrated in Figure 1, in which poverty has been depicted as a *confounder*. If we had concluded that migration causes disease, when, in fact, they have no true causal relationship, we would say that the relationship between migration and disease is confounded by poverty. People migrate because they are poor, and for the same reason they have higher rates of disease. Migration by itself does not cause disease.

However, not every factor associated with both the exposure and the disease is a confounder. It may also be a *mediating variable*. Mediating variables are associated with both the independent variable and the outcome, but are also part of the causal chain between them. In the diagram depicted in Figure 2, Z is the mediator between X and Y. If path c completely disappears when controlling for Z, then there is complete mediation. If, on the contrary, path c is not zero, then there is only partial mediation. Again, low socioeconomic status (SES) affects maternal nutrition and a deficient maternal nutrition increases the risk of LBW. However, there is a marginal effect of low SES on LBW, which cannot be completely accounted for by maternal nutrition.

The failure to distinguish between a confounder and a mediator is one of the most frequent errors in epidemiology. This distinction cannot be made on statistical grounds. An understanding of the process leading from the exposure to the disease is necessary.

When analyzing the probable causal relationship between an exposure and a disease, controlling for mediators can potentially lead to false conclusions. For instance, babies born to mothers with lower SES tend to have higher mortality rates. Controlling for birthweight reduces or nearly eliminates the differences between strata of SES. However, this does not mean that SES is not important as a causal factor of infant mortality. It just means that all or most of its

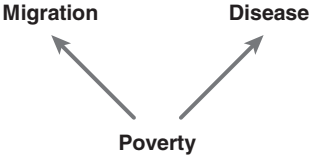


Figure 1 Poverty as a Confounder of the Relationship Between Migration and Disease

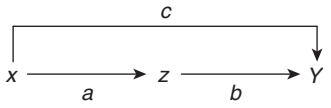


Figure 2 *Z* as a Mediating Variable Between *X* and *Y*

effect is expressed through the causal pathway represented by LBW.

There is yet a third way in which a factor *F* can influence the exposure-disease relationship. Factor *F* is said to be a *modifier or moderator* if it modifies the way in which the exposure and the disease are related. If exposure has different effects on disease at different levels of values of a variable, that variable is a modifier.

If a treatment is effective in cancer patients at Stages 1 or 2, and is ineffective in patients at Stages 3 or 4, then the stage of the disease modifies the effect of treatment on disease. In individuals with high cholesterol levels, smoking produces a higher relative risk of heart disease than it does in individuals with low cholesterol levels. Cholesterol is a modifier of the effect of smoking on heart disease. It can also be said that smoking *interacts* with cholesterol in its effects on heart disease.

As with confounding and mediation, the distinction between mediation and effect modification cannot be made on statistical grounds but on theoretical grounds.

Testing for Mediation

Several authors have provided algorithms to test for mediation. The best-known references are listed in the Further Readings. All these algorithms are based on linear regression. They can be summarized in the following steps (refer to Figure 2).

1. Show that the independent variable *X* is correlated with the outcome *Y*. To do this, regress *Y* on *X* and estimate the slope of the overall effect, which is given by the slope of the regression equation. By doing this, it has been established that there exists an effect that may be mediated.
2. Show that *X* is correlated with *Z*. For this, regress *Z* on *X*, which is equivalent to using the mediator as if it were an outcome, and estimate and test path *a*.
3. Show that *Z* is independently correlated with the outcome variable. To do this, regress *Y* on *X* and *Z*, and estimate and test path *b*. Observe that the regression is fitted with both *X* and *Z* as predictors. The purpose of this is to control for *X* in assessing the effect of the mediator on the outcome, and to avoid overlooking the fact that they could be both caused by the initial variable *X*. In this step, path *c* is also estimated and tested. If it does not significantly differ from zero, then complete mediation has been established. Otherwise, there is only partial mediation.

In Figure 2 and in the previous three-step algorithm, path *c* measures the direct effect of *X* on *Y*, while the product *ab* measures the indirect effect. It is important to note, however, that going through these steps and showing that *ab* is not zero does not conclusively prove that mediation has occurred, but only that it is consistent with the data.

—Jorge Bacallao Gallestey

See also Causal Diagrams; Causation and Causal Inference; Effect Modification and Interaction

Further Readings

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Hoyle, R. H., & Kenny, D. A. (1999). Statistical power and tests of mediation. In R. H. Hoyle (Ed.). *Statistical strategies for small sample research.* Thousand Oaks, CA: Sage.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602–619.

MacKinnon, D. P. (1994). Analysis of mediating variables in prevention and intervention research. In A. Cazares & L. A. Beatty (Eds.), *Scientific methods in prevention research* (pp. 127–153). NIDA Research Monograph 139. DHHS Pub. No. 94-3631. Washington, DC: Government Printing Office.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, *17*, 144–158.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*, 83–104.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445.

Medicaid

Medicaid, created by Title XIX of the Social Security Act of 1965, is a program that provides health insurance coverage for qualifying low-income individuals and their families. The program is administered through a state-federal partnership, with states having the authority to establish standards for eligibility, coverage of benefits, and payment rates. Over the past two decades, eligibility for Medicaid programs has been expanded to a variety of populations within the United States as a result of amendments to the original statute.

Overview of Benefits

Medicaid is administered by the states and territories of the United States. Each of the state Medicaid programs is financed jointly by the state and the federal government through a system of matching rate expenditures, as well as an allocation of federal funds to certain hospitals that treat a large number of Medicaid patients. Although the funding of the program is partly federal, the states have some control over how their particular program is structured and which health care services are covered by the program. Under the statute of Title XIX, each state has the ability to structure a Medicaid benefits package and payment system that fits their particular population needs.

While states are charged with designing and implementing their particular Medicaid program, all states must cover certain basic services. These core services include inpatient and outpatient care, visits to a physician, laboratory and imaging services, home health services, skilled nursing services, family planning, and general checkup and treatment services for children. Medicaid will also cover services provided in a nursing home or long-term care facility for those persons who have exhausted most of their financial assets. In addition to these services, states have the option to cover other services under Medicaid. These optional services include hospice care, other home and community-based services, dental care, physical and occupational therapy, rehabilitative services, coverage for eyeglasses, and prescription drug coverage. It is at the discretion of the state whether to include these optional services as part of the standard Medicaid benefits package.

Eligibility and Enrollment

Beneficiaries of the program include the categorically needy (families with children or certain groups of adults who meet specific income criteria), the medically needy (those who are blind and/or disabled), certain Medicare beneficiaries, the elderly, and persons who have very high medical bills. In recent years, federal legislation has expanded the traditional eligibility criteria to include additional groups that may receive Medicaid benefits.

Two pieces of legislation, in particular, give states the option of providing coverage to certain young adults; these eligible youth include those who are either in the foster care system or who are able to demonstrate financial independence. The Foster Care Independence Act of 1999 allows the state the ability to cover those individuals below the age of 21 who were in foster care on their 18th birthday. The other piece of legislation, the Benefits and Improvement Act of 2000, enables states to expand their eligibility criteria to include children below the age of 19 who qualify as Medicaid eligible based on information given to a school, homeless shelter, tribal program, or other qualified organization.

Additional federal legislation in the late 1990s and early 2000s expanded coverage to additional groups of disabled persons and women suffering from breast or cervical cancer. The Ticket to Work and Work Incentives Improvement Act of 1999 expands coverage to those disabled individuals who want to work and increase their earning capacity. Prior to this legislation, these individuals had been at risk for losing Medicaid coverage if they earned above a certain percentage of the federal poverty level. This act also establishes a program to allow states the option of providing Medicaid to people who will become blind or disabled as a consequence of their current illness. Additionally, the Breast and Cervical Prevention and Treatment Act of 2000 endows states with the ability to provide Medicaid assistance to women diagnosed with breast or cervical cancer and who are in need of treatment. Women who are eligible for Medicaid under this statute are entitled to all the services provided by the state Medicaid plan.

Each individual state is responsible for establishing the eligibility criteria for its Medicaid program, under the guidance of broad federal guidelines. In general, Medicaid eligibility is limited to the elderly, blind, and disabled, but may also include the parents or

caretakers of a child as well as pregnant women and their children. In addition to these qualifications, some beneficiaries must meet certain income criteria, with personal income and other resources below a percentage of the federal poverty level, as specified by the state. There are also state eligibility requirements that apply in certain circumstances, such as state residency and/or U.S. citizenship.

Administration of Medicaid

All the 50 states, Washington D.C., and the territories of Puerto Rico, Guam, American Samoa, the Virgin Islands, and the Northern Mariana Islands administer Medicaid programs. The state agency that administers this program is usually the state Department of Health and Human Services, which follows the guidelines set forth by the federal statute in Title XIX. In close coordination with this agency, the Centers for Medicare and Medicaid Services (CMS) assists in setting broad policy guidelines and monitoring the administration of the program.

While each state has a wide amount of discretion in the design and administration of the Medicaid program, each state must submit an administration plan for approval by CMS in accordance with the federal guidelines. States must specify eligibility criteria, benefits to enrollees, implementation guidelines and payment rates, and other requisite information. If a state wishes to change its current program, amendments must also be submitted to CMS for approval.

—Ashby Wolfe

See also Governmental Role in Public Health; Health Disparities; Medicare; Poverty and Health

Further Readings

Centers for Medicare and Medicaid Services. (2004). *CMS congressional guide: Medicare, Medicaid, SCHIP* (OCLC No. 57491652). Baltimore: Department of Health and Human Services, Center for Medicare and Medicaid Services, Office of the Administrator.

Centers for Medicare and Medicaid Services. (2005). *Medicaid-at-a-glance: A Medicaid information source* (Publication No. CMS-11024–05). Baltimore: Department of Health and Human Services. Retrieved July 26, 2007, from http://www.cms.hhs.gov/MedicaidGenInfo/Downloads/MedicaidAtAGlance2005.pdf.

Williams, B., Claypool, H., & Crowley, J. S. (2005, February). *Navigating Medicare and Medicaid: A resource guide for people with disabilities, their families and their advocates*. Washington, DC: Henry J. Kaiser Family Foundation. Retrieved July 26, 2007, from http://www.kff.org/medicare/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=50946.

Web Sites

Department of Health and Human Services, Centers for Medicare and Medicaid Services: http://www.cms.hhs.gov.

Medical Anthropology

Medical anthropology is a subdiscipline within anthropology that addresses sociocultural dimensions of health and illness, as well as the epistemologies and practices associated with diverse systems of healing. This entry examines medical anthropology's contribution to the study of the social production of health and illness. Although medical anthropologists move through the terrain of human health in various ways, this entry concentrates on a select few examples of theoretical and methodological contributions to the anthropological understanding of the political economy of health.

According to medical anthropologist Morgan (1987), the political economy of health is "a macroanalytic, critical, and historical perspective for analyzing disease distribution and health services under a variety of economic systems, with particular emphasis on the effects of stratified social, political, and economic relations within the world economic system" (p. 132). Political-economic medical anthropologists argue that health-threatening conditions are the result of historically based social, political, and economic systems of inequality. This perspective is bolstered by a methodological and conceptual commitment to the delineation of structures of inequality and their reproduction over time. The political-economic medical anthropology (PEMA) research framework thus engages the notion of change as an important part of the context of health.

PEMA research focuses on ideological and material foundations of inequality by examining the lived experiences of class relations and state-sponsored policies and practices. It expands microlevel, culturally based analyses of health and illness in particular communities or societies by illuminating the broader

context through which health phenomena are unevenly distributed among social groups. PEMA research thus explores the interaction among macrosocial forces and microlevel circumstances. As a result, PEMA studies emphasize the multifactorial nature of disease causality.

According to Morsy (1996), power is the central analytical construct within the PEMA framework. Power is a relational concept that describes the privileges of one group and concomitant subordination of others. Analyses of power allow researchers such as Ida Susser to make the connections between material and social resource distribution and structures of social inequality. In turn, focusing on the various and interconnecting dimensions of power enables PEMA scholars to understand population health and sickness as socially produced phenomena.

Political Economy in Population Health

On the basis of this political economic framework, PEMA scholarship raises a variety of questions about the social mechanisms of health and illness. Some biocultural anthropologists work to understand how environmental, social, and biological factors interact to produce differential health outcomes and the uneven spread of disease in populations. As Wiley (1992) explains, these researchers discern the effects of inequality on population health by focusing on biological variation and change that is associated with disease, psychophysiologic symptoms, and malnutrition. For example, Dressler (2005) elucidates the connections between hypertension, stress, and structurally mediated culture change for African Americans and Brazilian families. He finds that downward social mobility and the resulting inability to fulfill culturally based consumer norms induce stress-related illnesses such as hypertension and depression within these populations.

Another emerging trend in biocultural research involves the conceptual integration of social and human biological processes to examine health and illness as the direct outcome of political economic inequities. In their edited volume, Goodman and Leatherman (1998) provide a framework for applying political economic perspectives to biocultural studies. The contributions to this volume engage with diverse topics such as malnutrition, infant mortality, epidemic disease, and the impact that illness has on household production and reproduction. Swedlund and Ball (1998), for example, argue that poverty was the root cause of high infant mortality rates in a northeastern United States town in the early 20th century. Early studies that did not take into account the regional political-economic context blamed child death on the mothering skills of poor women rather than nutritional and ecological factors. Like Swedlund and Ball, other contributors working primarily on communities across the Americas illustrate how political-economic processes influence human biology, producing biological variation among groups that is expressed as illness and disease. Subsequently, human biological events affect the social fabric of communities and societies.

Focusing on the social construction of race, Mwaria (2001) argues that due to the history and ongoing patterns of racial/ethnic discrimination in the United States, researchers need to be aware of the potential for epidemiologic data to be misused. For example, Mwaria notes that certain genetic diseases are more common in some populations than others, as a result of historical and environmental factors. Thus, sickle cell disease (SCD) and sickle cell trait (SCT) are found in greater frequency among populations living in Africa, the Mediterranean, Saudi Arabia, and the Americas. However, to understand what this means in practice—in this instance, how SCT affects African Americans in the United States—requires moving from the level of medical ecology to that of critically informed biocultural anthropology. According to Molnar (1998) and Duster (1990), during the 1970s, institutions including the military considered African Americans as a population to be at high risk for developing sickle cell disease. Recruits who tested positive for SCT were considered unsuited for strenuous activities and especially for work conducted at high altitudes, including flying airplanes. The ban on recruits with SCT was ended in 1981, but the practice of testing prospective employees for SCT continues at some corporations. Put another way, Mwaria (2001) argues the importance of considering how medical information is used to understand the impact of genetic diseases that differentially affect specific racial/ethnic groups.

Political Economy in International Health

Medical anthropologists concerned with international health clarify the connections between macrolevel global power relations and population health. More generally, they show that global power relations, structured along the lines of financial and political

wealth, culminate in an uneven global distribution of adverse health effects. This body of research includes, but is not limited to, three broad research topics.

First, researchers such as Kim, Millen, Irwin, and Gershman (2000), Michele Rivkin-Fish (2005), and Paul Farmer (2001) participate in discussions concerning the uneven distribution of public health and health care resources among nations. For example, Farmer notes that prevention resources are concentrated in the United States, to address potential disease outbreaks, with considerably less attention given to ongoing epidemics in Haiti and elsewhere in the Global South.

Second, Imrana Qadeer, Nalini Visvanathan, and other medical anthropologists examine what kinds of health programs and issues are given priority within the context of international development programs. In their analysis of reproductive health programs in India, Qadeer and Visvanathan (2004) find that family planning services are designed to be compatible with strategies for national economic growth rather than the health of populations. International financial institutions encourage the Indian state to curb population growth to reduce internal expenditures, thereby freeing resources to service foreign debt.

Third, medical anthropologists draw attention to the ways in which international power structures shape local understandings of illness and disease. Adams (1998) explores how the Chinese-Tibetan conflict and international human rights discourses have affected the ways in which Tibetan dissidents define health and illness. Given these political, historical, and cultural factors, Adams argues, it is important to focus on the collective needs of Tibetans rather than health at the individual level.

The relationship between poverty and poor health is complicated. Accordingly, anthropologists, such as public health practitioners, point to the importance of research on multiple social forces that affect health. The topics discussed above are not discrete categories of analysis, but represent the broader discussions to which medical anthropology can contribute. Analyses often contribute to all three research foci and the interplay among them.

For example, in her book *Life Exposed: Biological Citizens after Chernobyl*, Petryna (2002) examines the making of "biological citizenship," or persons whose health needs and rights as citizens have been redefined as the result of their exposure to the 1986 Chernobyl nuclear disaster. In the context of international public health responses to the disaster, Petryna examines the ways in which scientific precision masked the arbitrariness of diagnosis with radiation illness. Narrowly defined categories of radiation-induced illness proved beneficial to strategies for economic development in post-Socialist Ukraine, insofar as they minimized the political and economic costs of this event. The use of these categories likewise meant that many survivors were not diagnosed with radiation illness and were thus considered ineligible for state-sponsored social support.

Political Economy at the Intersections of Social Life

Intersectionality theory is one of the primary theoretical tools that PEMA researchers use to guide political-economic analyses. This theoretical approach was developed by black feminist scholars as a way to understand the ways that racial/ethnic, class, and gendered inequalities work synergistically, producing outcomes that far exceed any of these constitutive factors. Consequently, intersectionality studies criticize the use of race, class, and gender as discrete variables. Applied to anthropology, intersectionality theorists use the ethnographic case study to understand the relationship between these various dimensions of social identity, and how they are experienced in daily life. The challenge in intersectionality studies is to move out of the abstract realm of social identities to identify specific processes through which health inequalities are produced. The following examples explore intersectionality as it relates to the political economy of health.

In her ethnography of health and healing in Egypt, Morsy (1993) focuses on the historical contexts of sickness and healing for poor women. Morsy explores the connections among the historical trajectories of state economic and social service policies, the unfolding of international development programs in Egypt, and transformations in production and the labor force. These changes are reflected in household gender dynamics that produce unequal chances for health and access to certain forms of healing. The consequence is that women, and poor women in particular, have fewer health care choices than men. As a result, Morsy's analysis brings into focus the interlocking nature of social locations and social inequality, macrolevel political economic relationships, and human health and illness.

Mullings and Wali's (2001) ethnographic research in central Harlem illustrates that African American

women's experiences with problematic birth outcomes are shaped by resource inequality, institutionalized racism, and gender discrimination. Stressors such as unemployment, impoverishment, occupational duress, violence, and the lack of affordable (and well-maintained) housing are the results of inadequate urban planning, along with changes in welfare policies implemented at national and local levels during the 1990s. These stressors are compounded by pregnancy, decreasing black women's chances for positive reproductive health outcomes. By examining the context of women's daily lives, Mullings and Wali highlight the multicausal nature of high infant mortality rates among black women.

In their edited volume, Parker, Barbosa, and Aggleton (2000) explore the complex interplay among social, cultural, political, and economic processes that sustain power imbalances at the level of communities and in the daily lives of residents. The contributors to this volume argue that it is insufficient to focus on sexuality or sexual risk taking in isolation; rather, sexual practices should be considered within broader political, cultural, and historical contexts. The chapters in this volume provide diverse examples from Indonesia, Brazil, Argentina, the United States, Costa Rica, Mexico, South Africa, and the Philippines to examine ways in which global relations of power, as well as local cultural tradition, articulate with experiences of sexuality. For example, in an analysis drawn from ethnographic research in developing countries, Mane and Aggleton (2000) examine the ways in which class and prevailing gender relations exacerbate women's vulnerability to sexually transmitted diseases. They find that, although the female condom is a potentially empowering technology, economic and social factors impinge on women's abilities to insist on protected sex. While sex workers generally reported higher use of condoms, especially the female condom, poor women who lacked power in marital relationships were most likely to be forced to have unprotected sex. Analyses such as Mane and Aggleton's reveal the complex interplay between gender and sexuality, the local and the global, power and resistance.

In her study among pregnant drug addicted women, Whiteford (1996) explores the ways in which poor, African American women experience discrimination in the medical and law enforcement systems. She focuses on a Florida statute that mandated drug testing and prison sentences for pregnant women who, on testing positive, could be charged with fetal endangerment. However, only the public hospitals that primarily served poor and African American women conducted drug testing at prenatal appointments. In contrast, middle- and upper-income women with access to private health care were shielded from this kind of surveillance during their pregnancies. The result of this policy was that low-income African American women were disproportionately subject to incarceration during which time they were denied access to health care. Rather than address important public health concerns, including the need for accessible prenatal care and drug treatment programs tailored to the specific concerns of pregnant women, such laws simply punish women of color for being poor and pregnant.

From Theory to Practice

In addition to theoretical contributions to medical anthropology, PEMA studies making use of the intersectionality approach seek to move beyond the realm of purely academic research. Their aim is to apply study findings to improve the material conditions in which health and illness are produced, and thereby reduce inequities in health. To this end, many PEMA scholars engage with the research process and research participants in a participatory manner.

Specifically, PEMA scholars seek to build collaborative relationships with community activists to understand the inner workings of health organizations and coalitions that are endeavoring to effect a movement for social change. Morgen (2002), for example, conducted ethnographic research in a feminist clinic in the United States to better grasp what it means for women to take control of their health. Studying women's experiences with running the clinic sensitized Morgen to the racialized politics of the women's health movement. In the clinic, women of color struggled for autonomy as well as inclusion within a dominantly white framework of activism and devised new strategies to address issues heretofore neglected by the women's movement.

According to Lock and Kaufert (1998), research that is guided by the perspectives of participants engenders new social scientific definitions and local strategies of resistance, agency, choice, and compliance. For example, Lopez (1998) points out that the reproductive "choices" of Puerto Rican women can only be understood when their perspectives concerning sterilization are placed in historical context. In the early 1900s, the

convergence of poverty, colonialism, the eugenics movement, and the commodification of family planning positioned sterilization as an economically and morally effective means of birth control. While the sterilization of Puerto Rican women has its roots in this context of coercion, the contemporary choice to undergo the procedure is often seen by women as a strategy for resisting the increased responsibilities that accompany large families. For Lopez, choice and resistance must be viewed through the perspectives of research participants.

The capacity for research to foster collaborative relationships and garner input from all participants is not serendipitous and must, instead, be built into the framework of the study. Inhorn (2001) argues that health research methodologies are most effective when they combine qualitative, quantitative, and community participatory research strategies. Participant observation, longitudinal case studies, and focus groups provide the means for data verification and enrichment of quantitative analyses. Various forms of community input can then guide ethnographic interpretations, the selection of fieldwork locations, and lines of inquiry. These diverse approaches to data collection, in turn, promote more holistic understanding of health-affecting circumstances and processes. Honing methodological dimensions of the PEMA framework are, for many political-economic medical anthropologists, an essential piece of the research process that warrants interdisciplinary dialogue.

Developing Interdisciplinary Dialogue

From the early work of Janes, Stall, and Gifford (1986) to the more recent efforts by Inhorn (1995) and Trostle (2005), medical anthropologists have explored points of convergence between medical anthropology and epidemiology. These include mutual interest in health as a human right, concerns about the health of populations at large, and attention to social dimensions in the prevention and treatment of disease. Medical anthropologists and epidemiologists alike examine multiple factors involved with the social production of disease. Through its focus on the contextualization of health inequities, medical anthropological research such as that of Carole Browner deepens understandings of the relationships between structures of inequality and individual health behaviors. Interdisciplinary exchange between medical anthropology and epidemiology contributes to the development of measures of health that link inequality with the physical expression of ill-health and disease. In this interdisciplinary framework, researchers redirect attention from the biomedical understanding of disease as solely or predominantly biological and toward an understanding of the social relations of poor health. This more holistic approach has the potential to create more effective social action to facilitate health and prevent or treat disease.

—Alyson Anthony, Mary Alice Scott, and Mary K. Anglin

See also African American Health Issues; Community Health; Health Disparities; Immigrant and Refugee Health Issues; Sexual Risk Behavior; Syndemics; Urban Health Issues; Women's Health Issues

Further Readings

General

Morgan, L. M. (1987). Dependency theory in the political economy of health: An anthropological critique. *Medical Anthropology Quarterly*, *1*(2), 131–154.

Morsy, S. A. (1996). Political economy in medical anthropology. In T. Johnson & C. F. Sargent (Eds.), *Medical anthropology: Contemporary theory and method* (2nd ed., pp. 21–40). Westport, CT: Praeger.

Susser, I. (1996). The construction of poverty and homelessness in U.S. cities. *Annual Review of Anthropology*, *25*, 411–435.

Political Economy in Population Health

Dressler, W. W. (2005). What's cultural about biocultural research? *Ethos*, *33*(1), 20–45.

Duster, T. (1990). *Backdoor to eugenics*. New York: Routledge.

Goodman, A. H., & Leatherman, T. L. (Eds.). (1998). *Building a new biocultural synthesis: Political-economic perspectives on human biology*. Ann Arbor: University of Michigan Press.

Kevles, D. J. (1995). *In the name of eugenics*. Cambridge, MA: Harvard University Press.

Molnar, S. (1998). *Human variation: Races, types, and ethnic groups*. Upper Saddle River, NJ: Prentice Hall.

Mwaria, C. (2001). Diversity in the context of health and illness. In I. Susser & T. C. Patterson (Eds.), *Cultural diversity in the United States* (pp. 57–75). Malden, MA: Blackwell.

Swedlund, A. C., & Ball, H. (1998). Nature, nurture, and the determinant of infant mortality: A case study from Massachusetts, 1830–1920. In A. H. Goodman &

T. L. Leatherman (Eds.), *Building a new biocultural synthesis: Political-economic perspectives on human biology* (pp. 191–228). Ann Arbor: University of Michigan Press.
Wiley, A. (1992). Adaptation and the biocultural paradigm in medical anthropology: A critical review. *Medical Anthropology Quarterly*, *6*(3), 216–236.

Political Economy in International Health

Adams, V. (1998). Suffering the winds of Lhasa: Politicized bodies, human rights, cultural difference, and humanism in Tibet. *Medical Anthropology Quarterly*, *12*(1), 74–102.
Farmer, P. (2001). *Infections and inequalities: The modern plagues*. Berkeley: University of California Press.
Kim, J. Y., Millen, J. V., Irwin, A., & Gershman, J. (Eds.). (2000). *Dying for growth: Global inequality and the health of the poor*. Monroe, ME: Common Courage Press.
Petryna, A. (2002). *Life exposed: Biological citizens after Chernobyl*. Princeton, NJ: Princeton University Press.
Qadeer, I., & Visvanathan, N. (2004). How healthy are health and population policies? The Indian experience. In A. Castro & M. Singer (Eds.), *Unhealthy health policy: A critical anthropological examination* (pp. 145–160). Walnut Creek, CA: Altamira Press.
Rivkin-Fish, M. (2005). *Women's health in post-Soviet Russia: The politics of intervention*. Bloomington: Indiana University Press.

Political Economy at the Intersections of Social Life

Mane, P., & Aggleton, P. (2000). Cross-national perspectives on gender and power. In R. Parker, R. M. Barbosa, & P. Aggleton (Eds.), *Framing the sexual subject: The politics of gender, sexuality, and power* (pp. 104–116). Berkeley: University of California Press.
Morsy, S. A. (1993). *Gender, sickness, and healing in rural Egypt: Ethnography in historical context*. Boulder, CO: Westview Press.
Mullings, L., & Schulz, A. J. (2006). Intersectionality and health: An introduction. In A. J. Schulz & L. Mullings (Eds.), *Gender, race, class, and health: Intersectional approaches* (pp. 3–17). San Francisco: Jossey-Bass.
Mullings, L., & Wali, A. (2001). *Stress and resilience: The social context of reproduction in Central Harlem*. New York: Kluwer Academic.
Parker, R., Barbosa, R. M., & Aggleton, P. (Eds.). (2000). *Framing the sexual subject: The politics of gender, sexuality, and power*. Berkeley: University of California Press.
Whiteford, L. M. (1996). Political economy, gender, and the social production of health and illness. In C. F. Sargent & C. B. Brettell (Eds.), *Gender and health: An international perspective* (pp. 242–259). Upper Saddle River, NJ: Prentice Hall.

From Theory to Practice

Inhorn, M., & Whittle, K. L. (2001). Feminism meets the "new" epidemiologies: Toward an appraisal of antifeminist biases in epidemiological research on women's health. *Social Science & Medicine*, *53*, 553–567.
Lock, M., & Kaufert, P. A. (Eds.). (1998). *Pragmatic women and body politics*. New York: Cambridge University Press.
Lopez, I. (1998). An ethnography of the medicalization of Puerto Rican women's reproduction. In M. Lock & P. A. Kaufert (Eds.), *Pragmatic women and body politics* (pp. 240–259). New York: Cambridge University Press.
Morgen, S. (2002). *Into our own hands: The women's health movement in the United States, 1969–1990*. New Brunswick, NJ: Rutgers University Press.
Singer, M. (1995). Beyond the ivory tower: Critical praxis in medical anthropology. *Medical Anthropology Quarterly*, *9*(1), 80–106.

Interdisciplinary Dialogue

Browner, C. (1999). Ethnicity, bioethics, and prenatal diagnosis: The Amniocentesis decisions of Mexican-origin women and their partners. *American Journal of Public Health*, *89*(11), 1658–1666.
Inhorn, M. (1995). Medical anthropology and epidemiology: Divergences or convergences? *Social Science & Medicine*, *40*(3), 285–290.
Janes, C., Stall, R., & Gifford, S. M. (1986). *Anthropology and epidemiology: Interdisciplinary approaches to the study of health and diseases*. Boston: Kluwer Academic.
Trostle, J. A. (2005). *Epidemiology and culture*. New York: Cambridge University Press.

Medical Expenditure Panel Survey

The Medical Expenditure Panel Survey (MEPS), conducted by the Agency for Healthcare Research and Quality (AHRQ), is an ongoing study conducted in the United States that collects data on health care utilization and costs and insurance coverage. It consists of four separate components: the Household Component (HC), the Nursing Home Component (NHC), the Medical Provider Component (MPC), and the Insurance Component.

The MEPS HC is a nationally representative survey of the U.S. civilian, noninstitutionalized population, using a sampling frame drawn from the previous year's National Health Interview Study sampling

frame. The HC, which has been conducted continuously since 1966, collects data on all members of a sampled household or family from a single member of the household and uses an overlapping panel design so that data are collected from each participating family or household for 2 years. Topics covered by the HC include family demographics; health conditions; health status; health care utilization, including physician visits, hospital utilization including inpatients care and visits to the Emergency Department and Outpatient Department, dental care, home health care, use of prescription and over-the-counter medications; health care expenditures; health insurance coverage; and household income and assets.

The MPC of the MEPS is a survey conducted with providers and facilities that provided health care to individuals included in the HC; this includes hospitals, physicians, and medical providers working under their supervision; home health care agencies; and long-term care institutions. Data collected by the MPC are used to verify and supplement data collected in the HC about charges, payments, and sources of payment for health services and are used to estimate the expenses of people enrolled in managed care plans; MPC data are not released as a stand-alone file. Collection of MPC data requires that the HC respondents give their consent to have MEPS contact their care providers, so not all providers of care to HC respondents are included, and the MPC sample for this reason is not nationally representative.

The IC, which collects data on employer-based health insurance, has been conducted annually since 1996. IC data were originally collected from two different samples, the *household sample* and the *list sample*. The household sample was originally selected from the insurance providers (unions and insurance companies) and employers of respondents to the previous year's HC: These providers and employers acted as proxy respondents who provided insurance information for the HC respondents. The data collected were then attached to the respondent's HC record. The definition of who was included in the household sample has changed several times, and this survey is no longer conducted: Data were collected in 1996, 1997, 1998, 1999, 2001, and 2002. The list sample collects information from a nationally representative sample of workplaces about the types and costs of health insurance offered through employers, including governments.

The NHC was conducted in 1996 only on a nationally representative sample of nursing homes and residents of nursing homes. Data collected about the facilities include structure (e.g., if it was part of a hospital or retirement center), ownership, staffing, number of beds, number of residents, and type and size of special care units. Data collected from residents include demographics, insurance coverage, health status, and medical conditions.

Most researchers will be interested in analyzing MEPS HC data, which are publicly accessible and available for download from the MEPS data site for the years 1996–2004. Due to confidentiality concerns, data from the NHC, MPC, and IC components are not publicly accessible. Researchers may apply for access to NHC and MPH data at the AHRQ's Center for Financing, Access, and Cost Trends Data Center; researchers cannot access the IC data directly but can produce tables using IC data through the MEPSnet online interface.

—Sarah Boslaugh

See also Health Care Delivery; Health Care Services Utilization; Health Economics; Secondary Data

Further Readings

Agency for Health Care Policy and Research. (1997). *Design and methods of the Medical Expenditure Panel Survey Household Component*. Rockville, MD: Author. Retrieved July 26, 2007, from http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr1/mr1.shtml.

Web Sites

Medical Expenditure Panel Survey: http://www.meps.ahrq.gov/mepsweb/index.jsp.

Medicare

Medicare, established under Title XVIII as part of the Social Security Act of 1965, is the federal health care financing program that provides health insurance for the elderly, the disabled, and those with end-stage renal disease (ESRD). Generally, Medicare covers care that is reasonable, necessary, and related to a diagnosed illness or injury. Medicare currently has four programs, which provide a variety of services to its enrollees (beneficiaries). These programs include Part A for inpatient services, Part B for outpatient and physician

services, Medicare Advantage for those services provided by private health plans, and Part D for those prescription drugs not covered under Parts A or B.

The Medicare program has evolved significantly since its inception, as a result of changes to the original statute. Originally, Medicare included Parts A and B and was available only to those above 65 years of age. The law was amended in 1972 to include those individuals entitled to disability benefits and in 1976 to include those with ESRD. The Medicare law was amended in 1997 as a result of the Balanced Budget Act to establish the Medicare + Choice program (Part C), which provides private health plan choices to beneficiaries. Additionally, as a result of the Medicare Prescription Drug, Improvement and Modernization Act of 2003, the Part D program was added to provide additional prescription drug benefits and to replace the Medicare + Choice program with Medicare Advantage.

Overview of the Medicare Programs

Medicare Part A

Medicare Part A is the portion of this program that pays for hospital, or inpatient, services. Covered services include inpatient hospital care, care at a skilled nursing facility, home health care, and hospice care. To receive these benefits, each person must meet certain eligibility qualifications. Most beneficiaries who receive coverage under Medicare Part A do not pay a premium, because the program is financed through payroll taxes that were deducted while beneficiaries were working. Part A is known as a fee-for-service benefit, whereby coverage is determined based on the services rendered by the health care provider.

Medicare Part B

The Medicare Part B program provides supplementary medical insurance to those electing to receive the benefit. There is a premium associated with this portion of the program, which is paid monthly by the beneficiary. Part B provides coverage for outpatient health services provided in local clinics or home health service organizations, as well as coverage for specific medical devices and equipment. Covered services include physician services, other outpatient care, drugs and biologicals, durable medical equipment, and preventive services (such as annual screenings). Services that are not covered by Part B include routine physicals, foot care, hearing aids or eyeglasses, dental care, and outpatient prescription drugs. Part B is also a fee-for-service program.

Medicare Advantage (Part C)

Formerly known as Medicare Part C or the Medicare + Choice program, Medicare Advantage is the Medicare program that allows beneficiaries to receive covered benefits through a private health plan. Specifically, if a person is enrolled in Medicare Advantage, there is no need to be enrolled in Medicare Parts A, B, or D. Medicare Advantage provides coverage of inpatient and outpatient services, in addition to prescription drug coverage, in one plan, which is administered by a variety of private health plans. Medicare Advantage is not a fee-for-service program.

Medicare Part D

Medicare Part D, the newest program within Medicare, provides coverage for prescription drugs for those who choose to enroll in the program. Benefits include coverage for certain prescription drugs not covered under Parts A or B of the Medicare program. Coverage is provided in two ways: (1) by private plans that offer drug-only coverage or (2) through the Medicare Advantage program, which offers both drug benefits and health insurance coverage. Covered drugs under this program include those drugs available by prescription, as well as vaccines, insulin, and the associated medical supplies required to administer the medication.

Eligibility and Enrollment

Using the Medicare guidelines, the Social Security Administration is responsible for determining those eligible to receive Medicare benefits. Each portion of the Medicare program has particular criteria. In general, those Americans aged 65 or older are entitled to receive hospital insurance under Medicare Part A. In addition to these beneficiaries, those with disability and most patients with ESRD or kidney failure are entitled to receive benefits under Part A. Persons are able to enroll 3 months prior to, and 3 months following, the day on which they become 65.

Medicare Part B is a voluntary program, and those who are able to receive Part A are also eligible to enroll in Part B to receive supplemental insurance for outpatient services. Additional eligible individuals

include those permanent residents of the United States who have lived in the United States for 5 years. Part B has a general enrollment period in January through March of each calendar year; beneficiaries are also able to enroll during the 3 months prior to, and 3 months following, their becoming 65. There are additional requirements, and certain penalties, for those eligible individuals who choose not to enroll during the period in which they are initially eligible.

Medicare Advantage (Part C) is available to those beneficiaries who are eligible for Parts A and B but choose instead to enroll in a private health plan. Beneficiaries do have the option of returning to their original Medicare coverage under Parts A and B if they elect to do so.

Medicare Part D is available to those beneficiaries who quality for Parts A and B. The initial enrollment period for this program occurred between November 2005 and May 2006; however, subsequent enrollment will occur annually between November 15 and December 31 of each calendar year. For those eligible beneficiaries who choose not to enroll during these periods, a late fee surcharge, similar to the penalty for Part B late enrollment, is assessed.

Administration of Medicare

Medicare is administered by the Centers for Medicare and Medicaid Services (CMS), a federal administration agency under the supervision of the Department of Health and Human Services. Formerly known as the Health Care Financing Administration (HCFA), today CMS coordinates both Medicare operations at the federal level and assists each state with the administration and operation of Medicaid.

CMS contracts with several agents outside of the federal government to provide payment and insurance services to its beneficiaries. Each of these companies processes claims relating to particular services under the separate parts of the Medicare program. Under Part A of the program, those companies processing claims for inpatient, skilled nursing or home health/hospice services are known as *fiscal intermediaries.* Those companies, which pay claims relating to Medicare Part B and all medical supplier claims, are known as carriers. Also contracting with CMS are organizations known as integrity program contractors, which investigate fraud and abuse claims.

For those beneficiaries seeking information about Medicare, agencies equipped to provide such information include the Social Security Administration, the Department of Health and Human Services, and the Centers for Medicare and Medicaid Services.

—Ashby Wolfe

See also Aging, Epidemiology of; Governmental Role in Public Health; Medicaid

Further Readings

Centers for Medicare and Medicaid Services. (2004). *CMS Congressional Guide: Medicare, Medicaid, SCHIP* (OCLC No. 57491652). Baltimore: Department of Health and Human Services, Center for Medicare and Medicaid Services, Office of the Administrator.

Centers for Medicare and Medicaid Services. (2006). *Medicare and Medicaid Statistical Supplement* (Publication No. 03469). Baltimore: Department of Health and Human Services, Center for Medicare and Medicaid Services, Office of Research, Development and Information.

Web Sites

Department of Health and Human Services, Center for Medicare and Medicaid Services: http://www.cms.hhs.gov. Medicare: http://www.medicare.gov.

Men's Health Issues

Men in the United States suffer more severe chronic conditions, have higher death rates for most of the leading causes of death, and die nearly 5 1/2 years younger than women. Why are there such gender differences, and why are some men healthy and others are not? The definition of health is complex, as is the answer to these questions. To improve the health of men, health care providers and public health professionals must better understand the determinants of men's health and become advocates for change of the social and economic factors that affect these determinants.

This entry presents an overview of (1) selected epidemiologic aspects of men's health; (2) the reported causes and "actual" causes of death for men; (3) the role of "gender" as a determinant of health; (4) the influence of selected dimensions of the social and economic environment, such as poverty, education, socioeconomic status, racism, and social capital, on health

status and outcomes in men; and (5) the role of stress as a mediator between these dimensions and health.

Epidemiology

Health, United States, *2005* (National Center for Health Statistics) provides extensive data on trends and current information on selected determinants and measures of health status relevant to the health of men and differences between men and women. These data raise many questions and suggest areas for further study and interventions related to differences in health outcomes by gender, race, education, and other variables. The information in this report includes the following:

- *Life Expectancy.* In 2002, life expectancy for males (at birth) was 74.5 years, while for females, 79.9 years. Between 1990 and 2002, life expectancy at birth increased more for the Black population than for the White population, thereby narrowing the gap in life expectancy between these two racial groups. In 1990, life expectancy at birth was 7.0 years longer for the white than for the black population. By 2003, the difference had narrowed to 5.2 years. However, for black men, the difference was 6.3 years.

- *Death Rates.* For males and females, age-adjusted death rates for all causes of death are three to four times higher for those with 12 years or less education compared with those of educational attainment of 13 years or more. Males continued to have higher death rates due to diseases of the heart (286.6 vs. 190.3), malignant neoplasms (233.3 vs. 160.9), chronic liver disease and cirrhosis (12.9 vs. 6.3), HIV disease (7.1 vs. 2.4), motor vehicle injuries (21.6 vs. 9.3), suicide (18.0 vs. 4.2), and homicide (9.4 vs. 2.6). In 2002, adolescent boys (15 to 19 years) were five times as likely to die from suicide as adolescent girls, in part reflecting their choice of more lethal methods, such as firearms.

- *Cancer Incidence.* Incidence rates for all cancers combined declined in the 1990s for males. Cancer incidence was higher for black males than for males of other racial and ethnic groups. In 2001, age-adjusted cancer rates for black males exceeded those for white males by 50% for prostate, 49% for lung and bronchus, and 16% for colon and rectum.

- *Tobacco Use.* In 2003, 24% of men were smokers, compared with 19% of women. Cigarette smoking by adults is strongly associated with educational attainment. Adults with less than a high school education were three times more likely to smoke than were those with at least a bachelor's degree or more from college.

- *Alcohol Use.* Among current drinkers 18 years and older, 40% of men and 20% of women reported drinking five or more alcoholic drinks on at least one day (binge drinking) in the past year. Among males in Grades 11 and 12, 22.4% drove after drinking alcohol, compared with 12.3% of females.

- *Seat Belt Use.* In 2003, 22% of male high school students rarely or never used a seat belt compared with 15% of female high school students.

- *Access to Health Care and Health Insurance.* Working-age males 18 to 64 years were nearly twice as likely as working-age females to have no usual source of health care (22% vs. 12%). Men of all ages, particularly between the ages of 18 and 54, are less likely than women to visit physician offices and hospital outpatient and emergency departments. For all persons below 65 years of age, males are less likely to have health insurance than are females.

"Real" Versus "Actual" Causes of Death

The mortality data presented above represent the reported causes of death on death certificates and indicates the primary pathophysiologic conditions identified at the time of death, as opposed to the root causes of the death. Major external (nongenetic) modifiable factors that contribute to death have been labeled "the actual causes of death." Half of the deaths that occurred among U.S. residents in 1990 were potentially preventable and could be attributed to the following factors: tobacco use (19%), diet/activity patterns (14%), alcohol (5%), microbial agents (4%), toxic agents (3%), firearms (2%), sexual behavior (1%), motor vehicles (1%), and illicit use of drugs ($< 1\%$). A similar analysis of the "actual causes of death" in 2000 showed that tobacco smoking remains the leading cause of mortality but diet and physical inactivity may soon overtake tobacco as the leading cause of death.

Gender

These striking differences in health status and outcomes for men and women result from a complex mix of beliefs, behaviors, biology, and socialization. Many sociocultural factors, including gender, are

associated with, and influence, health-related beliefs and behaviors.

There are gender differences in health beliefs. Compared with men, women rely less on "provider control" of health (doctors being in charge), express greater "nutritional consciousness," and believe more that psychological factors play an important part in the etiology of illness. There are gender differences in perceptions of cancer. Women are more frightened than men of cancer. Men are more frightened of heart disease than cancer. The greatest fear of cancer was its perceived incurability and the associated suffering. The greatest fear of heart disease was perceived susceptibility. Men are more likely than women to hold a more negative attitude toward cancer information and more likely to identify a cause of cancer as behavior rather than heredity. The greatest barrier to seeking services in male college students is their socialization to be independent and to conceal vulnerability.

The concept of masculinity may vary among communities and cultures, but the development and maintenance of male identity usually requires taking risks that are hazardous to health—more dangerous jobs, more homicides and car accidents, excess drinking of alcohol, smoking, and substance abuse. Unwillingness to admit weakness may prevent many men from consulting a doctor when an illness arises, from taking health promotion messages seriously, or admitting to and seeking care for mental illness. Factors such as ethnicity, economic status, educational level, sexual orientation, and social context influence the kind of masculinity that men construct and contribute to the differential health risks among men in the United States.

Poverty and Social Status

Poverty and social inequalities may be the most important determinants of poor health worldwide. Poverty is a multidimensional phenomenon that can be defined in both economic and social terms. Poverty leads to a person's exclusion from the mainstream way of life and activities in a society. Socioeconomic differences in health status exist even in industrialized countries where access to modern health care is widespread. There is convincing evidence of an increase in differential mortality rates according to socioeconomic level in the United States. Not surprisingly, mortality rates from most major causes are higher for persons in lower social classes. The Whitehall studies of British civil servants showed that mortality rates are three times greater for the lowest employment grades (porters) than for the highest grades (administrators). Conventional risk factors (smoking, obesity, low levels of physical activity, high blood pressure, and high plasma cholesterol) levels explain only about 25% to 35% of the differences in mortality rates among persons of different incomes. Income disparity, in addition to absolute income level, is a powerful indicator of overall mortality. Male mortality is more unequal than female mortality across socioeconomic groups.

Variables that have been postulated to intervene between income inequality and health status include civic engagement and levels of mutual trust among community members, and dimensions of social capital. One's control of the work environment may be an important connection between social and occupational class and mortality. Some researchers suggest that education is the critical variable.

Social Capital

Studies across the world indicate that social support (marriage, family, group affiliations) affects mortality after controlling for baseline differences in health status. A consistent pattern exists in which high levels of social capital are associated with desirable health outcomes.

Social capital has been shown to be associated with decreasing depression, suicide, colds, heart attacks, strokes, and cancer. It has also been associated with sociological factors such as reduced crime, juvenile delinquency, teen pregnancy, child abuse, drug abuse, and increased graduation rates/test scores.

Multiple studies have demonstrated an association between reported well-being and social connectedness. The protective effects of social connectedness have been shown for family ties, friendship networks, participation in social events, and association with religion and other civic organizations. These protective factors reduce the likelihood of developing colds, heart attacks, strokes, cancer, depression, and many sources of premature death. In fact, the strength of social integration and social support on health is believed to be as great as well-known risk factors such as smoking, obesity, and physical inactivity. How and/or why social connectedness is associated with well-being is uncertain, but researchers believe that the support (emotional and financial) that social networks provide decreases stress and, as such, reduces illness. Social networks increase communication between people, may support healthy norms in the community such

as not smoking and physical activity, and social cohesion may promote activism around important issues such as health insurance.

A nationwide study that related social capital and state-level health outcomes using data on trust and group membership in 39 states revealed that levels of social trust and group membership were significantly associated with heart disease, malignant neoplasms, and infant mortality. Increased trust and group membership decreased mortality rates even after controlling for income and poverty levels.

Racism, Discrimination, and Bias

Another important determinant of male health is access to, and the quality of, the health care "system." However, access and quality of care are not equal for all men. Numerous studies have found that racial and ethnic minorities tend to receive a lower quality of health care than nonminorities, even when access-related factors such as patients' insurance status and income are controlled. The sources of these disparities are complex and rooted in historic and contemporary inequities, and involve many participants at many levels. As one of the participants, health care providers may contribute to the racial and ethnic disparities found among men. Three mechanisms may be operative: bias or prejudice against minorities, greater clinical uncertainty when interacting with minority patients, and beliefs or stereotypes held by the provider about the behavior or health of minorities.

External stressors such as racism may contribute directly to the physiological arousal that is a marker of stress-related diseases. In addition, anger in young men has been associated with premature cardiovascular disease.

Stress as a Mediator

Individuals experience objective psychological and environmental conditions—such as discrimination, racism, mistrust, poverty, and diminished social capital—that are conducive to stress. The perception of stress is influenced by social, psychological, biophysical factors, genetics, and behavior. When the brain perceives an experience as stressful, physiological and behavioral responses are initiated, leading to allostasis (homeostasis) and adaptation. Wear and tear across multiple physiological systems becomes a significant contributor to overall health risk. Such wear and tear is hypothesized to result from repeated exposures to social relationship conflicts. Over time, allostatic load can accumulate, and overexposure to mediators of neural, endocrine, and immune system stress have adverse effects on various organ systems, leading to enduring negative health outcomes—physiological (e.g., hypertension and cardiovascular disease), psychological (e.g., depression), and behavioral (e.g., alcoholism and drug abuse).

—James Plumb, Rickie Brawer, and Lara Carson Weinstein

See also Cardiovascular Disease; Health Disparities; Social Capital and Health; Stress

Further Readings

Committee on Health and Behavior: Research, Practice, and Policy Board on Neuroscience and Behavioral Health. (2001). *Health and behavior: The interplay of biological, behavioral and societal influences.* Washington, DC: National Academy Press.

Doyal, L. (2001). Sex, gender, and health: The need for a new approach. *British Medical Journal*, *323*, 1061–1063.

Kaplan, G. A., Pamuk, E. R., Lynch, J. W., Cohen, R. D., & Balfour, J. L. (1996). Inequality in income and mortality in the United States: Analysis of mortality and potential pathways. *British Medical Journal*, *312*, 999–1003.

McCally, M., Haines, A., Fein, O., Addington, W., Lawrence, R. S., & Cassel, C. K. (1998). Poverty and health: Physicians can, and should, make a difference. *Annals of Internal Medicine*, *129*(9), 726–733.

Mustard, C. A., & Etches, J. (2003). Gender differences in socioeconomic inequality in mortality. *Journal of Epidemiology and Community Health*, *57*, 974–980.

National Center for Health Statistics. (2005). *Health, United States, 2005 with chartbooks on trends in the health of Americans.* Hyattsville, MD: Government Printing Office

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

van Ryn, M., & Fu, S. S. (2003). Paved with good intentions: Do Public Health and Human Service Providers contribute to racial/ethnic disparities in health? *American Journal of Public Health*, *93*(2), 248–255.

Wender, R. C., & Haines, C. (2006). *Men's health, an issue of primary care: Clinics in office practice.* A title in The Clinics: Internal Medicine series. Philadelphia: Saunders Elsevier.

Mercury

Mercury, also known as *quicksilver*, has been known since ancient times and is represented in the periodic

table by the symbol "Hg," which stands for *hydrargyrus*, or liquid silver, in Latin. It is a silvery transitional metal that is liquid at or near standard room temperature. Mercury has many uses, both in homes and in industry, and it has also been used as a medicine, although it has been acknowledged as being toxic to humans.

Mercury exists predominantly in three forms: elemental, inorganic, and organic. Methylmercury is the most important organic form of mercury in terms of human health effects. It has a high affinity for the brain, particularly the posterior cortex. It is neurotoxic (damaging to the nervous system), toxic to the developing fetus, and genotoxic (damaging to the DNA), and it can cause effects such as numbness and tingling, stumbling gait, weakness and fatigue, vision and hearing loss, spasticity and tremors, and in high enough concentrations, coma. Efforts have been made to decrease mercury exposure through public policy initiatives such as those of the U.S. Environmental Protection Agency.

History

There is evidence to suggest that mercury was known to the ancient Chinese and Hindus and was found in Egyptian tombs dating back to about 1500 BCE. Mercury has found many uses in ancient civilizations, including making of ointments, cosmetics, amalgams with other metals, and in alchemy. It was thought to prolong life, heal fractures, and preserve health. Indeed, it was named after the Roman god Mercury, known for his speed and mobility. It has been used as a diuretic, disinfectant, laxative, as a treatment for syphilis and worm infestation, in thermometers, in sphygmomanometers (blood pressure measuring devices), and as an antidepressant.

In the 18th and 19th centuries, mercury was used in the industrial process of carrotting, a method of treating fur in the process of making felt. Animal skins were rinsed in a solution of mercuric nitrate that helped open the sheaths surrounding each fur fiber and permitted matting (felting) of fibers in subsequent operations for making felt hats. The process, however, produced highly toxic mercury vapors and led to mercury poisoning among hatters. Many experienced tremors, emotional lability, insomnia, dementia, and hallucinations, and these symptoms led to the phase commonly used in medical parlance, "mad as a hatter," which refers to someone poisoned by mercury. It was also known as Danbury shakes, due to the effects seen in Danbury, Connecticut, a center of hat making. The U.S. Public Health Service banned the use of mercury in the felt industry in December 1941.

Some prominent historical personalities known or believed to be affected by mercury toxicity include Sir Isaac Newton, King Charles II, and Sir Michael Faraday. Their erratic behavior was thought to correspond to their work with mercury. Abraham Lincoln also exhibited erratic behavior that was thought to be due to the mercury in the "blue pill" he took for depression.

Toxicology and Clinical Manifestations

Metallic or elemental mercury volatilizes to odorless mercury vapor at ambient air temperatures, and it can be absorbed via inhalation, with concerns particularly in poorly ventilated spaces. Inhalation of mercury vapors may produce inflammation of the respiratory passages and a pneumonitis-like syndrome and the triad of excitability, tremors, and gingivitis (the mad hatter syndrome). Inorganic mercury salts can be divalent (mercuric salts) or monovalent (mercurous salts). They are generally white powder or crystals, with the exception of cinnabar (mercuric sulfide), which is red. The greatest concentrations of mercury after exposure to the inorganic salts or vapors can be found in the kidney. Mercuric salts are more corrosive and toxic than the mercurous salts. "Pink disease" has been seen in children when teething powders containing mercurous mercury has been used and is characterized by fever; pink rash; swelling of the spleen, lymph nodes, and fingers; constipation or diarrhea; hair loss and irritability. Organic mercury compounds are formed when mercury combines with carbon. As noted above, methylmercury is the most important organic form of mercury in terms of human health effects.

Mercury Exposure

The major sources of mercury exposure today are the natural degassing of the earth's crust, mining, and the consumption of fish containing mercury. In addition to miners, others subject to occupational risks of mercury exposure include technicians, nurses, those doing dental work, and machine operators. Workers involved in industrial production of elemental mercury, cinnabar (ore containing mercury) mixing and

processing, and the manufacturing and use of instrumentation containing elemental mercury are at a higher risk. Average urinary mercury among dentists in the United States was well above the general population mean, but it dropped after an educational campaign on mercury hygiene sponsored by the American Dental Association. Mercury exposure to patients filled with dental amalgam is somewhat controversial. Although it has been shown that routine activities such as tooth brushing, chewing gum, and cleaning and polishing of teeth result in high concentrations of mercury in the mouth in patients with amalgam tooth fillings, the average absorbed dose has been shown to be much less than environmentally absorbed mercury. Some other common uses of mercury today are in barometers, cell batteries, calibration instruments, fluorescent and mercury lamps, photography, silver and gold production, thermometers, fungicides, paper manufacturing, and wood preservatives.

On August 11, 2006, the U.S. Environmental Protection Agency (EPA) announced the National Vehicle Mercury Switch Recovery Program to remove mercury-containing light switches from scrap vehicles. During dismantling of discarded cars, a significant amount of mercury is released into the environment during melting of the scrap metal to make new steel and steel products. This can be prevented if mercury-containing switches are removed from scrap vehicles before they are shredded. This program aims at helping cut mercury air emissions by up to 75 tons over the next 15 years.

Some of the more recent mass exposures of mercury include the famous Minamata Bay incident in Japan (1960) and methylmercury-treated grain in Iraq (1960, 1970). The former refers to the contamination of the Minamata Bay in Japan by tons of mercury compounds dumped into it by a nearby company. Thousands of people whose normal diet consisted of fish from the bay developed mercury poisoning. The methylmercury incident in Iraq was one involving import of wheat grain contaminated with methylmercury used as a pesticide.

Diagnosis, Evaluation, Treatment, and Prevention

Acute inhalation exposure to high concentrations of elemental mercury vapors is a medical emergency. Aggressive supportive care is needed. Development of acute pneumonitis should be watched for. Mercury can be chelated with agents such as penicillamine, dimercaprol, or dimercaptosuccinic acid. Measurement of mercury in a 24-hr specimen is used to confirm exposure. Acute ingestion of mercuric chloride, usually with suicidal intent, is another medical emergency. Chelation can help excretion, and hemodialysis may be required in severe cases.

Chronic exposure is best assessed by measuring urine mercury, preferably from 24-hr collected urine. It should be remembered, however, that correlation of urine or blood mercury levels with toxicity is poor. For methylmercury, body burden can be estimated from measurement of mercury in whole blood or hair, although the reliability of these, especially the latter, is often questioned in clinical medicine.

Once the exposure and signs and symptoms are confirmed, the person should be removed from the exposure to avoid further ongoing toxicity. Pregnant women, or women intending to become pregnant in the near future, should limit the intake of food that contains elevated mercury levels, such as swordfish, shark, mackerel, and certain tuna.

—Abhijay P. Karandikar

See also Environmental and Occupational Epidemiology; Pollution

Further Readings

Diner, B., & Brenner, B. *Mercury toxicity.* Retrieved September 8, 2006, from http://www.emedicine.com/EMERG/topic813.htm.

Klaassen, C. D., & Watkins, J. B., III. (1999). *Casarett & Doull's toxicology companion handbook* (5th ed.). New York: McGraw-Hill.

Rosenstock, L., Cullen, M. R., Brodkin, C. A., & Redlich, C. A. (2004). *Textbook of clinical occupational and environmental medicine* (2nd ed.). Philadelphia: Elsevier Saunders.

U.S. Environmental Protection Agency. (2007). *Mercury.* Retrieved September 8, 2006, from http://www.epa.gov/mercury.

Meta-Analysis

In epidemiology, the proliferation of multiple and sometimes contradictory studies can be a challenge

for interpretation of health risk and health policy formulation. One approach to synthesizing the results of separate but related studies is meta-analysis—the systematic identification, evaluation, statistical synthesis, and interpretation of separate study results. For example, for many years, conflicting results were reported in observational studies of the effect of diet on breast cancer risk. The lower rate of breast cancer incidence for women in Asian countries suggested a protective effect for soy-based diets; yet migration patterns and changes in diet yielded conflicting results. A synthesis of epidemiologic studies showed a moderately protective effect for soy intake (odds ratio $[OR] = 0.89$, 95% confidence interval $[CI]$ $0.75 - 0.99$), with a stronger effect among premenopausal women ($OR = 0.70$, 95% $CI = 0.58 - 0.85$).

This entry reviews the elements of a well-conducted meta-analysis, summarizes recent research, and discusses two important examples of the use of meta-analysis in epidemiology: *The Guide to Community Preventive Services* and the Human Genome Epidemiology Network.

The technique of using a quantitative synthesis probably was used first by Karl Pearson in 1904 to increase statistical power in determining the efficacy of a vaccine for enteric fever; Gene Glass coined the term *meta-analysis* in 1976 to apply to systematic review and quantitative synthesis. From the social sciences, use of meta-analysis quickly spread to medicine in the 1980s. Later, meta-analysis was used increasingly to combine results from observational studies.

Over time, meta-analysis has become more prominent in epidemiology, extending to important policy decisions and determining the effectiveness of interventions. To address the quality of reporting of meta-analytic reviews, guidelines were developed for reporting randomized controlled trials (RCTs) to facilitate synthesis, meta-analysis of RCTs, and meta-analysis of observational studies.

Elements of a Well-Conducted Meta-Analysis

Stating the Problem and Conducting the Literature Search

A well-conducted meta-analysis should start with an explicit statement of the research problem, which can be framed by population, intervention (or exposure), comparison, or outcome. After specifying the study question, the next step is a systematic search for relevant studies. Computerized databases have aided this step, particularly in meta-analyses of RCTs. However, limiting a search to two or three electronic databases might produce incomplete evidence. A comprehensive search will include multiple databases, the reference lists of recent review articles and meta-analyses, and frequently contact with experts to find unpublished results.

With the proliferation of meta-analyses in the epidemiologic literature and the availability of electronic repositories of research, variably skilled researchers are conducting searches. Recognizing the importance of knowledge and skills in complex bibliographic retrieval and verification of information, the Medical Library Association has developed a policy that health science librarians should contribute to the search process for health and information research.

Collection of Data

Abstraction of data from the search should begin with explicit inclusion and exclusion criteria for studies. Commonly used criteria include period covered in the review, operational definitions of the variables, the quality of a study, and the language of publication. To the extent possible, inclusion or exclusion criteria should be based on valid scientific principles (e.g., treatment changes over time), not the convenience of the researcher.

The procedures for abstracting data should be developed in detail and documented. Blinding the abstractor(s) to the identity of the journal or the results, for example, can reduce bias; however, blinding is difficult to achieve, time-consuming, and might not substantially alter results. If possible, multiple abstractors should assess the data, and the report should include a calculation of interrater reliability.

Assessment of Study Quality

One of the most controversial questions related to meta-analysis is the question of whether to include studies that are of doubtful or poor quality. Critics argue that any meta-analysis that summarizes studies of widely differing quality is likely to be uninformative or flawed. Indeed, studies with methodologic flaws have been demonstrated to overestimate accuracy in studies of a medical test. One study of published cost-effectiveness studies reported that studies

funded by industry, studies of higher methodologic quality, and those published in Europe and the United States were more likely to report favorable cost-effectiveness ratios.

Other researchers counter by noting that assessing methodologic quality is often difficult, and researchers often disagree on what constitutes quality. Despite a researcher's best attempts to provide an objective measure of quality, decisions to include or exclude studies introduce bias into the meta-analysis.

Still others note that the quality of a study might not have an effect on the study's outcome. When in doubt, include the study in the meta-analysis and use an independent variable to code the quality of a study; for example, using quality to stratify the estimates. Guidelines have been developed to assess quality of RCTs, but even here, improvement is needed.

Evaluation of the Body of Evidence Collectively

Assessment of Publication Bias

Reporting of publication bias, the tendency to publish findings (or not) based on bias at the investigator or editorial level (e.g., failure to publish results of studies demonstrating negative results), is a major problem for meta-analysis in epidemiology. This bias can be related to strength or direction of results, author's native language, the sex of the author, or country of publication.

Of the methods developed to address publication bias, perhaps the simplest is the funnel plot, a type of scatter plot with estimate of sample size on one axis and effect size estimate on the other (Figure 1). The utility of the funnel plot to assess publication bias is based on the statistical principle that sampling error decreases as sample size increases. Other statistical tests can help assess deviation from symmetry, but these are controversial because of high Type I error rates. A more robust approach includes a comprehensive search and estimating contribution from the components of publication bias.

Assessment of Heterogeneity

When studies to be combined in a meta-analysis are heterogeneous, the interpretation of any summary effect might be difficult. Tests for heterogeneity most often use a formulation of Hedges Q statistic; however, this method has been reported to have low power when the variance of the effect size differs among studies (which is the case most of the time). This problem can be addressed with meta-regression techniques and graphical approaches. Statistical methods have been developed to assist with determining

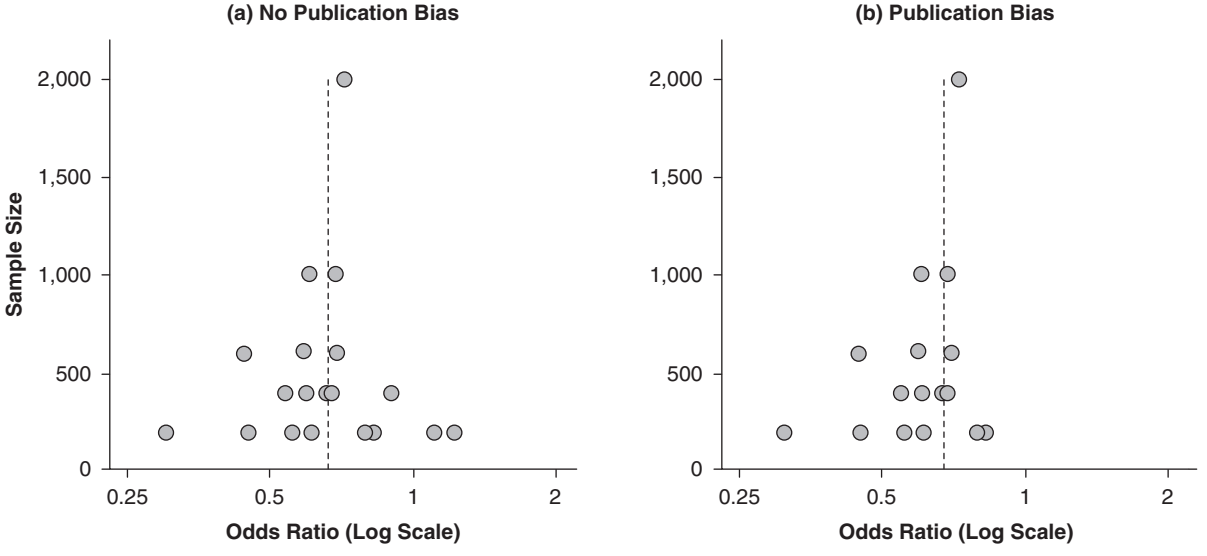


Figure 1 Sample Funnel Plots Illustrating Publication Bias

the source and nature of heterogeneity. Whether the heterogeneity is important, however, requires judgment beyond the statistics.

Consideration of Quantitative Synthesis of Evidence

A quantitative synthesis might not be useful if the number of studies is limited, the studies are of low quality, or important conceptual or empirical heterogeneity exists. If the studies located are appropriate for a quantitative synthesis, fixed-effects or random-effects models can be used in a meta-analysis, depending on the presence or absence of heterogeneity. The fixed-effects model applies to a situation that assumes each study result estimates a common (but unknown) pooled effect. The random-effects model assumes that each study result estimates its own (unknown) effect, which has a population distribution (having a mean value and some measure of variability). Thus, the random-effects model allows for between- and within-study variability. Summary estimates from heterogeneous studies should be interpreted with caution, even when using a random-effects model, bearing in mind that the source of heterogeneity might be as important to identify estimate of risk.

Recently, Bayesian meta-analysis has been used more frequently, allowing both the data and the model itself to be random parameters. Bayesian methods allow the inclusion of relevant information external to the meta-analysis. Although use of Bayesian meta-analysis methods is hindered by software limitations and not easily understood *p* value, it leads easily to a framework that can consider costs and utilities when making decisions based on epidemiologic studies.

Cumulative meta-analysis is the process of performing a new (or updated) meta-analysis as results become available. For certain public health problems, and indeed for the majority of situations with limited resources, adding additional information for an already established effect is of limited value. Thus, methods to determine when sufficient evidence exists to establish intervention effectiveness are critical.

Emerging Methods

As meta-analysis becomes more widely used, numerous aspects can be handled by emerging methods. For example, meta-analysis is often complicated by lack of information on standard deviation of estimates in reports; the validity of various methods of imputing this information from other sources has been studied. Combining information from different study designs or evidence on multiple parameters is of interest to researchers. Models for this situation are complex but provide opportunities to assess whether data are consistent among studies. Finally, availability of computer packages has aided in the conduct of meta-analysis in epidemiology.

Meta-Analysis in Epidemiologic Practice: Two Examples

The Guide to Community Preventive Services (*Community Guide*) provides decision makers recommendations regarding population-based interventions to promote health and to prevent disease, injury, disability, and premature death. The Task Force on Community Preventive Services makes its recommendations on the basis of systematic reviews of topics in three general areas: changing risk behaviors; reducing diseases, injuries, and impairments; and addressing environmental and ecosystem challenges. Systematic reviews (and quantitative synthesis where appropriate) are conducted for selected interventions to evaluate the evidence of effectiveness that is then translated into a recommendation or a finding of insufficient evidence. For those interventions where evidence of effectiveness is insufficient, the *Community Guide* provides guidance for further prevention research.

The completion of sequencing of the human genome and advances in genomic technology provide tremendous opportunities for using genetic variation in epidemiology. The Human Genome Epidemiology Network (HuGENet™) is a global collaboration committed to the assessment of the impact of human genome variation on population health and how genetic information can be used to improve health and prevent disease. To address the quality of reporting of genotype prevalence and gene disease association, an interdisciplinary group of epidemiologists has developed guidelines for appraising and reporting such studies. Such guidelines will contribute to the quality of meta-analyses. Furthermore, systematic reviews of genetic studies and publication of results increase the utility of this discipline.

The use of meta-analysis in epidemiology has moved beyond just recognizing the need to stay current with the literature in a field experiencing an

explosion of studies. The method has driven recommendations about improved reporting of abstracts and primary studies, identified research gaps, and shifted attention away from statistical significance to consideration of effect size and confidence interval. The term *evidence-based* will continue to have varying interpretations, but the role of meta-analysis in epidemiology is pushing the science of public health further.

—Stephen B. Thacker and Donna F. Stroup

See also Bayesian Approach to Statistics; Evidence-Based Medicine; Publication Bias

Further Readings

Cochrane Collaboration. (2006). *The Cochrane Collaboration: The reliable source of evidence in health care*. Oxford, UK: Author. Retrieved April 20, 2006, from http://www.cochrane.org/index.htm.

Dickersin, K. (2002). Systematic reviews in epidemiology: Why are we so far behind? *International Journal of Epidemiology*, *31*, 6–12.

Egger, M., Stern, J. A. C., & Smith, G. D. (1998). Meta-analysis software [Electronic version]. *British Medical Journal*, *316*. Retrieved April 20, 2006, from http://bmj.bmjjournals.com/archive/7126/7126ed9.htm.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *The Educational Researcher*, *10*, 3–8.

Glasziou, P., Irwig, L., Bain, C., & Coldtiz, G. (2001). *Systematic reviews in health care: A practical guide*. New York: Cambridge University Press.

Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficacy of search methods in systematic reviews of complex evidence: Audit of primary sources. *British Medical Journal*, *331*, 1064–1065.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Jadad, A. R., Moher, D., & Klassen, T. P. (1998). Guides for reading and interpreting systematic reviews. II. How did the authors find the studies and assess their quality? *Archives of Pediatric & Adolescent Medicine*, *152*, 812–817.

Little, J., Bradley, L., Bray, M. S., Clyne, M., Dorman, J., Ellsworth, D. L., et al. (2002). Reporting, appraising and integrating data on genotype prevalence and gene-disease associations. *American Journal of Epidemiology*, *156*(4), 300–310.

McGowan, J., & Sampson, M. (2005). Systematic reviews need systematic searchers. *Journal of the Medical Library Association*, *93*, 74–80.

Muellerleile, P., & Mullen, B. (2006). Efficiency and stability of evidence for public health interventions using cumulative meta-analysis. *American Journal of Public Health*, *96*, 515–522.

Mullen, P. D., & Ramírez, G. (2006). The promise and pitfalls of systematic reviews. *Annual Review of Public Health*, *27*, 81–102.

Petitti, D. B. (1994). *Meta-analysis, decision analysis, and cost-effectiveness analysis: Methods for quantitative synthesis in medicine*. New York: Oxford University Press.

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., et al. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Journal of American Medical Association*, *283*, 2008–2012.

Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, *10*, 277–303.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester, UK: Wiley.

Truman, B. I., Smith-Akin, C. K., & Hinman, A. R. (2000). Developing the guide to community preventive services: Overview and rationale. *American Journal of Preventive Medicine*, *18*(Suppl. 1), 18–26.

Web Sites

The Community Guide: http://www.thecommunityguide.org/.

Human Genome Epidemiology Network, Centers for Disease Control and Prevention: http://www.cdc.gov/genomics/hugenet/default.htm.

Migrant Studies

Migrant studies are an extension of the ecologic study design, which compares disease rates in different locations. In migrant studies, the disease rate among persons who have migrated from one location to another is compared with the disease rate in persons who did not migrate. Ideally, the rate for migrants is compared both with persons remaining in the country of origin and with lifelong residents of the destination or host country. Migrants share their genetic makeup and early life environment with persons remaining in the country of origin. They share recent environmental exposures with residents of the host country. Thus, the comparison of disease rates between the migrants and the nonmigrating populations is used to generate hypotheses about the relative importance of genetics and environment in determining disease risk. Migrant

disease rates that remain similar to those in the country of origin suggest that genetic factors play a role in geographic variation. Migrant disease rates that converge on that of the host country suggest an important role for the environment. Comparisons between persons migrating at different ages, or between migrants and their offspring, may identify a critical age of exposure for environmental factors.

Migrant studies are likely to be informative when there is marked geographic variation in disease and little is known about the etiologic factors responsible for the variation. The diseases most extensively studied with migrant populations are multiple sclerosis and cancer. Multiple sclerosis has an unusual geographic pattern, with prevalence generally higher with increasing distance from the equator, in both hemispheres. Migrants from higher to lower risk areas, such as Europeans to Israel, South Africa, or Australia, or internal migrants from higher to lower latitudes within the United States, have disease risk intermediate between their place of origin and destination. However, there is little change in risk for migrants moving from lower to higher risk areas, such as Africa to Israel or southern to northern United States. Studies able to ascertain age at migration suggest disease risk is largely determined by age 15 or 20. These studies have been interpreted as most compatible with an environmental exposure in childhood being an important etiologic factor, such as a protective effect from early infection by an agent endemic in areas closer to the equator.

Most cancers show significant geographic variations. Breast and stomach cancers offer two contrasting patterns. Breast cancer rates are highest in the most developed Western countries. Rates for white migrants to the United States converge to rates for U.S.-born whites within 20 years of migration. However, although breast cancer rates for Asian migrants to the United States and their U.S.-born daughters are higher than rates for women in Asia, they are not as high as U.S.-born whites. This suggests the importance of environmental factors in breast cancer risk, but it does not exclude the possibility that there might be genetic differences in risk for Asian women, since even U.S.-born Asian women have lower rates of breast cancer than whites.

The international pattern for stomach cancer is quite different: Japan has one of the highest rates and the United States one of the lowest. Migrant studies, including those of Japanese migrants to Hawaii and their offspring, found that migrants had rates somewhat lower than persons remaining in Japan, while their offspring had much lower stomach cancer rates. This pattern suggests the importance of environmental factors, both in childhood and adulthood.

In addition to estimating change in disease risk at the population level, migrant populations also offer the opportunity to investigate the etiologic role of environmental factors that change dramatically with immigration, such as diet. For such studies, migrants are recruited to participate in case-control or cohort studies, and risk factors are directly assessed at the individual level.

Limitations of Migrant Studies

The two most serious limitations to migrant studies are data quality and selection bias. Migrant studies often use data routinely collected for surveillance, and this raises concerns about data quality and the comparability of data from different countries. Diagnostic criteria and the completeness of disease ascertainment may vary between countries. When death certificates are used to identify migrants, place of birth may be incomplete. Neither year of migration nor parents' birthplaces are generally recorded. Thus, death certificate studies are limited to first-generation migrants. Sample sizes of migrants with a particular disease may be small, limiting analyses to standardized mortality ratios or standardized incidence ratios. Ethnic identification may be inconsistent between sources of numerator data, such as death certificates or cancer registries, and denominator sources, usually a census.

International migration is not a random event, and the opportunity to carry out migrant studies depends on historical circumstances. Frequent migration destinations such as the United States, Australia, and Israel are usually the study settings, although there are also migration studies within a country, often from rural to urban areas. A fundamental limitation of migrant studies is that persons who migrate are different from persons who do not. They may differ systematically by religion, education, ethnicity, or social class. International migration is often a difficult process, and successful migrants are likely to be persons with intellectual, emotional, health, and financial resources. For these reasons, it is impossible to know whether the migrants would have experienced the same disease risk as persons remaining in the country of origin had they not migrated. Studies have often observed that

migrants experience generally good health, which is sometimes called the "healthy immigrant" effect. In some cases, such as migrants to Israel, there may be no ethnically similar people remaining in the country of origin for comparison. If migrants systematically differ from nonmigrants with respect to underlying disease risk, even in the absence of the migration experience, then migrant studies are subject to selection bias.

—Diane Sperling Lauderdale

See also Asian American/Pacific Islander Health Issues; Bias; Immigrant and Refugee Health Issues; Latino Health Issues; Secondary Data

Further Readings

Gale, C. R., & Martyn, C. N. (1995). Migrant studies in multiple sclerosis. *Progress in Neurobiology, 47*, 425–448.

Haenszel, W., Kurihara, M., Segi, M., & Lee, R. K. (1972). Stomach cancer among Japanese in Hawaii. *Journal of the National Cancer Institute*, *49*, 969–988.

Parkin, D. M., & Khlat, M. (1996). Studies of cancer in migrants: Rationale and methodology. *European Journal of Cancer, 32A*, 761–771.

Thomas, D. B., & Karaga, M. R. (1996). Migrant studies. In D. Schottenfeld & J. F. Fraumeni Jr. (Eds.), *Cancer epidemiology and prevention* (2nd ed., pp. 236–254). New York: Oxford University Press.

Missing Data Methods

Virtually all epidemiologic studies suffer from some degree of missing or incomplete data. This means that some of the cases have data missing on some (but not all) of the variables. For example, one patient in a study may have values present for all variables except age. Another may have missing data on blood pressure and years of schooling. Missing data create problems for most statistical methods because they presume that every case has measured values on all variables in whatever model is being estimated. This entry surveys some of the many methods that have been developed to deal with these problems.

The most common method for handling missing data is complete case analysis (also known as listwise or casewise deletion). In this method, cases are deleted from the analysis if they have missing data on any of the variables under consideration, thereby using only complete cases. Because of its simplicity and because it can be used with any kind of statistical analysis, complete case analysis is the default in nearly all statistical packages. Unfortunately, complete case analysis can also result in the deletion of a very large fraction of the original data set, leading to wide confidence intervals and low power for hypothesis tests. It can also lead to biased estimates if data are not missing completely at random (to be defined below).

To avoid these difficulties and to salvage at least some of the discarded information, many different methods have been developed. Most of these methods are crude, ad hoc, and may only make things worse. Although more principled and effective methods have appeared in the past 20 years, they are still woefully underutilized.

Assumptions

Before examining various missing data methods, it is important to explain some key assumptions that are often used to justify the methods. The definitions given here are intended to be informal and heuristic.

Missing Completely at Random (MCAR)

Many missing data methods are valid only if the data are missing completely at random. Suppose that only one variable Y has missing data and that there are other variables (represented by the vector X) that are always observed. We say that data on Y are missing completely at random if the probability of missing data on Y is completely unrelated either to X or to Y itself. Symbolically, this is expressed as

$$Pr(Y \text{ missing } |X, Y) = Pr(Y \text{ missing}).$$

Note that missingness on Y can depend on other *unobserved* variables. It just can't depend on variables that are observed for the model under consideration. This definition can easily be extended to situations in which more than one variable has missing data, but the representation gets more complicated. Note also that MCAR is not violated if missingness on one variable is related to missingness in another variable. In an extreme but common situation, two variables may be always missing together or always present together; this does not violate MCAR.

Missing at Random (MAR)

MCAR is a very strong assumption. Some of the newer missing data methods are valid under the weaker assumption of missing at random. Again, let's suppose that only one variable Y has missing data, and that there are other variables X that are always observed. We say that data on Y are missing at random if the probability of missingness on Y may depend on X, but does not depend on Y itself after adjusting or controlling for X. In symbols, we say $\Pr(Y \text{ missing}|X, Y) = \Pr(Y \text{ missing}|X)$. Here's an example of a violation of this assumption: People with high income may be less likely to report their income, even after adjusting for other observed variables.

Although the MAR assumption is still fairly strong, it's much weaker than MCAR because it allows missingness to be related to anything that's observed. It's also an untestable assumption because a test would require that we know the values that are missing. If the data are MAR and if the parameters governing the missing data mechanism are distinct from the parameters in the model to be estimated, the missing data mechanism is said to be *ignorable*. That means that one can make valid inferences without explicitly modeling the mechanism that produces the missing data. Because the distinctness requirement is almost always satisfied, the terms *missing at random* and *ignorability* are often used interchangeably.

Goals for Missing Data Methods

To compare missing data methods, we need some criteria for evaluating them. Nowadays, most experts stress the quality of parameter estimates that are obtained when using a missing data method. In that regard, the goals for missing data methods are just the usual goals for good estimation:

1. Minimize bias.
2. Minimize sampling variability, that is, make the true standard errors as small as possible.
3. Get good estimates of uncertainty, that is, estimated standard errors and confidence intervals.

Most conventional methods fail to accomplish one or another of these goals. Nearly all of them (except for complete case analysis) are deficient in the estimation of standard errors and confidence intervals.

Conventional Methods

Here is a brief look at some conventional missing data methods and how they stack up on the three goals.

Complete Case Analysis

If the data are MCAR, the subset of the cases that are complete can be regarded as a simple random sample from the target sample. It follows that complete case analysis will not introduce any bias into parameter estimates, and the standard error estimates will be accurate estimates of the true standard errors. On the other hand, the standard errors may be much larger than necessary because much usable data have been discarded. Furthermore, depending on the analytic method and the parameters to be estimated, complete case estimates may be severely biased if the data are MAR but not MCAR. Surprisingly, however, neither MCAR nor MAR is required for complete case analysis when data are missing only on predictor variables in any kind of regression analysis.

Available Case Analysis

For linear models, a popular missing data method is available case analysis (also known as pairwise deletion). This method rests on the fact that, in most applications, the parameters in linear models can be expressed as functions of means, variances, and covariances. The basic idea is to estimate these "moments" by conventional methods, using all available data (so different numbers of cases might be used to estimate different moments), then substitute these estimates into formulas for the parameters of interest. If the data are MCAR, this method produces estimates that are consistent and, hence, approximately unbiased. However, the method may break down completely for some patterns of missing data. More important, there is no good way to get accurate estimates of standard errors using this method.

Dummy Variable Adjustment

For regression analysis with missing data on a predictor variable, a popular method is to substitute some constant for the missing data (e.g., the mean or 0) and then include in the regression an indicator (dummy) variable for whether or not the data are missing. For categorical predictors, a related method is to create an

additional category for missing data. Unfortunately, these methods have been shown to produce biased estimates, even when the data are MCAR.

Imputation

A wide variety of missing data methods fall under the general heading of imputation, which simply means to fill in some reasonable guesses for the missing values, based on other available information. There are many ways to do this, the simplest of which is to replace missing values with the means of the observed values. Once all the missing data have been imputed, the analysis is done with standard statistical methods.

Virtually all conventional methods of imputation suffer from two problems. First, variances tend to be underestimated, leading to biases in any parameters that depend on variances (e.g., regression coefficients). Second, and arguably more serious, standard errors tend to be underestimated, leading to confidence intervals that are too narrow and p values that are too low. The reason for this should be intuitively obvious. Imputation substitutes "made up" data for real data, but standard software has no way of distinguishing imputed from real data. Consequently, standard error estimates assume more information than is really present.

Maximum Likelihood

One of the most general and effective methods for handling missing data is maximum likelihood estimation (MLE). Of course, MLE is widely used for all sorts of applications that do not involve missing data, for example, logistic regression. What makes it so attractive for missing data applications is that it satisfies all the three goals mentioned above. Estimates are approximately unbiased; all the available information is used to produce estimates that have minimum sampling variability; and standard errors are consistent estimates of the true standard errors.

Most software for MLE with missing data is based on the assumption that the data are missing at random. To do MLE, one first needs a probability model for the joint distribution of all the variables in the model of interest. For categorical variables, for example, one might specify a log-linear model. For continuous variables, a multivariate normal model is commonly assumed. Next, one must construct the likelihood function, which expresses the probability of the data as a function of the unknown parameters. Once the likelihood function is constructed, the final step is to find values of the parameters that maximize the likelihood.

If the missing data mechanism is ignorable (and therefore MAR) the construction of the likelihood function is fairly simple. When observations are independent, the likelihood for the whole sample is just the product of the probabilities of observing each of the observations. Suppose the model of interest has three variables X, Y, and Z. Suppose, further, that a particular observation has observed values of X and Y but is missing a value on Z. Let $p(x, y, z)$ be the probability function for all three variables. Then, if Z is discrete, the likelihood for that observation is found by summing $p(x, y, z)$ over all possible values of z. If Z is continuous, one integrates over the full range of values for z. If data are missing on that observation for both Y and Z, one integrates or sums over both variables.

Once the likelihood function has been constructed, maximization usually requires some numerical algorithm. For missing data applications, one popular maximization method is known as the expectation-maximization (EM) algorithm. This method involves iteration between two steps until convergence. The E-step consists of calculating the expected value of the log-likelihood, where the expectation is taken over variables with missing data, given the current values of the parameters. The M-step consists of maximizing this expected log-likelihood to get new values of the parameters. Most of the better-known statistical packages contain procedures for applying the EM algorithm under the multivariate normal model. The output of these procedures consists of ML estimates of the means, variances, and covariances.

Although the EM algorithm has its attractions, there are many other numerical methods that will yield the same ML estimates. For estimating the parameters of log-linear models with missing data, there are several freeware and commercial software packages available. For estimating linear models, it is a common practice to first estimate the means and covariance matrix by the EM algorithm and then use those statistics as input to a linear modeling program. Although this two-step procedure will give the correct parameter estimates, it will not yield correct standard errors, confidence intervals, or test statistics. To get correct values of these statistics, one must use "direct maximum likelihood" for the specified model. This

method is currently available in many structural equation modeling programs.

Multiple Imputation

When feasible, MLE is an excellent method for handling missing data. Unfortunately, for many applications either the theory or the software is lacking. For example, it's not easy to find MLE software for doing logistic regression or Cox regression. In such cases, multiple imputation is an attractive alternative. Like ML, the estimates produced by multiple imputation are approximately unbiased with accurate standard errors. The statistical efficiency is *almost* at the maximum level achieved by MLE, but not quite.

The advantages of multiple imputation over MLE are that (a) it can be used with virtually any kind of data or model and (b) conventional software can be used for estimation. There are two disadvantages, however. First, there are many different ways to do multiple imputation, leading to some uncertainty and confusion. Second, every time you use it, you get a (slightly) different result, because a random element is deliberately introduced into the imputation process.

A fundamental principle of multiple imputation is that imputed values should be random draws from the predictive distribution of the variable with missing data, conditional on the values of the observed variables. For example, if imputations are based on linear regression (one of the more popular approaches to multiple imputation), the imputed values are generated by, first, calculating predicted values as one usually does with linear regression and, second, adding a random draw from the residual distribution for the regression equation. Introduction of this random element avoids the biases that typically occur with deterministic imputation.

The other key principle of multiple imputation is that more than one randomly imputed value should be generated for each missing datum. How many? Three is the minimum, and many multiple imputation software programs use a default of five. More is always better, but the returns diminish rapidly. Two things are accomplished by making the imputations multiple: (a) parameter estimates are considerably more efficient (i.e., have less sampling variability) and (b) having multiple imputations makes it possible to get good standard error estimates that accurately reflect the uncertainty introduced by the imputation process.

In practice, the multiple imputations are used to construct multiple data sets, and each data set contains different imputed values for the missing data. Conventional software is then applied to each data set to estimate the parameters (and their standard errors) for the model of interest. Using a few simple rules, these multiple estimates are then combined to get a single set of parameter estimates, standard errors, and test statistics. For the parameter estimates themselves, a simple average of the multiple estimates is sufficient. Combination of the standard errors requires the following steps: (a) square the estimated standard errors (producing variance estimates) and average them across the data sets; (b) calculate the variance of the parameter estimates across the multiple data sets; (c) add the results of (a) and (b) (applying a small correction factor to the variance); and (d) take the square root of that sum.

As noted, the most popular method for generating the imputed values is linear regression with random draws from the residual distribution. This is straightforward when there is only a single variable with missing data, but typically runs into difficulty when several variables have missing data. One solution, based on Bayesian principles, is an iterative Markov chain Monte Carlo (MCMC) algorithm that involves long sequences of repeated iterations.

Of course, many variables with missing data are categorical, and linear regression may not be a plausible method of imputation in those cases. Although ad hoc fix-ups of the linear model often work well for such variables, imputation can alternatively be based on logistic regression, Poisson regression, log-linear models, and other semi- or nonparametric approaches. The downside is that the MCMC algorithm may be difficult or impossible to implement for such models. Consequently, software based on these approaches typically use iterative methods that have no theoretical foundation.

Nonignorable Missing Data

As with ML, most methods and software for doing multiple imputation are based on the assumption that the missing data mechanism is ignorable and, hence, that the data are MAR. It's important to stress, however, that both multiple imputation and ML estimation can handle applications where the data are *not* MAR. Because there are often reasons to suspect that the data are not MAR (e.g., sicker patients may be more

likely to drop out of a study than healthier patients), there has been no shortage of attempts to develop specialized models and methods for the nonignorable case. But doing this successfully requires a correct mathematical model for the missing data mechanism, and that is something that is usually quite difficult to come by. An incorrect choice can produce highly misleading results, and there is no way to test the fit of the model. For those who want to pursue such approaches, it is strongly advised that one try out several different missing data models to determine whether the results and conclusions are sensitive to the model choice.

—Paul D. Allison

See also Confidence Interval; Cox Model; Dummy Variable; Logistic Regression; Regression

Further Readings

Allison, P. D. (2001). *Missing data.* Thousand Oaks, CA: Sage.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* New York: CRC Press.

Molecular Epidemiology

Although the recognition of molecular epidemiology as an epidemiology subspecialty is relatively recent, laboratory methods have long been used to classify disease and determine exposure in epidemiologic studies. For example, the ability to detect and identify bacteria was essential to the success of studies illuminating the epidemiology of typhoid by Wade Hampton Frost and others, as was the ability to measure blood lipids in the identification of an association between cholesterol and heart disease risk in the Framingham study. The distinction implied in the term *molecular epidemiology* arises from the challenges and opportunities of applying the rapidly expanding array of modern molecular techniques to studies of health and disease in populations. Modern molecular techniques include the ability to directly study genes (genomics), gene expression (transcriptomics), proteins (proteomics), and metabolites left behind by cell processes (metabolomics). These techniques are applied in the rapidly expanding number of epidemiology specialties, most notably genetic, cancer, environmental, and infectious disease epidemiology.

The application of molecular techniques to epidemiology gives epidemiologists the tools to move beyond risk factor epidemiology and gain insight into the overall system of the disease. For infectious diseases, the system includes the transmission system, pathogenesis, and virulence of the agent. The inclusion of molecular tools has the potential to enhance diagnosis of outcome, and to detect low levels of exposure or markers of previous exposure, decreasing misclassification of outcome and exposure. Molecular epidemiologic studies can identify previously undetectable agents, enhance outbreak investigation, help describe disease transmission systems, and give insight into pathogen gene function and host-agent interaction. When applied in combination with appropriate epidemiologic methods, modern molecular techniques allow us to identify novel methods of disease prevention and control, markers of disease diagnosis and prognosis, and fertile research areas for potential new therapeutics and/or vaccines. However, the success of these studies depends not only on the molecular measure chosen, but also on whether the strengths and limitations of the chosen measure are considered in the design, conduct, and analysis, and interpretation of the study results.

Examples in this entry focus on applications in infectious disease epidemiology, which provide the additional challenges and opportunities of at least two genomes (sets of transcripts, proteins, and metabolites), that of the infectious agent and that of the host(s). However, the general principles described are applicable to all epidemiologic studies that incorporate molecular techniques.

Opportunities

The potential of modern molecular techniques in epidemiologic studies has been best explored with genomics. The promise of transcriptomics, proteomics, and metabolomics for increasing understanding of the distribution of health and disease in human populations is great, but at this writing, not much realized for infectious diseases. Thus, the opportunities for applying molecular tools to epidemiologic studies in infectious diseases described below involve primarily the use of genomic techniques.

Gaining Insight Into Gene Importance and Function

Many major human pathogens have been genetically sequenced, and hundreds of genomes will be sequenced in the near future. Unlike genes from multicelled organisms, single-celled organisms often vary greatly in genetic content and expression—that is, genes may be present or absent as well as expressed or silent. Once a single strain of an infectious agent has been sequenced, the sequence can be used as a reference for comparison with others in the same species, providing insight into the heterogeneity of the species. Sequence information can also be mined to identify risk factor genes of unknown function that are structurally similar to genes whose function is known, thereby giving insight into the function of these risk factor genes. Epidemiologic screening of collections of infectious agents for the prevalence of genes that alter the transmission, pathogenesis, and virulence of the agent can provide important insights into the potential importance and putative function of those genes. For example, a gene found more frequently in strains causing severe disease (virulent strains) than among strains that colonize without causing symptoms (commensal strains) suggests that the gene is worthy of more detailed laboratory analyses of its function.

Transcriptomics extends and refines this concept by enabling the identification of genes that are expressed rather than genes that are simply present in the genome. These technologies may be used to identify which genes an infectious agent expresses at different stages of pathogenesis. Expression profiling adds to our knowledge of disease pathology by separating the mechanisms associated with initiation of infection from those associated with disease progression. Furthermore, transcriptomics has the potential to greatly enhance our understanding of interactions between the infectious agent and the human host. For example, in a model system, we can identify which genes are expressed in response to infection. Understanding how virulence is regulated in vivo can point to new targets for therapeutics or vaccines.

Determining a Molecular "Fingerprint"

Infectious agents can be classified into unique groups based on direct or indirect measures of genetic sequence, yielding what is known as a *molecular fingerprint.* Molecular fingerprints for infectious agents are less subject to misclassification than typing systems based on phenotype, which may vary with growth conditions. For example, bacteria that are genetically identical may have different appearance (“morphology”) depending on growth conditions. Furthermore, extrachromosomal material, such as plasmids, that code for antibiotic resistance or other characteristics used in typing may be gained or lost during storage and handling. Typing systems based on chromosomal material often focus on genetic regions that are less likely to develop mutations, known as highly conserved regions.

Several molecular typing techniques for infectious agents are based on genetic sequence. Techniques range from gel-based methods such pulsed-field gel electrophoresis (PFGE) used by the Centers for Disease Control and Prevention in PulseNET, PCR-based methods based on repeated elements, to sequencing conserved areas of the genome, such as multilocus sequence typing, to comparing entire genomes with that of a reference sequenced strain, known as genomotyping. These methods vary in cost, reliability, and adaptability to high throughput formats. When properly applied, molecular typing allows the investigator to confirm that strains are identical at the genetic level, generate hypotheses about epidemiologic relationships between strains in the absence of epidemiologic data, and describe distribution of strain types and identify the determinants of that distribution.

A primary application of molecular fingerprinting is in outbreak investigations. Molecular fingerprints confirm or refute epidemiologic hypotheses, for example, confirming that a particular food item is the common source for a widely disseminated foodborne outbreak. Molecular techniques also are used to confirm transmission events, particularly when epidemiologic data suggest limited contact, thus giving us new information about potential transmission modes. As part of a surveillance system, the identification of common molecular fingerprints can suggest potential epidemiologic linkages requiring further investigation. For example, there are sporadic cases of *Escherichia coli* O157:H7 that can occur in space-time clusters. If these cases share a common molecular fingerprint, an outbreak investigation is in order. Isolates from multiple clusters in different areas occurring in the same time period with a common molecular fingerprint can suggest a common source outbreak from a widely disseminated vehicle, such as occurred with spinach in 2006. Molecular fingerprints can distinguish between

new infection and recurrence of existing infection, such as studies that confirmed exogenous reinfections with tuberculosis. Comparing the distribution of transposable genetic elements among different strain types can provide insight into whether the spread of factors coded for by genes in the transposable elements is primarily clonal or due to horizontal gene transfer. Understanding these types of evolutionary relationships is key to designing effective prevention and control strategies.

Identifying Previously Unknown or Uncultivable Infectious Agents

The vast majority of infectious agents cannot be cultured using standard laboratory techniques; thus, we have only limited knowledge about the numbers and types of bacteria, viruses, and fungi that live with us and in us. The ability to make a copy of genetic material and determine the genetic sequence, which can then be compared with known genetic sequence, has led to a radical reassessment of the amount of life around us, and has facilitated the identification of previously unknown infectious agents. The identification of the causative agent of HIV and Kaposi's sarcoma are attributable to the use of modern molecular techniques combined with epidemiologic principles. A great triumph of this technique was the identification of human papillomavirus as the cause of cervical cancer, leading ultimately to the development of an effective vaccine.

Understanding Host-Agent Interactions

Infectious diseases are often classified into those that primarily affect humans and those that primarily affect nonhumans. As we have increased our understanding of the genetics of human pathogens, there is increasing evidence that subsets of a particular human pathogen may be better adapted to human hosts with certain genetic characteristics. For example, there is some evidence that variants of tuberculosis from particular geographic areas spread more easily among persons from the same geographic area. Strains of human papillomavirus seem to persist longer among individuals whose genetic origin is the same as the origin of the virus: Variants of African origin persist longer among African Americans, and variants of European origin persist longer among European Americans. As we refine our understanding of both human and pathogenic genomes, we might discover answers to why some individuals suffer chronic, recurring infections and others are apparently resistant. These answers will provide important insights for the development of prevention and control strategies.

Implications of Adding Molecular Techniques to Epidemiologic Studies

The choice of a molecular technique should not be made lightly, because the choice not only affects study conduct and analysis but also can profoundly affect study design efficacy. A technique might be a direct or indirect measure of the exposure or outcome of interest. It might be able to characterize disease stage or history of exposure. Each molecular technique has its own reliability, validity, sensitivity to specimen collection and processing, and cost, which affect its suitability for a particular project. Furthermore, one must consider the ability of the molecular technique to capture the desired level of variance within and between individuals. For example, some hormone levels may vary greatly through out the day in one individual, while levels of another biological marker may vary little between individuals over long periods of time.

General Comments on Selecting a Molecular Tool

In some ways, selecting a molecular measuring tool is no different from choosing the appropriate way to measure any item of interest in an epidemiologic context. A list of considerations is shown in Table 1. Because many molecular tools are sensitive to variance in technique, instruments, reagents used, and technician skill, the reliability and validity of any selected molecular tool will be different in different laboratories and should be determined prior to beginning any studies. If there is more than one laboratory, the investigator should also assess interlaboratory variation in addition to intralaboratory variation.

Some methods, such as gel-based typing methods, can be so sensitive to individual laboratory conditions that gel-to-gel variation makes it difficult to compare gel results between gels from the same laboratory, much less between different laboratories. Although running standards and using software that normalizes to those standards increase comparability, assessments

Table 1 Considerations When Choosing Between Molecular Typing Systems

Validity	High sensitivity (the probability that test is positive given that the sample is truly positive) and high specificity (the probability that test is negative given that the sample is truly negative)
Reliability	The assay is both repeatable: (same result in the same laboratory under conditions) and reproducible: same result in different laboratory
Transportability	Results are easily transported and compared between laboratories
Level of Discrimination	The number of categories that result from testing. Whether categories have biologic rationale such that categories can be collapsed in an interpretable meaningful way
Rapidity	The results will be available in a timely manner for the desired investigation
Cost	The resulting measures answer the question for a reasonable cost

of identity of bacterial strains using even gold-standard, gel-based typing methods such as PFGE are best made by running the strains on the same gel. This is true for all methods based on comparing visual band patterns. In contrast, sequence information can be ascertained with extremely high reliability and validity. Furthermore, sequence information is easily recorded and compared.

Other methods can be sensitive to how a specimen is collected, the media in which the specimen is stored, the time since collection, and the length and temperature of storage. The sensitivity may be a function of the specimen itself, as some molecules can degrade quickly, such as bacterial RNA. The specimen may include other material that interferes with the assay or degrades the substance of interest, so that considerable processing may be necessary prior to testing and/or storage. Some bacteria grow, albeit slowly, even when frozen, such as *Listeria* species. Therefore, a pilot study of all laboratory procedures is essential to identify any such problems prior to implementing the study.

Even valid and reliable methods vary in their level of discrimination, that is, the number of categories that result from a measurement. The same tool can be more or less discriminatory depending on how it is applied. For example, PFGE is not very discriminatory if the chosen restriction enzyme cuts the DNA into only one or two bands, but it is highly discriminatory if 30 bands result. The investigator needs to decide what level of information is required to answer the research question. While it may be attractive to use the most recently developed technique, a simpler technique may give the answer at the level required at a much cheaper cost. Furthermore, the type of inferences the investigator wishes to make is an important consideration. Multilocus sequence typing was developed to study genetic lineages, which occur over thousands of years; thus, it may not be appropriate for determining genetic linkage in an outbreak, where genetic changes not captured by multilocus sequence typing may distinguish outbreak from nonoutbreak strains.

The investigator should also consider whether there is a biologic rationale for collapsing the separate categories derived from a highly discriminatory technique. A change in number of bands in a gel-based method may result from a variety of different events, some of which are relevant to the question at hand, some of which are not. For example, band patterns in PFGE may be the same but the genetic content of the bands may vary. Whether categories can be collapsed may depend on epidemiologic information. For example, a one- or two-band difference in PFGE pattern may be considered the same organism in the context of an outbreak investigation, but not when comparing a large collection of organisms collected over time and space. Any technique based on band patterns, such as PFGE, can exhibit homoplasia, meaning that the same pattern can evolve by chance in two different groups. The probability of homoplasia increases as larger geographical and temporal frames are sampled, which becomes important when examining international or long-standing databases.

Results may be recorded on different types of scales that give the investigator more or less power in

the analysis. The results may be qualitative (a simple positive/negative), nominal (putting results into groups but there is no order to groups), ordinal (putting results into groups where there is an implied order), or ratio (where the distances between points are equidistant and there is a meaningful zero point). Ratio scales give most power for the analysis. Ordinal and ratio scales are preferred, as categories might be reasonably collapsed in the analysis. Collapsing of categories is problematic for nominal variables, such as those that occur with gel-based typing techniques based on band patterns.

The investigator also needs to consider the timeliness of the results. In an outbreak investigation, having a timely answer may be essential, whereas in research settings, a more definitive but less timely answer might be desirable. Finally, cost is an important consideration. The investigator often must trade off the added precision and power that might be gained from using a more expensive test, with the loss of power from enrolling fewer participants.

Study Design

Obtaining biological specimens—especially those that require needle sticks or other procedures that are uncomfortable—often decreases response rates, thus increasing the potential for selection bias, and adversely affecting the study validity and generalizability. Investigators need to stick to sound epidemiologic principles for identifying, recruiting, and following study participants, and for determining sample size, and avoid being seduced by the latest laboratory method. The gain in power from decreased misclassification afforded by using a newer molecular technique can be rapidly offset by the increased cost per unit of the new technique, resulting in a need to reduce sample size.

After these basic design considerations, the investigator should consider how the choice of molecular tool might affect the study design. For example, if a test must be conducted on fresh samples, the sampling of cases and controls in a case-control study should be done such that the groups are sampled and tested in similar time periods to minimize potential biases resulting from assay drift, which is where a method gives increasingly higher or lower results with time. For nested case-control studies, how specimens are collected may determine whether controls can be sampled from the base population (case-based, also called case-cohort, sampling) or at time of incidence disease (incidence density sampling).

Study Conduct

Molecular epidemiologic studies have added complexities associated with the collection and handling of specimens that may contain infectious agents. Study personnel must be protected from infection by vaccination and proper training, and appropriate precautions must be followed at all stages to minimize risk of infecting study personnel or others. There are substantial regulations about shipping and handling of diagnostic specimens and infectious agents; investigators should acquaint themselves with these prior to any data collection. Information about these regulations is available from the Centers for Disease Control and Prevention Web site (www.cdc.gov). Laboratories and facilities must be inspected and appropriate certifications obtained from the appropriate institutional biosafety committee.

Biological Versus Technical Variability

Prior to testing specimens, any assays or techniques should be tested and the reliability and validity in the hands of the investigators' team determined. Assays should be perfected so that any technical variation is minimized. The objective is to minimize variability in results stemming from technical issues so that the true biological variability will be measured. Sometimes, there is substantial diurnal biological variability. In this case, the investigator should consider strategies to minimize the effects of this variability, such as collecting all specimens at the same time of day or pooling specimens over the course of the day. For example, if testing urine for a metabolite, the investigators might use the first urine void of the day, or pool all urine from an individual over the course of a single day.

Quality Control and Quality Assurance

Study protocols should include ongoing quality control and quality assurance procedures to detect any problems with the protocol as the study proceeds, for example, including specimens with known values and/or duplicate specimens when material is shipped as a check that proper shipping procedures are followed. Laboratory personnel should be blinded to

which specimens are standards or duplicates. Assays should be run in duplicate and positive and negative controls included in each run. Equipment needs to be calibrated frequently, and the reliability within and between technicians assessed.

Specimen Collection, Handling, and Storage

Tracking and logging of specimens is essential. The tracking must continue as long as specimens are stored, and items that might affect future assays noted, such as the number of times frozen specimens are thawed. Ideally, specimens are divided into separate vials (aliquoted) and stored in different locations. Aliquoting minimizes contaminations and problems with freezing and thawing. Storage in different locations helps minimize loss from untoward events such as equipment failure due to power losses. Specimen handling should be meticulous to avoid contamination. If testing infectious agents, the investigator should determine if the agent might change following successive cultures or under certain storage conditions. If specimens are to be frozen at low temperatures, it is essential to use labels that will not fall off on extended freezing. The labels should include all information that may help the laboratory technician to minimize error. Alternatively, the investigator can use bar codes linked to a database with all required information.

Recording Data

Specimens may be tested for several different substances, or several different specimens may be obtained from the same individual. If cultured for infectious agents, multiple isolates may be recovered from the same individual. The development of a study identifier that allows linkage to questionnaire information, but also includes information useful about the assay, is essential. For example, the identification number used in the laboratory might also include an indicator of when the specimen was collected, or if separated into components or tested for different substances, an additional indicator of the substance. For example, the first three numbers might indicate the individual, and the fourth number the site of collection (567-2, where 567 is study identification number and 2 indicates collected from urine). The investigator should consider how the data would be used in the analysis, so the various databases can be merged with minimal amounts of data management.

Data Analysis

Epidemiologists are used to using either an individual or person-time as the denominator for estimating rates and proportions. However, individuals may be colonized with multiple strains of a single infection agent or at multiple sites that may or may not be independent. Depending on what relationship(s) the investigator intends to demonstrate, the strain of the infectious agent may be used as the unit of analysis. For example, if the outcome is transmission between couples, the denominator might be couples or the number of isolated organisms, some of which are and some of which are not transmitted.

The investigator might also wish to determine if an individual can be colonized with multiple strains of a single species or if host or agent factors inhibit colonization by multiple strains. For these analyses, the investigator must be aware of sampling errors, that is, the potential that a strain might truly be present and have been missed because of laboratory procedures. If culturing bacteria, an investigator might choose to test only the most common isolate, which precludes testing whether colonization inhibits coinfection with other strains (“super infection”). Multiple viral strains might also be cultured from a single individual, but only if the laboratory procedures are set up accordingly. Potential sampling errors and sensitivity of chosen technique to low levels of the organism of interest should also be taken into account when interpreting prevalence and incidence estimates.

In outbreak investigations, molecular evidence is often used to support or refute epidemiologic evidence gathered via questionnaire or medical records. Thus, the investigators need to be certain that the laboratory technique is appropriately classifying organisms into related groups. If an infectious organism changes rapidly—for example, if it uptakes or loses genetic material that codes for antibiotic resistance—it may still be part of the same chain of infection but may show a remarkedly different molecular type depending on the typing system used. The speed of change depends on the characteristics of the organism itself; thus, investigators should take care to understand the molecular biology of the organism under study, as well as the strengths and limitations of any typing system when drawing inferences.

Future Opportunities and Challenges

As researchers gain experience in the application of molecular tools to epidemiologic studies of human pathogens, there is tremendous opportunity to increase understanding of the transmission, pathogenesis and evolution of human pathogens, and the interaction of human and pathogen genes. But many challenges lie ahead before this potential will be fully realized.

First, we will need to increase our understanding of the ecology of microorganisms that are normal inhabitants of the human bowel, skin, vaginal cavity, nose, and throat. Although we generally consider human pathogens in isolation, transmission and pathogenesis are not solely a function of a particular pathogen but are often modified by the presence of other pathogens. For example, the presence of genital ulcer disease increases transmission of HIV. A less obvious example is that bacterial vaginosis also increases transmission of HIV. Bacterial vaginosis is a disruption of the vaginal ecology that may or may not cause clinical symptoms. Similar disruptions of the ecology of other human microflora may be associated with increased disease risk. However, as of this writing, our understanding of the ecology of the human microflora is limited.

Second, current analytic methods are inadequate to deal with the complicated interactions inherent in molecular epidemiologic studies of agent-host interactions. We have the capacity to generate large amounts of data on the genetics of infectious agents, the expression of these genes during different aspects of the infection process, and the expression of human genes in response to infection. Thus, description is the order of the day. However, to truly advance the field, we will need to develop more complete theories to explain our observations. Molecular epidemiology of infectious diseases is one area of epidemiology that has obvious ties to evolutionary theory. Evolutionary theory is highly developed and there are associated analytic methods that we have just begun to apply in an epidemiologic context.

The second problem is tied to a final problem: Our understanding of the system is so limited that we fail to collect the correct data. This challenge can be addressed by the development of appropriate mathematical models. Even a simple conceptual model forces the investigator to explicitly specify the relationships between various aspects of the system, providing insight into what we do and do not know, and what additional data are lacking. Mathematical models also provide insight into the relative importance of various aspects of the system, and can take into account known theories and thus help develop new theories to explain the transmission, pathogenesis, and evolution of infectious diseases.

—Betsy Foxman

See also Genetic Epidemiology;Genomics; Outbreak Investigation

Further Readings

Bender, J. B., Hedberg, C. W., Besser, J. M., Boxrud, D. J., MacDonald, K. L., & Osterholm, M. T. (1997). Surveillance by molecular subtype for *Escherichia coli* O157:H7 infections in Minnesota by molecular subtyping. *New England Journal of Medicine*, *337*(6), 388–394.

Blahna, M., Zalewski, C., Reuer, J., Kahlmeter, G., Foxman, B., & Marrs, C. F. (2006). The role of horizontal gene transfer in the spread of trimethoprim-sulfamethoxazole resistance among uropathogenic *Escherichia coli* in Europe and Canada. *Journal of Antimicrobial Chemotherapy*, *57*(4), 666–672.

DeRiemer, K., & Daley, C. L. (2004). Tuberculosis transmission based on molecular epidemiologic research. *Seminars in Respiratory and Critical Care Medicine*, *25*(3), 297–306.

Foxman, B., Manning, S. D., Tallman, P., Bauer, R., Zhang, L., Koopman, J. S., et al. (2002). Uropathogenic *Escherichia coli* are more likely than commensal *E. coli* to be shared between heterosexual sex partners. *American Journal of Epidemiology*, *156*(12), 1133–1140.

Foxman, B., Zhang, L., Koopman, J. S., Manning, S. D., & Marrs, C. F. (2005). Choosing an appropriate bacterial typing technique for epidemiologic studies. *Epidemiologic Perspectives & Innovations*, *2*, 10.

Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., de Jong, B. C., Narayanan, S., et al. (2006). Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(8), 2869–2873.

Gerner-Smidt, P., Hise, K., Kincaid, J., Hunter, S., Rolando, S., Hyytia Trees, E., et al. (2006). PulseNet USA: A five-year update. *Foodborne Pathogens and Disease*, *3*(1), 9–19.

Graham, M. R., Virtaneva, K., Porcella, S. F., Gardner, D. J., Long, R. D., Welty, D. M., et al. (2006). Analysis of the transcriptome of Group A *Streptococcus* in mouse soft tissue infection. *American Journal of Pathology*, *169*(3), 927–942.

Manning, S. D., Tallman, P., Baker, C. J., Gillespie, B., Marrs, C. F., & Foxman, B. (2002). Determinants of

co-colonization with group B streptococcus among heterosexual college couples. *Epidemiology*, *13*(5), 533–539.

Porta, M., Malats, N., Vioque, J., Carrato, A., Soler, M., Ruiz, L., et al. (2002). Incomplete overlapping of biological, clinical, and environmental information in molecular epidemiological studies: A variety of causes and a cascade of consequences. *Journal of Epidemiology and Community Health*, *56*, 734–738.

Rundle, A., Vineis, P., & Ashsan, H. (2005). Design options for molecular epidemiology research within cohort studies. *Cancer Epidemiology, Biomarkers, & Prevention*, *14*(8), 1899–1907.

Tenover, F. C., McDougal, L. K., Goering, R. V., Killgore, G., Projan, S. J., Patel, J. B., et al. (2006). Characterization of a strain of community-associated methicillin-resistant *Staphylococcus aureus* widely disseminated in the United States. *Journal of Clinical Microbiology*, *44*(1), 108–118.

Trottier, H., & Franco, E. (2006). The epidemiology of genital human papillomavirus infection. *Vaccine*, *24*(Suppl. 1), S1–S15.

Vineis, P., Schulte, P., & Vogt, R. (1993). Technical variability in laboratory data. In P. Schulte & F. Perera (Eds.), *Molecular epidemiology. Principles and practices.* San Diego, CA: Academic Press.

Zhang, L., Foxman, B., Manning, S. D., Tallman, P., & Marrs, C. F. (2000). Molecular epidemiologic approaches to urinary tract infection gene discovery in uropathogenic *Escherichia coli*. *Infection and Immunity*, *68*(4), 2009–2015.

Monitoring the Future Survey

Monitoring the Future (MTF) is an annual survey of the attitudes, values, and behaviors of a nationally representative sample of 15,000 to 19,000 American high school students and young adults. It is conducted by the Survey Research Center of the University of Michigan, with funding from the National Institute on Drug Abuse of the National Institutes of Health. MTF began collecting data, originally on 12th graders only, in 1975; since 1991, 8th and 10th graders have also been surveyed. Currently, approximately 50,000 students in the 8th, 10th, and 12th grades are surveyed each year. Participation rate has been 66% to 85% over all years of the study. In addition, follow-up mail questionnaires are sent biannually to a randomly selected sample from each senior class.

The primary focus of MTF is the use and abuse of tobacco, alcohol, and drugs by young adults and their perceptions and attitudes toward these substances. The MTF consists of two parts: core questions that include demographic information and basic questions about substance use, which are asked of every respondent, and ancillary questions on a variety of topics such as social and political attitudes, health behaviors, and educational aspirations, which are administered to different subsamples of respondents through the use of different questionnaire forms. MTF data, documentation, and supporting materials are available from the ICPSR and Monitoring the Future Web sites listed below.

Most MTF data are gathered through an annual cross-sectional survey of students currently attending school in the 8th, 10th, or 12th grades. These data are collected through self-administered questionnaires filled out by individual students, usually during a normal class period at their school. The survey administration is supervised by University of Michigan staff members and data are not shared with either the students' parents or school officials. Questionnaire forms are optically scanned and stored as an electronic data file.

MTF uses a probability sample design with three selection stages:

1. Broad geographic area (Northeast, North central, South, or West)
2. Schools or linked groups of schools within a geographic area
3. Students within schools—if a school has less than 350 students in the relevant grade, all are selected; if there are more than 350, participants are randomly selected

Schools who decline to participate are replaced with schools similar in type (public, Catholic or private/non-Catholic), geographic location, and size. Specific questionnaire forms (six were used in 2004) are administered in an ordered sequence so a nearly identical subsample of students completes each form.

A longitudinal component was added to the MTF in 1976: Since then, a random sample of about 2,400 students from that year's 12th-grade participants has been selected to participate in follow-up surveys. These participants are divided into two groups, who are mailed questionnaires in alternating years (so half the participants receive a questionnaire in odd-numbered years following 12th grade, i.e., Years 1, 3, 5, and so on, while half receive follow-up questionnaires in

even-numbered years. Retention for the first year of follow-up averages 77%.

The greatest strength of MTF is the availability of data on the same questions over multiple years and the use of scientific sampling procedures to allow the computation of nationally valid estimates of responses. This allows researchers to address questions such as the prevalence of tobacco use among 12th graders and how that number has changed over the years. The large number of questions included in each survey, the extremely detailed examination of substance use, and the inclusion of a range of other types of questions also increases its usefulness. The most obvious limitation is that the MTF sample is not representative of all young Americans in the age groups included, only of those attending school: Young people who are home schooled, have dropped out, or are not attending school for some other reason (such as health problems or incarceration) are not included in the sample, and there is every reason to believe that they would differ systematically from young people attending conventional schools.

—*Sarah Boslaugh*

See also Alcohol Use; Child and Adolescent Health; Drug Abuse and Dependence, Epidemiology of; Tobacco

Further Readings

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2006). *Monitoring the future national results on adolescent drug use: Overview of key findings, 2005* (NIH Publication No. 06-5882). Bethesda, MD: National Institute on Drug Abuse.

Web Sites

Monitoring the Future (MTF) Series: http://webapp.icpsr.umich.edu/cocoon/ICPSR-SERIES/00035.xml.
Monitoring the Future: A Continuing Study of American Youth: http://www.monitoringthefuture.org

Morbidity and Mortality Weekly Reports

See Centers for Disease Control and Prevention

Mortality Rates

Mortality rates (synonym: death rates) are used to quantify the tendency to die in a population during a given time period. Since death can be considered as the utmost form of "unhealthiness" or "disease," mortality rates are major (inverse) health indicators. Because measuring morbidity is often difficult, mortality rates or mortality-based indicators like life expectancy remain the major indicators used to ascertain the level of health in a society or social group.

Mortality rates always refer to a time period, usually 1 year, though monthly or even daily mortality rates are sometimes calculated for particular situations or processes—a war, epidemics, and so on. The annual crude death rate (m) can be thought as the proportion of the population dying during 1 year—presently not much above or below 1% throughout the world—and is computed by dividing the count of deaths during the year, d, by the total population or "population at risk," p, generally approximated by the population at mid year, and expressing the result per thousand or, more generally, per some power of 10. Therefore, $m = (\mathrm{d/p}) \times 10^k$, and $k = 3$, if the rate is expressed per 1,000.

Mortality rates can be conceptualized as an approximate measure of the probability of death during a given period of time in members of the group for which the rate is calculated, or as a special type of incidence rate, where the "disease" is death. If we know that 48,700 deaths occurred over 2 years in a population of 2.1 million people, we can estimate the annual mortality rate during that period as approximately 11.6 per 1,000, since (48,700/2)/2,100,000 $= 0.011595$. However, when computing death rates for small groups, for instance, in a longitudinal study or a clinical trial, it is usually needed to consider properly the exact period of observation, expressing the rate per person-time units, and taking into account that those who die are no longer "at risk" for death. If 100 persons at the start of the observation period were followed for 3 years, during which 5 persons died, 2 at the end of the first year, and other 3 when 1.3, 2.2, and 2.6 years had passed, we have exactly

$$1.0 \times 2 + 1.3 \times 1 + 2.2 \times 1 + 2.6 \times 1 + 3 \times 95 = 293.1 \text{ person-years of observation.}$$

$$1 \times 2 = 2 + 1.3 = 3.3 \times 1 = 3.3 \times 1 = 3.3 + 2.2$$
$$= 5.5 \times 1 = 5.5 + 2.6 = 8.1 \times 1 = 8.1 + 3$$
$$= 11.1 \times 95 = 1054.5.$$

Since $5/293.1 = 0.0171$, the mortality rate can be expressed as 0.0171 deaths per person-year, or 17.1 deaths per 1,000 person-years, or as an annual death rate of 17.1 per 1,000.

An age-specific mortality rate is a death rate in a given age stratum. If the subindex i refers to the particular age stratum, the age-specific mortality rate will be $m_i = 10^k \times d_i/p_i$, where m_i is age-specific mortality in the age stratum i, d_i are total deaths in the age stratum i, and p_i is the population in that age.

Since the total death count in a population is the sum of all deaths in all age strata,

$$\begin{aligned} m = d/p &= [d_1 + d_2 + d_3 + \cdots + d_k]/p \\ &= [p_1 m_1 + p_2 m_2 + p_3 m_3 + \cdots + p_k m_k]/p \\ &= [p_1/p]m_1 + [p_2/p]m_2 + [p_3/p]m_3 \\ &\quad + \cdots + [p_k/p]m_k \\ &= s_1 m_1 + s_2 m_2 + s_3 m_3 + \cdots + s_k m_k, \end{aligned}$$

where s_i is the proportion of the whole population living in the age stratum i and, therefore, $s_1 + s_2 + s_3 + \cdots + s_k = 1$. In compact notation,

$$m = \sum_{i=1}^{k} s_i \cdot m_i.$$

This means that the crude death rate is a weighted average of the age-specific death rates, with the weights being the shares of each age stratum in the whole population.

Since the probability of death is much higher during the first year of life than during the other years of childhood and then grows exponentially with age (Table 1), the crude death rate is strongly affected by the age structure of the population (the weights s_i), and will be large in societies with excellent health conditions but a high proportion of elderly in the population. For this reason, the crude mortality rate is not a good health indicator.

Age-specific, sex-specific, or age-and-sex-specific mortality rates are often computed to gauge health conditions in specific demographic groups. For instance, in 1990, 2,573 females aged 55 to 64 died in Sweden, out of a total of 429,914 females in this age group, which makes for a specific mortality rate of 6.0 deaths per 1,000 females aged 55 to 64, compared with 11.2 deaths per 1,000 males in the same age group. Save exceptional circumstances, male mortality is larger than female mortality in each age stratum (Table 1). Since differences in age-specific or age-and-sex-specific mortality rates cannot be caused by differences in age distribution, they can be used to compare health conditions across time in a given geographical region or across geographical regions in the same point in time. For example, mortality rates in females and males aged 55 to 64 in the United States were in 1990, 8.8 and 15.6 per 1,000, respectively, compared with much lower rates, 6.0 and 11.2, observed for females and males in that same age stratum in Sweden.

Table 1 Age-and-Sex-Specific Death Rates (per 1,000) in the United States in 1995

Ages	*Males*	*Females*
Below 1	8.44	6.90
5 to 14	0.27	0.18
15 to 24	1.41	0.48
25 to 34	2.10	0.80
55 to 64	14.17	8.41
75 to 84	73.77	48.83

Source: U.S. Census Bureau (2005).

When mortality levels of entire populations across time or across regions need to be compared, the influence of the age-structure of the population is excluded through age-adjustment (synonym: age-standardization). Two methods of age-adjustment are used, direct and indirect adjustment. In the direct method, the most commonly used, the age-specific mortality rates of population and year of interest are applied to the age structure of a standard population (Figure 1).

If population A has the age-specific mortality rates $m_1, m_2, m_3, \ldots, m_k$, and the population proportions $s_1, s_2, s_3, \ldots, s_k$ in k age intervals, and population B has the age-specific mortality rates $\rho_1, \rho_2, \rho_3, \ldots, \rho_k$, and the population proportions $\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_k$, then the age-adjusted mortality rate of population A with population B as standard is

$$m_1\sigma_1 + m_2\sigma_2 + m_3\sigma_3 + \cdots m_k\sigma_k,$$

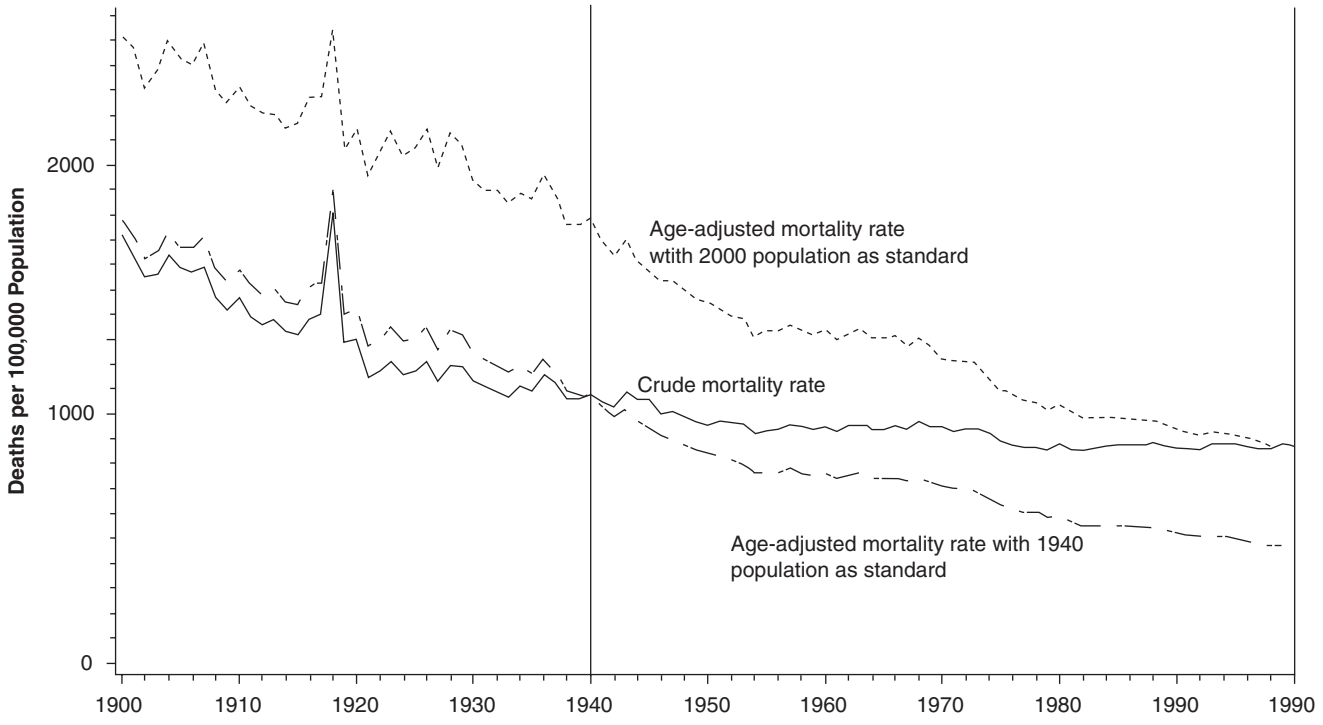


Figure 1 Mortality Rates in the United States Throughout the 20th Century

Source: U.S. Census Bureau (2007).

that is, $\sum_{i=1}^{k} m_i \cdot \sigma_i$. Using the 1940 U.S. population as a standard, in 1990 the age-adjusted mortality rate was 5.2 per 1,000 for the general population of the United States, and 4.9 and 6.9 per 1,000 for whites and nonwhites, respectively. Age-adjusted mortality rates are often used to compare health conditions between different nations, regions, ethnic groups, or social classes. In every country in which age-adjusted or age-specific mortality has been compared among social classes, a gradient has been found with mortality increasing when going from high to low levels of income, wealth, education, or other social class indicator.

Since estimating the number of infants (and even children in poor countries) is difficult, the infant mortality rate is computed by dividing the annual death count of infants (children less than 1 year old) by the annual count of live births, and multiplying the result by 1,000, so that infant mortality is expressed usually as a rate of infant deaths per 1,000 live births. A similar rationale applies to the under-5 mortality rate, often called child mortality rate, computed as deaths below 5 per 1,000 live births. Neither the infant mortality rate nor under-5 mortality rate are age-specific death rates in a strict sense, since the denominator of the rate is not a population count. However, both infant mortality and child mortality are good summary indicators of population health.

Cause-specific mortality rates are computed by dividing counts of deaths attributed to specific causes by the size of the specific group considered. Therefore, they can be age specific, sex specific, age-and-sex specific, and so on. A particular type of cause-and-sex-specific mortality rates is maternal mortality rates, which are referred to "all deaths ascribed to deliveries and conditions of pregnancy, childbirth, and the puerperium." Maternal mortality is much higher in poor countries and is usually sensitive to the availability of obstetric care, as well as the legal status of abortion and the social condition of women.

Because of the strong decline in deaths caused by infectious disease at all ages and mostly in infancy and childhood, age-specific mortality rates secularly declined in all countries in the world. This process started about one and a half or two centuries ago in Western Europe, a little bit later in North America

(Figure 1) and the rest of the industrialized world, and only in the past century in most countries of Latin America, Asia, and Africa. Except in very poor countries where infectious diseases such as malaria, tuberculosis, and AIDS are presently major killers, in most countries of the world the major causes of death and, therefore, the largest cause-specific death rates are presently illnesses of the heart and the circulatory system, malignancies, and injuries, mostly related to traffic.

—José A. Tapia Granados

See also Incidence; Life Expectancy; Person-Time Units; Prevalence; Rate; Ratio

Further Readings

Division of Health Statistics, World Health Organization. (1980). *Manual of mortality analysis: A manual on methods of analysis of national mortality statistics for public health purposes*. Geneva, Switzerland: Author.

Gordis, L. (2000). *Epidemiology* (2nd ed., pp. 42–56). Philadelphia: Saunders.

Lilienfeld, A. M., Lilienfeld, D. E., & Stolley, P. D. (1994). *Foundations of epidemiology* (3rd ed., chaps. 4 and 5). New York: Oxford University Press.

U.S. Census Bureau. (2005). *Statistical abstract of the United States*. Retrieved July 26, 2007, from http://www.census.gov/prod/www/statistical-abstract-1995_2000.html.

U.S. Census Bureau. (2007). *Statistical abstracts*. Retrieved July 26, 2007, from http://www.census.gov/prod/www/abs/statab.html.

Multifactorial Disorders

See Genetic Disorders

Multifactorial Inheritance

The multifactorial inheritance model applies to diseases that depend on multiple genetic loci (polygenic) and the additional contribution of environmental factors. Multifactorial diseases are the result of the interplay of multiple environmental risk factors with more than one gene, where these multiple genes are viewed as susceptibility genes. In this model, genes may increase an individual's susceptibility to a particular disease, but the actual expression of the disease depends on the extent to which the individual encounters certain environmental exposures during embryogenesis or throughout his or her life.

Multifactorial diseases include birth defects such as neural tube defects, developmental disabilities such as autism and common adult-onset diseases such as cancer, diabetes mellitus, hypertension, and heart disease. In fact, most geneticists and epidemiologists today believe that the vast majority of diseases are inherited in a multifactorial fashion or their outcomes are influenced by multiple genetic and environmental factors. For example, the phenotypic outcomes in seemingly straightforward single gene or monogenic disorders (such as the classic autosomal recessive disorder, phenylketonuria, or PKU) are increasingly viewed from a multifactorial perspective as a result of new evidence of the complex relationship between genotype and phenotype in PKU.

Multifactorial diseases are labeled non-Mendelian because they do not exhibit the typical pedigree patterns we observe in Mendelian or monogenic disorders that depend on genotypic expression at one genetic locus (e.g., a recessive disorder such as sickle cell disease). An individual is at an increased risk for developing a multifactorial disease if one or more of his or her relatives are affected as well. This risk is greatest among first-degree relatives. However, patterns of multifactorial conditions are less predictable than diseases that are caused by single gene mutations. For example, multifactorial diseases such as asthma, coronary heart disease, and diabetes mellitus are heterogeneous in their etiology. This means that while the final phenotypic outcome of each disease is similar among individuals, each disease is really a group of diseases, with each subtype having variable genetic and environmental causes.

The complex etiology of multifactorial diseases cannot be discussed without a basic understanding of polygenic inheritance. Also known as quantitative inheritance, polygenic inheritance refers to an inheritance pattern in which traits are controlled by two or more genes, each having a small effect on phenotypic expression. These genes contribute some cumulative effect on the phenotype. These effects may be simple, as in a specific gene adding or subtracting from the measured value of the phenotype, or more complex when other genes or environmental factors influence the phenotype. Unlike basic Mendelian traits, polygenic traits show a continuous distribution of phenotypic values in a population and display a bell-shaped,

or Gaussian, curve when their frequency distribution is plotted on a graph. Polygenic characteristics in humans include height, skin color, dermatoglyphics, blood pressure, head circumference, and intelligence. For continuous traits such as height and head circumference, abnormality is defined arbitrarily (usually 2 *SD* over or below the mean).

The polygenic threshold theory attempts to explain how the inheritance of continuous traits may be used to conceptualize the occurrence of dichotomous characters such as the presence or absence of a birth defect. According to the polygenic threshold theory, an individual has a certain degree of polygenic susceptibility, or genetic liability, for a particular disease. Theoretically, this susceptibility is normally distributed within populations, with individuals displaying varying degrees of genetic susceptibility. Individuals who exceed the critical threshold value for susceptibility to a disease will develop it, while those below this value will not.

The polygenic threshold theory is applied in genetic counseling to explain the observed patterns of recurrence risk for multifactorial diseases among family members. Individuals who are affected with a multifactorial condition have genotypes composed of many high-susceptibility alleles. Because they share more genes in common with the affected individual, close relatives will be more likely to fall above the threshold of susceptibility than unrelated individuals, and hence will be more likely to develop the disease themselves.

As mentioned previously, the recurrence risk for multifactorial diseases does not follow basic Mendelian patterns. For example, among parents who are carriers for a recessive disorder such as sickle cell disease, the risk of having an affected pregnancy is 1 in 4, or 25%. On the other hand, the recurrence risk for parents who have a child with a multifactorial disease typically falls around 3% to 5%. While data among populations and families vary, this risk generally increases if more than one member of the family is affected with the multifactorial disease in question.

For multifactorial diseases to develop, genetic predispositions must interact with some environmental exposure or trigger, or perhaps an array of environmental factors over the entire course of an individual's life. In addition to genetic liability, an individual's environmental liability must also be taken into account when attempting to explain the etiology of multifactorial diseases. In theory, genetically susceptible individuals will not develop a multifactorial disease until they exceed the threshold values for exposure to environmental risk factors. For example, specific genes have been identified as risk factors for many birth defects in infants. However, the vast majority of, if not all, birth defects are now thought to result from the combination of both genetic and environmental factors. The most well-documented environmental factor in the etiology of birth defects is the role of maternal folic acid intake in the risk of neural tube defects (NTDs). Studies have shown specific genes in the developing infant that increase the risk of developing an NTD. Nevertheless, women can significantly reduce the risk of occurrence of a child with an NTD by consuming adequate amounts of folic acid both prior to conception and during pregnancy.

There are no tests to determine the genetic predispositions of individuals to most multifactorial diseases. The best approach for combating multifactorial diseases is to identify and modify the environmental factors that interact with susceptible genotypes. As of yet, the environmental risk factors for many multifactorial diseases are still unknown. In addition, it is unclear how environmental factors interact with specific genes to cause disease. Clearly, future research is needed to identify (1) the underlying genes that predispose individuals or populations to multifactorial diseases, (2) the environmental factors that increase or decrease the expression of predisposing genotypes, and (3) the complex mechanisms by which gene-gene and gene-environment interactions result in multifactorial diseases.

—Cynthia Taylor and F. John Meaney

See also Gene; Gene-Environment Interaction; Genetic Counseling; Genetic Epidemiology; Phenotype

Further Readings

King, Richard, A., Rotter, Jerome, I., & Motulsky, A. G. (2002). *The genetic basis of common diseases* (2nd ed.). New York: Oxford University Press.

Strachan, T., & Read, A. P. (2004). *Human molecular genetics 3* (3rd ed.). New York: Garland Science.

Turnpenny, P., & Ellard, S. (2004). Polygenic and multifactorial inheritance. In *Emery's elements of medical genetics* (12th ed., pp. 137–146). London: Churchill Livingstone.

Multilevel Modeling

Multilevel modeling is the simultaneous use of more than one source of data in a hierarchical structure of units

and is useful for analysis of clustered or longitudinal data. Single-level models are unable to accommodate the characteristics of hierarchical data structures, such as unit of analysis, aggregation bias, state dependency, and within-group and between-group heterogeneity, resulting in misestimations. By accommodating these characteristics, multilevel modeling allows researchers to use data more fully and efficiently and to assess the direction and magnitude of relationships within and between contextual and individual factors.

In the past few years, multilevel modeling, also known as hierarchical linear models, mixed-effects models, random-effects models, and random-coefficient models, has become increasingly common in public health research, partly due to growing interest in social determinants of health and partly due to the growing availability of multilevel statistical methods and software.

Multilevel modeling is a powerful research method that promises to complicate and explicate the field of epidemiology. The inclusion of multiple levels of data in simultaneous analysis models allows for more efficient use of data and provides greater information about within-group and between-group effects and relationships. However, attention must be paid to group-level measurement issues and complex data structures, and better social epidemiology theoretical models are needed to guide empirical research.

With the development of research into social inequalities in health in the late 20th century came an interest in social epidemiology that sought to draw focus away from decontextualized individual characteristics and toward their social or ecological context. Such research can be thought of as encompassing multiple layers surrounding and intersecting with the individual, as illustrated by the multiple layers or levels of the ecological model, including intrapersonal, interpersonal, institutional, community, and policy. At the same time, significant contributions to statistical theory for multilevel methods were being made by Dennis V. Lindley and Adrian F. M. Smith, Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin, and others.

Statistical Model

As in other research fields, notably education in which children are clustered within classrooms and classrooms are clustered within schools, data for epidemiologic analysis are not limited to measurements of the individual, but can include measurements within or outside of individuals. For example, blood pressure and heart rate can be measured within an individual, individuals measured within a family, families measured within neighborhoods, and neighborhoods measured within geographic regions. Similarly, repeated observations are measurements nested or clustered within individuals over time, and also produce a measurement hierarchy. Such hierarchies or clusters, even if random in origin, mean that except at the highest level of measurement there will be subgroups of observations that are similar to each other, because they come from the same group. For example, individuals within families are likely to be more similar to each other on observable and unobservable characteristics than are individuals across families. Such correlations violate the assumption of independence on which most traditional statistical techniques are based, and can result in incorrect standard errors and inefficient parameter estimates. While group level characteristics can be included in single-level contextual analyses of individual outcomes, this method requires group characteristics to be fixed, or averaged, within and between groups. Within-group and between-group heterogeneity are the hallmark of hierarchically structured data and require specialized statistical methods to simultaneously compute within-group and between-group variances.

The mathematical equation for a multilevel model can be thought of as separate equations for each level of measurement. For example, in a two-level analysis of a normally distributed individual outcome predicted by one independent variable measured at the individual level (Level 1), and one independent variable measured at a group level (Level 2), one can write first a single-level equation for individuals in groups (e.g., neighborhoods), and then write a single-level equation for the group effect, as follows:

$$Y_{ij} = b_{0j} + b_{1j}X_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2),$$

where $i = 1,\ldots,n$ individuals, $j = 1,\ldots,n$ groups, Y_{ij} is the individual outcome for the ith individual in the jth group and X_{ij} is the individual (Level 1) independent variable measurement for ith individual in the jth group. Individual-level errors within each group (ε_{ij}) are assumed to be normally distributed with a mean of zero and a variance of σ^2. Regression coefficient b_{0j} is the group-specific intercept and b_{1j} is the group-specific effect of the individual-level variable(s); these vary across groups j.

Next, equations for these group-specific (Level 2) regression coefficients are written using a group-level measurement as the independent variable, as follows:

$$b_{0j} = \gamma_{00} + \gamma_{01} C_j + U_{0j}, U_{0j} \sim N(0, \tau_{00}),$$
$$b_{1j} = \gamma_{10} + \gamma_{11} C_j + U_{1j}, U_{1j} \sim N(0, \tau_{11}),$$

where C_j is the group-level independent variable. The common intercept across groups is γ_{00}, and γ_{01} is the effect of the group-level independent variable on the group-specific intercepts. Similarly, γ_{10} is the common slope associated with the individual-level variable across groups, and γ_{11} is the effect of the group-level independent variable on the group-specific slopes. U_{0j} and U_{1j} are the random error terms for the Level 2 equations, and are assumed to be normally distributed with a mean of zero and variances of τ_{00} and τ_{11}. The Level 2 random error terms allow for the assumption that the group-specific intercepts and slopes are actually random samples from a larger, normally distributed population of group-specific values. Without the random effects, this model becomes a one-level model with averaged fixed group-level effects. Finally, these equations can be combined into one multilevel equation by replacing the regression coefficients in the Level 1 model with the Level 2 equations, as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{01} C_j + \gamma_{10} X_{ij} + \gamma_{11} C_j X_{ij} + U_{0j} + U_{1j} X_{ij} + \varepsilon_{ij}.$$

The parameter estimates resulting from this equation can address questions such as the following: Are individual and group variables related to outcomes when tested simultaneously? Do individual characteristics associated with outcomes vary across groups? How much variation is explained at each level? Estimation of group-level effects and between-group differences depends on the number of groups included at Level 2. Statistical inferences have generally been based on the method of maximum likelihood. More recently, Bayesian methods of inference have been applied to multilevel models. The intraclass correlation provides a measure of the similarity among Level 1 measurements within each Level 2 group.

Multilevel models can be applied to various and complex data structures as dictated by the data or conceptual models. Multilevel models can incorporate multiple independent variables at different levels as well as interaction terms. Models can be expanded to three or more levels, such as repeated blood pressure measures (Level 1) in individuals (Level 2) in clinics (Level 3). The relationships between levels, known as cross-level effects, can be various, including linear, quadratic, or nonexistent in either or both directions. Some hierarchical data are complicated by cross-classification, such that lower-level measures get classified in unexpected or multiple higher-level groups. For example, students may be in multiple classes within the same school, individuals may attend clinics outside their neighborhoods, or people may move or change experimental assignment in the course of a longitudinal study.

The linear random-effects model can be transformed for nonnormal response data, including binary, count, ordinal, multinomial data, and survival analysis. Multilevel models also have been applied to the analysis of latent variables and meta-analytic data. Software commonly used for multilevel analysis includes HLM, MIX, MLWIN, MPLUS, SAS, and recent versions of SPSS.

An Example From Psychiatric Epidemiology

Twentieth-century epidemiologic research in the United States has consistently shown that people who live with mental illness are far more likely to live in poverty than the general population. In 1969, Barbara and Bruce Dohrenwend published a review of available studies of psychiatric epidemiology, *Social Status and Psychological Disorder: A Causal Inquiry*, and reported that the only consistent correlation with severe mental illness they found across different cultures and times was low socioeconomic status. In the 1980s, the National Institute of Mental Health's Epidemiologic Catchment Area (ECA) study conducted psychiatric diagnostic interviews with approximately 20,000 adult respondents (anyone above age 18) in five cities: New Haven, Connecticut; Baltimore, Maryland; St. Louis, Missouri; Durham, North Carolina; and Los Angeles, California. The ECA found that about 20% of respondents had an active mental disorder, with a lifetime reported prevalence of 32%. In this study, higher prevalence of active mental disorder was associated with being African American, being unemployed, and other measures of social class. Bruce, Takeuchi, and Leaf (1991) used a measurement of individual poverty status in a multivariate logistic regression analysis of the 12-month incidence

of mental disorders. They found individual poverty to be associated with greater likelihood of any mental disorder, adjusted odds ratio (95% confidence interval) = 1.92 (1.12, 3.28); and greater likelihood of major depression, adjusted odds ratio (95% confidence interval) = 2.29 (1.19, 4.43).

Over a decade later, Goldsmith, Holzer, and Manderscheid (1998) analyzed ECA data in conjunction with the neighborhood data from the 1980 decennial census. In addition to individual characteristics, their one-level logistic regression models included what they called social area dimensions of neighborhoods. These dimensions were social rank or economic status (median household income of census tract), family status (percentage of households with husband-wife families), residential lifestyle (percentage of single dwelling units), and ethnicity (90% or more white, mixed, and 90% or more minority). Individual risk factors included age, gender, marital status, race, and education. They found that controlling for individual characteristics, only living in a low economic status neighborhood (compared with medium or high status) was statistically significantly associated with greater likelihood of having a 12-month mental disorder (prevalence), adjusted odds ratio (95% confidence interval) = 1.43 (1.18, 1.74); living in a majority ethnic minority neighborhood (compared with mixed or majority nonminority) was associated with greater likelihood of having a past mental disorder. By assessing the magnitude of statistically significant odds ratios and changes in model fit, Goldsmith and colleagues posited that with the exception of economic status, inclusion of neighborhood characteristics contributed little to individual-level explanations of psychiatric epidemiology.

Like Goldsmith et al., Silver, Mulvey, and Swanson (2002) combined ECA and census data to develop a multilevel model of the 12-month prevalence of specific mental disorders, including schizophrenia and major depression. In this analysis, nine census tract measures of neighborhood structure were selected: (1) percentage of persons living below the poverty line; (2) percentage of husband-wife families; (3) percentage of families with children that are female-headed; (4) percentage of households with public assistance income; (5) adult unemployment rate in the tract; (6) percentage of families with income above $30,000; (7) percentage of adults employed in executive or managerial jobs; (8) percentage of housing units that are rentals; and (9) percentage of persons above 5 years old who did not live at that address 5 years earlier. A factor analysis of these measures was used to derive two variables: neighborhood disadvantage and neighborhood residential mobility. Finally, neighborhoods were coded as either racially homogeneous (90% or greater of one race) or heterogeneous. In multivariate logistic regression models including both individual and neighborhood characteristics, neighborhood mobility was associated with greater likelihood of schizophrenia, adjusted odds ratio (95% confidence interval) = 1.27 (1.02, 1.59) and greater likelihood of major depression, adjusted odds ratio (95% confidence interval) = 1.16 (1.03, 1.29). The index of neighborhood disadvantage was also associated with greater likelihood of major depression, adjusted odds ratio (95% confidence interval) = 1.14 (1.01, 1.31). In these models, individual-level poverty (household income less than $10,000 per year) was associated with greater likelihood of schizophrenia, adjusted odds ratio (95% confidence interval) = 2.66 (1.30, 5.42), and major depression, adjusted odds ratio (95% confidence interval) = 1.69 (1.20, 2.36). In their analysis, they used the multilevel equation $y_{ti} = \alpha + \beta' \chi_{ti} + \nu_{ti}$, where t indexes census tracts and i indexes individuals within tracts, and $\nu_{ti} = \varepsilon_t + \upsilon_{ti}$. The authors report that when the analyses were reestimated using hierarchical linear regression, there was no significant census tract level variation in the distribution of mental disorders after the individual- and neighborhood-level variables were added to the model, and therefore argue that Level 2 variance (tract level) does not warrant use of multilevel levels with these data.

In a multilevel model of incident cases of schizophrenia in 35 neighborhoods in Maastricht, the Netherlands, Van Os, Driessen, Gunther, and Delespaul (2000) found that controlling for individual characteristics, deprived neighborhoods (characterized by relatively high unemployment and welfare dependence) were associated with greater incidence of schizophrenia, with relative risk and 95% confidence interval of 1.07 (1.01, 1.14) and 1.04 (1.00, 1.08), respectively. They also report that the Level 2 (neighborhood) variance is not statistically significant at 0.14 (95% confidence interval 0.00, 0.29; $p = 0.055$), yet still constitutes about 12% of the total observed variance. Unlike Silver et al. (2002), they argue that this level of random neighborhood variance is not a chance finding and is an argument in support of using multilevel methods.

A number of other researchers have used multilevel models to examine effects of neighborhood

deprivation or disorder on symptoms of depression or psychological distress and identified significant associations. There has been a lack of consistency in the application of theoretical frameworks and operationalization of measures, which may have detracted from the usefulness of multilevel measures in psychiatric epidemiology thus far.

Theoretical and Measurement Challenges

Multilevel modeling allows researchers to use powerful statistical techniques to incorporate multiple measurement levels of data. However, there is concern that the theoretical and conceptual developments are lagging behind computational abilities. To date, more multilevel empirical research has been conducted than theoretical models of contextual influence have been developed. As a result, epidemiologists have used theories from other fields such as sociology and community psychology for explanatory models. However, given the potential complexity of multilevel relationships as described above, more specific theories and hypotheses may be needed for epidemiologic models. For example, the appropriate size or scope of a group is debatable: Overly large group identities, such as cities or states, may be too large to detect between-group variation, but very small identities, such as census block or family may lack variation at the individual level. Choice of group size should be dictated by theory or hypothesis rather than empirically.

Group-level definition and measurement are other challenges to multilevel modeling. For example, neighborhood can be defined geographically or as an abstract concept based on dynamic interaction. These types of group characteristics have been called derived variables and integral variables. Derived variables, also known as analytical, aggregate, or compositional variables, summarize the characteristics of the individuals in the group (mean proportions, or measures of dispersions), such as median household income or proportion of household members with a high school education. A special type of derived variable is the average of the dependent variable within the group, for example, prevalence of infection or prevalence of a behavior. Integral variables, also known as primary or global variables, describe characteristics of the group that are not derived from characteristics of individuals, such as availability of services, certain regulations, or political systems. A special type of integral variable refers to patterns and networks of contacts or interactions among individuals within groups. Although distinct, derived and integral variables are closely related. For example, the composition of a group may influence the predominant types of interpersonal contacts, values, and norms or may shape organizations or regulations within the group that affect all members.

Traditional psychometric methods of evaluating scale reliability and validity are inadequate for the assessment of nonindividual neighborhood, or ecological, measures. Raudenbush and Sampson (1999) have proposed a methodology for understanding the psychometric properties of ecological measures, called ecometrics. For example, in developing a neighborhood-level measure from a survey of individual residents, an individual's item responses are aggregated to create a single scale, and then all individuals' scales are aggregated into a neighborhood measure. In this case, unlike in traditional psychometrics, scale reliability depends not only on the number of items in a scale and item consistency within a respondent but also on the number of respondents and the scale consistency with respondents, or intersubjective agreement. Furthermore, ecometric assessment of a neighborhood measure requires examination of potential sources of bias, such as nonrepresentative sampling, which should be adjusted for statistically. Similar techniques can be applied to assessing the reliability of observational data, where the unit of observation is aggregated, rather than individuals' survey responses, interrater agreement can be calculated, and biases arising from the sample (e.g., time of day of observations) adjusted. In addition to reliability analyses, ecometrics includes methodologies for other aspects of scale construction, including analysis of ecological scale dimensionality and validity, although use of these techniques is not often reported.

—Jane K. Burke-Miller

See also Geographical and Social Influences on Health; Geographical and Spatial Analysis; Social Capital and Health; Social Epidemiology; Social Hierarchy and Health

Further Readings

Bruce, M. L., Takeuchi, D. T., & Leaf, P. J. (1991). Poverty and psychiatric status. *Archives of General Psychiatry*, *48*, 470–474.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm.

Journal of the Royal Statistical Society, Series B, *39*, 1–38.
Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health*, *21*, 171–192.
Diez-Roux, A. V. (2002). A glossary for multilevel analysis. *Journal of Epidemiology & Community Health*, *56*, 588–594.
Diez-Roux, A. V. (2003). The examination of neighborhood effects on health: Conceptual and methodological issues related to the presences of multiple levels of organization. In I. Kawachi & L. F. Berkman (Eds.), *Neighborhoods and health* (pp. 45–64). New York: Oxford University Press.
Dohrenwend, B. P., & Dohrenwend, B. S. (1969). *Social status and psychological disorder: A causal inquiry.* New York: Wiley-Interscience.
Duncan, C., Jones, K., & Moon, G. (1998). Context, composition, and heterogeneity: Using multilevel models in health research. *Social Science & Medicine*, *46*, 97–117.
Goldsmith, H. F., Holzer, C. E., & Manderscheid, R. W. (1998). Neighborhood characteristics and mental illness. *Evaluation and Program Planning*, *21*, 211–225.
Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.
Lindley, D., & Smith, A. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society Series B (Methodological)*, *34*(1), 1–41.
O'Campo, P. (2003). Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health. *American Journal of Epidemiology*, *157*(1), 9–13.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology*, *29*, 1–41.
Silver, E., Mulvey, E. P., & Swanson, J. W. (2002). Neighborhood structural characteristics and mental disorder: Faris and Dunham revisited. *Social Science & Medicine, 55,* 1457–1470.
Subramanian, S. V., Jones, K., & Duncan, C. (2003). Multilevel methods for public health research. In I. Kawachi & L. F. Berkman (Eds.), *Neighborhoods and health* (pp. 65–111). New York: Oxford University Press.
Van Os, J., Driessen, G., Gunther, N., & Delespaul, P. (2000). Neighborhood variation in incidence of schizophrenia. *British Journal of Psychiatry*, *176*, 243–248.

Multiple Comparison Procedures

The issue of multiple comparisons has created considerable controversy within epidemiology. The fundamental questions are which procedure to use and whether probabilities associated with multiple tests should be adjusted to control Type I errors. The latter topic appears the most contentious.

It is helpful to make a distinction between *multiple comparisons*, which usually involve comparison of multiple groups or treatment arms on one dependent variable, and *multiple testing*, which usually involves the comparison of two (or more) groups on multiple dependent variables. Although both procedures raise many of the same questions, they differ in some important ways. Multiple comparison procedures are more formalized than those for multiple testing.

A Type I error, also referred to as alpha or by the Greek letter α, refers to the probability that a statistical test will incorrectly reject a true null hypothesis (H_0). In most cases, α is set at .05 or .01. When we test multiple differences (whether between groups or for different dependent variables), we can talk about Type I errors in two different ways. The *per comparison* error rate is the probability of an error on each of our comparisons, taken separately. Alternatively, the *experiment-wise* or *family-wise* error rate is the probability of making at least one Type I error in a whole set, or family, of comparisons. An important argument in epidemiology is which of these error rates is the appropriate one.

Multiple Comparisons

One approach to making multiple comparisons is to define a set of linear contrasts that focus specifically on important questions of interest. Normally, these questions are defined before the start of an experiment and relate directly to its purpose. For a clinical trial with two control groups and several treatment groups, we might, for example, create a contrast of the mean of the control groups versus the mean of the combined treatment groups. Or we might ask whether the mean of the most invasive medical procedure is significantly different from the mean of the least invasive procedure. We usually test only a few contrasts, both to control the family-wise error rate and because other potential contrasts are not central to our analysis. Generally, though not always, researchers will use some variant of a Bonferroni procedure (described below) to control the family-wise error rate over the several contrasts.

An alternative approach to multiple comparisons is to use a procedure such as the Tukey HSD ("honestly

significant difference") test. (There are many such tests, but the Tukey is a convenient stand-in for the others. These tests are discussed in any text on statistical methods and can be computed by most software programs.) The Tukey is a range test and is based on the Studentized range statistic. It modifies the critical value of the test statistic depending on the number of levels of the dependent variable. Other tests differ in how that critical value is determined, often in a sequential manner. The Tukey procedure performs all possible comparisons between pairs of groups and places the groups into homogeneous subsets based on some characteristic, in this example their mean. For instance, in an experiment with six groups there might be three homogeneous subsets, such that within each subset the mean values of each group do not differ significantly from each other. These homogeneous subsets may overlap; for instance, $\mu_1 = \mu_2 = \mu_3$; $\mu_3 = \mu_4 = \mu_5$; $\mu_5 = \mu_6$. The presence of overlapping sets is often confusing to users, but is inherent in the test. In addition, detection of homogeneous, but overlapping, subsets is seldom the goal of a statistical analysis, and it may be difficult to use this information. A third difficulty is posed by multiple outcomes; in fact, most researchers using multiple comparison procedures either do not measure more than one dependent variable, or they consider different dependent variables to be distinct and treat them separately. The final way of making comparisons among groups is to use some sort of Bonferroni correction. The Bonferroni inequality states that the probability of the occurrence of one or more events cannot exceed the sum of their individual probabilities. If the probability of a Type I error for one contrast is α, and we create k contrasts, the probability of *at least* one Type I error cannot exceed $k\alpha$. So if we run each contrast at $\alpha' = \alpha/k$, then the family-wise error rate will never exceed α. The Bonferroni procedure and the sequential tests based on it are widely applicable.

Multiple Testing

The more tests you do, the more likely you are, purely by chance, to find a significant difference when the null hypothesis is actually true. That was the rationale behind the development of multiple comparison procedures, and it is the rationale behind the current debate over how to treat the comparison of two groups on multiple dependent variables. This is a situation in which Bonferroni corrections are often advocated.

It should be apparent that the Bonferroni correction is applicable to any set of multiple tests, whether they be multiple comparisons on k means, multiple t tests on k different dependent variables, or tests on correlations in a matrix. We can hold the family-wise error rate at .05 for two contrasts by evaluating each test at $\alpha = .05/2 = .025$, and we can do the same thing when we have two t tests on two dependent variables. The big question is whether we should.

When Should We Adjust Probabilities?

Given that we have good methods for controlling the family-wise error rate, when should we employ them—and when should we not do so? When computing pairwise comparisons of each mean against each other mean, there is not much of an argument. The Tukey, or one of its variants, should be used, and the family-wise error rate should be set at α. However, when the researcher is running only a few very specific contrasts that derive directly from the nature of the study, and that were identified before the data were analyzed and were not chosen simply because they were the largest difference, then a good case can be made for running each one at α per comparison, although some would argue for a corrected α.

There is certainly room for debate here, but there is nothing sacred about $\alpha = .05$. If there are three contrasts and one person adjusts α and another does not, they are simply working at different significance levels. It is no different from the general case where one person may prefer $\alpha = .01$, while another chooses $\alpha = .05$ for a single t test. However, it is easier to find significant results at a higher α level, so sometimes the discussion becomes contentious because results that are significant without Bonferroni adjustment become nonsignificant when the adjustment is applied.

When it comes to multiple outcome measures, things are somewhat less clear. Consider a psychologist who compares two groups in differentially stressful environments using a symptom checklist having multiple subtest scores. That researcher could use Hotelling's T^2 test, treating all dependent variables simultaneously. However, this will not yield specific information about the variables that are affected. Alternatively, he or she could run an unadjusted t test on each variable, but that would strike most people as a fishing expedition, and the associated Type I error rate would be high. Finally, he or she could run those

t tests but adjust the significance levels with a Bonferroni adjustment. In this example, the last option would seem to make the most sense.

Next, consider a study of cancer treatments with two dependent variables—tumor reduction and survival. These are two quite different outcome measures, and there is no obvious reason why they should be treated together, so an adjustment does not seem necessary. As others have pointed out, a treatment that reduces tumors but has no effect on survival is quite different from a treatment that has no effect on tumors but increases survival. As a general rule, when a clinical trial is designed to look at two or three different questions, especially if those questions lead to different outcome behaviors on the part of a researcher or physician, then it makes sense to treat those separately. If there is no clear hierarchy of questions, or no obvious measure of outcome (as in the case of a symptom checklist with multiple subscales), then the prudent thing to do would be to control family-wise error.

It is extremely rare for a study to exist on its own without a context. Choices about corrections for multiple comparisons should take into account the following facts: (1) Studies are designed for specific reasons against a background literature that is relevant to subsequent actions. (2) The fact that a particular treatment is effective may be trumped by the fact that it is outrageously expensive. (3) Not every statistically significant difference will lead to future implementation of the procedure. (4) A single study is not likely to change medical practice. In that context, it is often reasonable to take a more liberal approach and not restrict the study to a family-wise error rate.

—David C. Howell

See also Analysis of Variance; Clinical Trials; Hypothesis Testing; *p* Value; Significance Testing

Further Readings

Cook, R. J., & Farewell, V. T. (1996). Multiplicity considerations in the design and analysis of clinical trials. *Journal of the Royal Statistical Society, Series A*, *159*(Pt. 1), 93–110.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Savitz, D. A., & Olshan, A. F. (1995). Multiple comparisons and related issues in the interpretation of epidemiologic data. *American Journal of Epidemiology*, *142*, 904–908.

Thompson, J. R. (1998). Invited commentary: Re: "Multiple comparisons and related issues in the interpretation of epidemiologic data." *American Journal of Epidemiology*, *147*, 801–806.

Multivariate Analysis of Variance

Multivariate analysis of variance (MANOVA) is a statistical technique used extensively in all types of research. It is the same thing as an analysis of variance (ANOVA), except that there is more than one dependent or response variable. The mathematical methods and assumptions of MANOVA are simply expansions of ANOVA from the univariate case to the multivariate case.

A situation in which MANOVA is appropriate is when a researcher conducts an experiment where several responses are measured on each experimental unit (subject) and experimental units have been randomly assigned to experimental conditions (treatments). For example, in a double-blind study systolic and diastolic blood pressures are measured on subjects who have been randomly assigned to one of two treatment groups. One group receives a new medication to treat high blood pressure, and the other group receives a placebo. The group that receives the placebo is considered a control group. The researcher wants to know whether the new medication is effective at lowering blood pressure.

There are several compelling reasons for conducting a MANOVA instead of an ANOVA. First, it is more efficient and economical in the long run to measure more than one response variable during the course of an experiment. If only one response is measured, there is the risk that another important response has been ignored. The measurement of several response variables provides a more thorough understanding of the nature of group differences given the response variables.

Another good reason to use MANOVA is that analyzing multiple responses simultaneously with a multivariate test is more powerful than analyzing the individual responses separately with multiple univariate tests. The chances of incorrectly rejecting the null hypothesis are inflated with multiple univariate tests, because the Type I error rate (α) increases with each additional test. For instance, the overall Type I error

rate for two univariate tests each with α set at .05 is .10 $(1 - (.95)^2)$ rather than .05.

Finally, the correlation between the response variables is taken into account in a multivariate test. The result is that differences between groups that are not detected by multiple univariate analyses may become obvious. Figure 1 illustrates the hypothetical univariate distributions of two response variables, X_1 and X_2, for two study groups. The distributions appear to overlap such that no significant difference in means between groups is expected. In Figure 2, the multivariate distributions for the same response variables are illustrated with 95% confidence ellipses drawn about the group means. The figure shows that the two groups do not overlap as much as might be expected given the univariate distributions. Figure 3 illustrates the same multivariate distributions, except that the response variables are negatively correlated rather than positively correlated as in Figure 2. Given the degree to which the ellipses overlap in Figure 3, the null hypotheses may not be rejected.

When conducting a MANOVA, several assumptions are made about the data. When these assumptions do not hold, conclusions based on the analysis may be erroneous. The assumptions are as follows:

- The experimental units are random samples of target populations.
- The observations are independent of each other.
- The response variables are distributed multivariate normal. There is no test for multivariate normality commonly available. Generally, multivariate normality can be assumed when the individual variables are normally distributed; however, it is not guaranteed. Additionally, MANOVA is particularly sensitive to outliers so it is important to check for them prior to analysis. Outliers should be transformed or omitted from the analysis. Deviation from multivariate normality has less impact in larger samples.
- All populations have the same covariance matrix (homogeneity of covariance). This assumption is made because the error sums of squares are computed by adding the treatment sums of squares weighted by $(n_i - 1)$, where n_i is the number of experimental units in each treatment. Otherwise, adding the treatment sums of squares would be inappropriate.

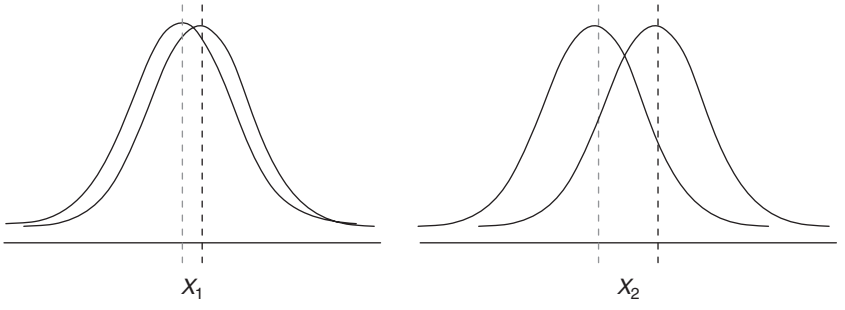


Figure 1 Univariate Distributions for Two Populations, Two Responses

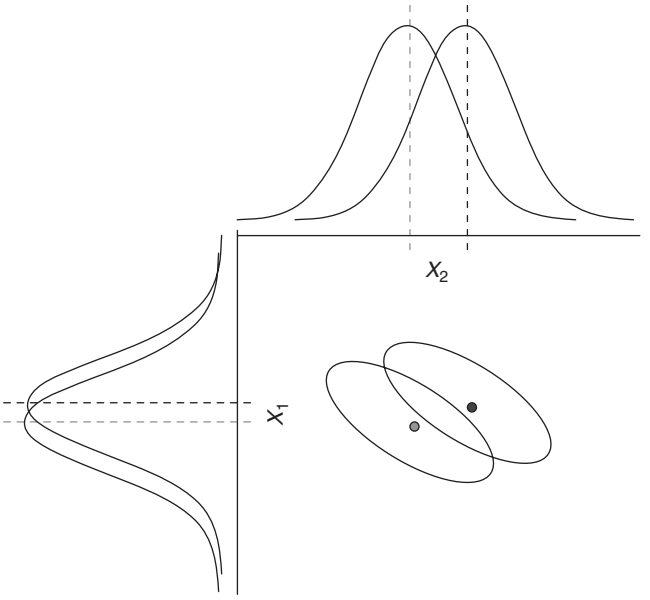


Figure 2 Multivariate Distribution for Two Populations, Two Positively Correlated Responses

In the blood pressure experiment, there is one factor; medication type. The number of factor levels or treatments (k) for medication type is two, medication and placebo. The number of responses (p) is two, systolic blood pressure and diastolic blood pressure. The appropriate model for this experiment is a one-way MANOVA. The model is constructed as follows:

$$\boldsymbol{x}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{e}_{ij}$$

where

$\boldsymbol{x}_{ij}$ are observations randomly sampled from p-variate populations representing the ith treatment on the jth unit, $j = 1, 2, \ldots, n_i$ and $i = 1, 2, \ldots, k$.

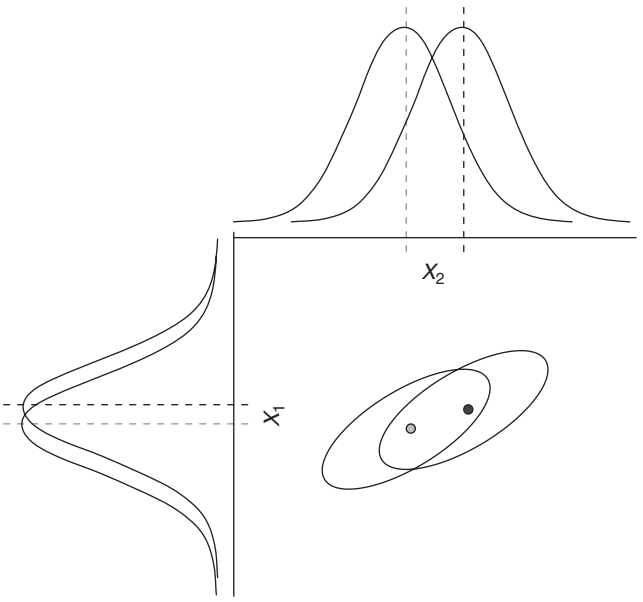


Figure 3 Multivariate Distribution for Two Populations, Two Positively Correlated Responses

$\boldsymbol{e}_{ij}$ are independent random variables with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}$. $\boldsymbol{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$.

$\boldsymbol{\mu}$ is a vector of overall means.

$\boldsymbol{\tau}_i$ represents the main effect of the ith treatment subject to the following restriction: $\sum_{i=1}^{k} n_i \tau_i = 0$.

All the vectors are $p \times 1$.

The null hypothesis is that there are no differences in treatment means for either response due to the main effect of medication type.

Suppose there was another factor in the blood pressure example, say gender. In addition to the parameters for $\boldsymbol{\mu}$ and the main effect of medication type, $\boldsymbol{\tau}_i$, there would be a parameter for the main effect of gender, $\boldsymbol{\beta}_l$, and a parameter for the interaction effect of the two factors, $\boldsymbol{\gamma}_{il}$, where $i = 1, \ldots, k, l = 1, \ldots, b$. The error parameter becomes $\boldsymbol{e}_{ilj}$, where $j = 1, \ldots, n$ and n is the number of experimental units in each treatment. The model is subject to the following restriction:

$$\sum_{i=1}^{k} \tau_i = \sum_{l=1}^{b} \beta_l = \sum_{i=1}^{k} \gamma_{il} = \sum_{l=1}^{b} \gamma_{il} = 0.$$

Now, the three null hypotheses are that there are no differences in treatment means for either factor due to (1) the main effect of treatment type, (2) the main effect of gender, or (3) the interaction between treatment type and gender.

The goal of MANOVA, as in ANOVA, is to test for differences between treatment means. However, instead of testing the equality of means for one response variable as in ANOVA, the means for all response variables are tested for equality simultaneously. Suppose there are k treatments and p response variables. The null hypothesis is

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k,$$

where $\boldsymbol{\mu}_i$ is a $p \times 1$ vector, $i = 1, \ldots, k$.

In ANOVA, the overall test of significance is based on the ratio of the between-subject sums of squares to the within-subject sums of squares (SS_B/SS_W) adjusted for degrees of freedom. The multivariate case is more complex, because there is information for p variables that must be considered jointly to obtain an overall test of significance. There are four overall multivariate tests of significance designed specifically for this purpose: Wilks's lambda, Pillai's trace, Roy's greatest characteristic root, and Hotelling's T.

Overall MANOVA tests are conducted for the main effect of each factor on the response variables. When there is more than one independent variable, overall MANOVA tests are conducted for the effect of the interaction of the factors on the response variables. If the mean vector for at least one treatment is not equal to all the others, then the null hypothesis that there is no difference between mean vectors is rejected.

To decide whether to reject the null hypothesis, the overall MANOVA test statistics are transformed into an approximate F distribution. Most statistical analysis computer programs execute the transformations specific to each test and output the results. The null hypothesis can be rejected if the p value of the observed F statistic is less than the set α level. The results of these four tests will disagree as to whether to reject the null hypothesis at the set α level, except in the case where $k = 2$ or $p = 1$. It is not always easy to know which test to use. While Wilks's lambda is used the most often, researchers have studied each test under conditions violating one or more assumptions. They have found that the tests are generally robust to departures from multivariate normality. They have also found that the tests are not robust to departures

from homogeneity of variance/covariance when there are unequal sample sizes. When there are equal sample sizes, Pillai's trace was found to be the most robust against violation of assumptions than any of the other tests.

When a MANOVA has produced a significant main effect, there are two approaches to exploring differences between groups to determine the source. One approach is to interpret the contributions of the response variables. There are several ways of doing this, but the two most common are interpreting univariate *F* tests on each of the response variables and interpreting the coefficients of the discriminant function. Univariate *F* tests may indicate which response is causing the significance. Discriminant functions are linear combinations of the *p* original variables designed to maximize the differences between groups. The magnitude of the coefficients indicates the importance of the variable to the separation of groups relative to the other variables. In a multivariate environment, the discriminant function approach is preferable; however, caution must be exercised when using these methods of interpretation. The univariate *F* tests do not take the correlation between the response variables into account. The magnitude of the discriminant function coefficients may be misleading if the original variables are highly correlated.

The other approach to exploring group differences is to perform contrasts on the response variables. Multivariate contrasts are performed on the vector of responses. There are several multivariate statistics available to test for mean differences between the two groups. Univariate contrasts are performed on each variable individually as in ANOVA. The same contrasts used in ANOVA are appropriate in this application. Univariate contrasts can be used to further explore significant multivariate contrasts.

—Mary Earick Godby

See also Analysis of Variance; Discriminant Analysis

Further Readings

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance.* Beverley Hills, CA: Sage.

Hand, D. J., & Taylor, C. C. (1987). *Multivariate analysis of variance and repeated measures: A practical approach for behavioral scientists*. New York: Chapman & Hall.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis.* Englewood Cliffs, NJ: Prentice Hall.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis.* San Diego, CA: Academic Press.

Wickens, T. D. (1995). *The geometry of multivariate statistics.* Hillsdale, NJ: Lawrence Erlbaum.

Mutation

A mutation is a transmissible or heritable change in the nucleotide sequence of the genetic material of a cell or virus. Mutations are either spontaneous, occurring naturally due to errors in DNA or RNA replication, or induced by external agents. When identifying the etiology of a disease and the factors that alter a person's risk for disease, epidemiologists must often determine the unique contributions of environmental and genetic factors. Increasingly, we are made aware of the importance of mutations in the development of disease and the evolution of pathogens.

Mutations are often involved in the etiology of diseases attributed to host genetic factors. Mutations of the breast cancer susceptibility genes *BRCA1* or *BRCA2* result in an increased risk of developing breast cancer or ovarian cancer and account for up to half of hereditary breast cancers. It is believed that these genes normally play a role in repairing breaks in double-stranded DNA induced by radiation and that mutations in these genes hinder this mechanism, resulting in DNA replication errors and cancer. The effect of host mutations need not be deleterious as certain mutations confer protection from disease. A mutant variant of the chemokine receptor 5 resulting from a 32 base pair deletion (CCR5Δ32) is associated with nearly complete resistance to HIV-1 infection in homozygous individuals and partial resistance with delayed disease progression in heterozygous individuals. CCR5 is a necessary coreceptor for HIV-1 infection; individuals with at least one mutant allele do not express the receptor on their cell surfaces and are thereby protected.

Mutations acquired by pathogens may alter infectivity and virulence, and therefore, affect disease in the host. Influenza virus lacks a proofreading mechanism and thus allows errors during replication to remain undetected and uncorrected, resulting in an accumulation of point mutations and the ultimate emergence of a new antigenic variant. This process is

referred to as *antigenic drift* and is the reason why the human influenza vaccine must be updated on an annual basis. Avian influenza viruses undergo limited antigenic drift in their aquatic bird reservoirs; however, the accumulation of mutations becomes more pronounced when the virus spreads through domestic poultry, and continued accumulation could support human-to-human transmission as witnessed with the 1918 Spanish influenza pandemic.

Epidemiologists are recognizing with increasing frequency the contributions of genetic factors such as mutations in the development of both chronic and communicable diseases.

—Margaret Chorazy

See also Association, Genetic; Gene; Gene-Environment Interaction; Genetic Epidemiology; Genetic Markers

Further Readings

Andersen, W. R., & Fairbanks, D. J. (1999). Mutation. In *Genetics: The continuity of life* (pp. 124–155). Pacific Grove, CA: Brooks/Cole.

Blanpain, C., Libert, F., Vassart, G., & Parmentier, M. (2002). CCR5 and HIV infection. *Receptors and Channels*, *8*, 19–31.

National Library of Medicine, National Center for Biotechnology Information. (n.d.). Cancers. In *Genes and disease*. Retrieved January 31, 2007, from http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=gnd.chapter.10.

Webster, R. G., & Hulse, D. J. (2004). Microbial adaptation and change: Avian influenza. *Revue Scientifique et Technique (International Office of Epizootics)*, *23*(2), 453–465.

N

National Ambulatory Medical Care Survey

The National Ambulatory Medical Care Survey (NAMCS) is a national survey that provides information about the delivery and use of ambulatory medical care services provided by nonfederal, office-based physicians in the United States. The NAMCS captures information about patient demographics, sources of payment, the reason for visit, and diagnosis and drug information. Epidemiologists, physician associations, and policymakers frequently use the gathered data to identify patterns of care in the United States, discover emerging trends and issues, or support and inform decision making.

Since 1989, the NAMCS has been conducted annually; prior to that, it was conducted annually from 1973 to 1981 and also in 1985. Each year approximately 3,000 physicians are randomly selected to provide information on approximately 30 patient visits seen at their practice over a 1-week period.

Included within the scope of the NAMCS are physician services provided in locations such as free-standing clinics or urgent care centers, neighborhood mental health centers, or a health maintenance organization. Services provided in locations such as hospital emergency departments, institutional settings such as schools or prisons, and locations that specialize in laser vision surgery are not included in the survey. Physicians specializing in anesthesiology, pathology, or radiology services are also excluded. The NAMCS data are collected by the National Center for Health Statistics of the Centers for Disease Control and Prevention.

Physicians complete the questionnaires that are used to compile the NAMCS data, and their participation is entirely voluntary. The NAMCS data are *de-identified*, meaning all personal identifying information is removed to protect the confidentiality of the respondents and their patients. During the past 10 years, participation rates have varied between 65% and 75%. Various strategies have been attempted to improve response rates such as publicity campaigns and eliminating questions with high item nonresponse. Physicians are not remunerated for their participation in the survey, but are motivated to take part so that their practice and others that are similar will be represented in the data.

When conducting analyses on the NAMCS, it should be noted that it is a record-based survey that captures information about episodes of care. The unit of analysis is therefore episodes of care, that is, patient visits, not number of patients. Hence, population incidence and prevalence rates cannot be calculated from the NAMCS data. However, it is possible to compute estimates for things such as the most common reasons patients visit their doctors or the percentage of office visits that include mention of particular pharmaceuticals. Estimates that are based on at least 30 individual records and also have a relative standard error less than 30% are considered reliable. To improve reliability, data from the NAMCS can be combined with data from its sister survey the National Hospital Ambulatory Medical Care Survey. Because

of the sampling design, the survey should be used primarily to determine national estimates: Meaningful assessments cannot be made at lower levels of geography such as states or counties. The sample data must be weighted to produce national estimates.

—Alexia Campbell

See also Centers for Disease Control and Prevention; Health Care Services Utilization; National Center for Health Statistics

Web Sites

Centers for Disease Control and Prevention, Ambulatory Health Care Data: http://www.cdc.gov/nchs/about/major/ahcd/namcsdes.htm.

National Center for Health Statistics

The National Center for Health Statistics (NCHS) is the principal U.S. Federal agency responsible for collecting, analyzing, and disseminating statistical information relevant to the health and well-being of the American public. The purpose of the NCHS is to provide information that may be used to guide actions and policies to improve the health of Americans. It is a center within the Centers for Disease Control and Prevention, which is located within the Department of Health and Human Services. The NCHS, along with nine other federal statistical agencies, is represented on the Interagency Council on Statistical Policy (ICSP), part of the Office of Management and Budget (OMB). The purpose of the ICSP is to coordinate statistical work and provide advice and counsel to the OMB.

The NCHS conducts a wide range of annual, periodic, and longitudinal sample surveys and administers the National Vital Statistics Systems. Among the best known of the NCHS surveys are the National Health and Nutritional Examination Survey, the National Health Interview Survey, the National Immunization Survey, the Longitudinal Studies on Aging, and the National Survey of Family Growth. Through the National Vital Statistics System, the NCHS collects information on vital events from the vital registration systems of the jurisdictions responsible for collecting them, that is, the systems within each of the 50 states, Washington, D.C., New York City, and the territories of Puerto Rico, the U.S. Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands. The NCHS then compiles and analyzes this information and in some cases links it to other NCHS data sets. As a rule, the NCHS does not release information about individuals and does not provide copies of birth and death certificates, a responsibility that remains with the local jurisdiction.

The NCHS produces a number of publications and reports, ranging from simple one-page fact sheets about particular health topics to detailed descriptions of the NCHS surveys and analyses of data from those surveys. The *Advance Data* report series provides timely analyses of data from current survey, usually focused on a narrow topic such as HIV risk in young adults and are relatively brief (20 to 40 pages). The *National Vital Statistics Reports* series, which replaces the *Monthly Vital Statistics Reports*, provides timely reports of birth, death, marriage, and divorce statistics, and includes four to six special issues per year that focus on specialized topics such as trends in Cesarean birth rates or births to women aged between 10 and 14 years. The *Vital Health and Statistics Series* produces reports in 24 different series, which cover a range of topics from technical descriptions of the survey design and data collection methods of the NCHS surveys to analysis of data from those surveys and vital statistics data. Often, *Advance Data* reports are followed up by a more detailed report in the *Vital Health and Statistics Series.* Most of these publications are available for free download from the NCHS Web site.

Much of the data collected by the NCHS are available for download through the NCHS Web page or may be requested on CD or tape. Some of the NCHS data that are not publicly accessible due to confidentiality restrictions are available through the NCHS Research Data Center (RDC). To gain access to restricted data, which includes NCHS survey data at lower levels of geographic aggregation or which include more contextual information than can be included in publicly released files, researchers must submit a proposal to the RDC and agree to follow specified procedures to protect the confidentiality of survey respondents.

—Sarah Boslaugh

See also Centers for Disease Control and Prevention; Governmental Role in Public Health; National Health and Nutrition Examination Survey; National Health Interview Survey; National Survey of Family Growth; Public Health Surveillance

Further Readings

Hetzel, A. M. (1997). *U.S. Vital statistics system: Major activities and developments, 1950–95.* Hyattsville, MD: National Center for Health Statistics. Retrieved November 24, 2006, from http://www.cdc.gov/nchs/products/pubs/pubd/other/miscpub/vsushist.htm.

Web Sites

National Center for Health Statistics: www.cdc.gov/nchs

National Death Index

The National Death Index (NDI) is a computerized index of information compiled by the National Center for Health Statistics (NCHS) from death record information submitted to the NCHS by the state offices of vital statistics. This national file of death records contains a standard set of identifying information for each decedent, which includes first and last names, middle initial, father's surname, social security number, sex, race, date of birth, state of birth, state of residence, marital status, and age at death.

The NDI enables investigators to ascertain if the participants in their studies have died by matching the identifying information for an individual with the NDI database. The NDI Retrieval Program searches the NDI file to identify possible matches between a particular NDI death record and a particular user record. An NDI match can identify the state in which death occurred, the date of death, and the death certificate number. To qualify as a possible match, both records must satisfy at least one of seven criteria. The complete social security number alone would provide a match; the other six matching criteria consist of various combinations of birth date, name, and father's surname. When a user record matches one or more NDI records, an NDI Retrieval Report is generated, listing all the identifying information for each of the possible matches, and indicating items that match exactly, items that do not match, and items that possibly match. However, it is up to the users to determine whether NDI records match the individuals in their studies.

In designing a study in which ascertaining death might be important, investigators should collect as many of the NDI data items as possible to optimize the assistance available through the NDI. Using NDI-Plus, investigators can obtain the ICD-9 codes for the cause of death (underlying cause and multiple causes). The NDI does not provide copies of death certificates.

The NDI database contains death record information (beginning with 1979 deaths) for all 50 states, the District of Columbia, Puerto Rico, and the Virgin Islands. Death records are added to the NDI file annually, about 12 months after the end of each calendar year. Approximately 2 million death records are added each year. Deaths during the 2005 calendar year will be available for the NDI search in April 2007.

Use of the NDI is restricted to statistical purposes involving medical and health research. Investigators planning to use the NDI must complete an application and review process, which usually takes 2 to 3 months. Users must meet confidentiality requirements and must submit data on study subjects in a manner that meets the NCHS technical specifications. The fees for routine NDI searches, as of June 2007, consist of a $350 service charge, plus $0.21 per study subject for each year of death searched if vital status of subject is unknown, and $5.00 per decedent if subjects are known to be deceased. Fees for the optional NDI-Plus are slightly higher. A free user's manual can be requested from the NDI; it includes a sample application form.

—Judith Marie Bezy

See also Death Certificate; International Classification of Diseases; National Center for Health Statistics

Further Readings

National Center for Health Statistics. (1997). *National Death Index user's manual* (Publication No. 7–0810). Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

National Center for Health Statistics. (1999 revised). *National Death Index Plus: Coded causes of death, supplement to the NDI user's manual* (Publication No. 0–0190). Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Web Sites

National Death Index, National Center for Health Statistics: www.cdc.gov/nchs/ndi.htm.

National Health and Nutrition Examination Survey

The National Health and Nutrition Examination Survey (NHANES) is a series of cross-sectional, nationally representative surveys conducted by the National Center for Health Statistics. The NHANES uniquely combines in-person interviews with physical examinations and is the authoritative source of objective information on the health and nutritional status of the U.S. population. In 2002, the Continuing Survey of Food Intakes by Individuals (CSFII) of the Department of Agriculture was incorporated into the NHANES. The integrated dietary component is now called *What We Eat in America*.

Findings from the survey provide estimates of important health conditions and risk factors in the U.S. children and adults, such as blood-lead levels, obesity, oral health, sexually transmitted diseases, and smoking. Survey data are also used to assess rates of previously undiagnosed conditions, monitor trends in risk behaviors and environmental exposures in the overall population and subgroups, analyze risk factors for selected disease, explore emerging public health issues, and develop appropriate public health policies and interventions related to health and nutrition.

Survey History

The NHANES resulted in the late 1960s from the addition of a nutritional component to the Health Examination Survey, established by the National Health Survey Act of 1956. While the NHANES was conducted on a periodic basis from 1971 to 1994, the current survey has been carried out continuously since 1999. The data are now released for public use in 2-year increments.

In the current survey, approximately 7,000 individuals of all ages are interviewed in their homes each year. Of these, approximately 5,000 also complete a health examination component. Fifteen primary sampling units (PSUs), which are counties or small groups of counties, are visited annually. To ensure reliable estimates for important population subgroups, the current survey includes oversamples of low-income persons, adolescents between 12 and 19 years of age, persons 60+ years of age, African Americans, and Mexican Americans. The current continuous design of the NHANES allows annual estimates to be calculated for some health indicators. However, data combined over several years are often necessary to produce reliable statistics in the overall sample or in subgroups.

Previous surveys in the NHANES series included the NHANES I (1971–1975), the NHANES II (1976–1980), the NHANES III (1988–1994), the Hispanic HANES (1982–1984), and the NHANES I Epidemiologic Follow-Up Study (NHEFS) (1982–1984, 1986, 1987, and 1992). The NHANES I interviewed a sample of 31,973 persons aged between 1 and 74 years, 23,808 of whom were also given a health examination. The sample was selected to include oversampling of population subgroups thought to be at high risk of malnutrition, including low-income persons, preschool children, women of childbearing age, and the elderly. The NHANES II included a sample of 25,286 persons aged between 6 months and 74 years who were interviewed and 20,322 who were also examined. Children and persons living at or below the poverty level were oversampled. The NHANES III further expanded the age range to include infants as young as 2 months of age, with no upper age limit on adults. Information was collected on 33,994 persons, of whom 31,311 were examined. Younger and older age groups, as well as blacks and Hispanics, were oversampled during the NHANES III. The Hispanic HANES (HHANES) was conducted to obtain reliable estimates of the health and nutritional status of Puerto Ricans, Mexican Americans, and Cuban Americans residing in the United States. The survey included 7,462 Mexican Americans from Texas, Colorado, New Mexico, Arizona, and California; 1,357 Cuban Americans from Dade County, Florida; and 2,834 Cuban Americans from the New York area, including parts of New Jersey and Connecticut.

The NHEFS is a longitudinal study designed to investigate the relationships between clinical, nutritional, and behavioral factors assessed in the NHANES I and subsequent morbidity, mortality, and hospital utilization, as well as changes in risk factors, functional limitations, and institutionalization. The NHEFS cohort includes all 14,407 persons aged between 25 and 74 years who completed a medical examination for the NHANES I. Four follow-up

studies have been conducted, and mortality data collection is scheduled to continue indefinitely.

Data Collection

The NHANES employs a stratified, multistage probability sample of the civilian noninstitutionalized U.S. population. Four sampling stages are used to produce a nationally representative sample: PSUs, which are counties or small groups of counties; area segments of PSUs such as a block or group of blocks containing a cluster of households; households within segments; and finally, one or more persons within households.

The data are collected from two primary sources, an initial household interview and a standardized health examination conducted by trained medical personnel. The household interview includes demographic, socioeconomic, dietary, and health-related questions. Health examinations for eligible participants usually occur in mobile examination centers, which are tractor trailers specially outfitted for physical and dental examinations and laboratory tests.

Data Uses

The data from the NHANES have been used in a large number of epidemiological studies and health-related research. An electronic search conducted using the National Library of Medicine Pub Med on July 17, 2006, produced 10,499 scientific journal articles identified with the search term *NHANES*. Findings from the NHANES have contributed directly to the U.S. public health policies and health services. The data from NHANES documented the considerable increase in obesity in the United States since 1980 and substantiate the current national effort aimed at addressing this epidemic in adults and children. When nutrition data from the NHANES I and II indicated that many Americans were consuming inadequate amounts of iron, the government responded by fortifying grain and cereal with iron. The NHANES has also contributed to other health-related guidelines and reforms such as folate fortification of grain, reduction of lead in gasoline, and development of growth charts for children.

—Helen L. Kwon and Luisa N. Borrell

See also National Center for Health Statistics; Nutritional Epidemiology; Prevalence; Probability Sample

Further Readings

National Center for Health Statistics. (n.d.). *National health and nutrition examination survey*. Retrieved September 7, 2006, from www.cdc.gov/nchs/nhanes.htm.

National Center for Health Statistics. (1963). Origin, program, and operation of the U.S. National Health Survey. *Vital Health Statistics*, *1*, 1–41.

National Center for Health Statistics. (1994). Plan and operation of the Third National Health and Nutrition Examination Survey, 1988–94. *Vital Health Statistics*, *1*, 1–39.

National Research Council. (2005). Federal data sets on food and nutrition. In *Improving data to analyze food and nutrition policies* (pp. 26–52). Washington, DC: National Academies Press. Retrieved September 7, 2006, from http://www.nap.edu/catalog/11428.html.

National Health Care Survey

The National Health Care Survey (NHCS) is a group of eight related surveys conducted under the auspices of the National Center for Health Statistics (NCHS). Each collects data from health care providers or establishments (such as hospitals) or their records, rather than from patients, and each is based on a multistage sampling plan that allows the computation of valid national estimates of different aspects of health care utilization. These surveys gather information about patients, caregivers, and institutions, as well as on health care events such as physician office visits, hospitalizations, and surgeries. The data for many of the surveys included in the NHCS are available for public use and may be downloaded from the NHCS Web site.

The four original parts of NHCS were the National Hospital Discharge Survey (NHDS), the National Ambulatory Medical Care Survey (NAMCS), the National Nursing Home Survey (NNHS), and the National Health Provider Inventory (NHPI). Later surveys added to the NHCS are the National Survey of Ambulatory Surgery (NSAS), the National Hospital Ambulatory Medical Care Survey (NHAMCS), the National Home and Hospice Care Survey (NHHCS), and the National Employer Health Insurance Survey (NEHIS).

The NAMCS and NHAMCS gather information about ambulatory care: visits to physician's offices in the case of the NAMCS and to hospital outpatient departments (OPDs) and emergency departments (EDs) in the case of the NHAMCS. The NAMCS

began collecting data in 1973 and has been conducted annually since, except for the years 1982 to 1984 and 1986 to 1988. The information collected by the NAMCS, which is reported by physicians, includes patient demographics, expected payment source, patient's complaint, procedures and services performed, and medications. The NHAMCS has been conducted annually since 1992; it was begun in recognition of the fact that an increasing number of physician visits took place in OPDs and EDs rather than physician offices. It collects similar information to the NAMCS plus information about characteristics of the hospital, such as type of ownership. The NHAMCS information is reported by hospital staff, and slightly different forms are used for ED and OPD visits.

The NHDS, which has been conducted annually since 1965, was one of the first facility-based surveys conducted by the NCHS. It collects data from a sample of inpatient records from a national sample of about 500 hospitals: Federal, military, institutional, and Veteran's Affairs hospitals are excluded from the sample, as are hospitals with fewer than six beds or with an average length of stay longer than 30 days. The data collected include patient demographics, diagnoses, procedures and discharge status, and admission and discharge dates.

The NNHS was first conducted in 1973 to 1974 and has since been conducted for the years 1977, 1985, 1995, 1997 and 1999, and 2003. It collects data on both facilities, such as size, ownership, and occupancy rate, and on the residents (current and discharged), such as demographics, health status, and services received.

The NHPI, as the name suggests, is an inventory rather than a survey. It was conducted once, in 1991, and provides a comprehensive national listing of health care providers as of that year. The data were collected via mail questionnaires, and two different forms were used, for the two types of providers included. Nursing homes and board and care homes were sent Facility questionnaires, while home health agencies and hospices were sent Agency questionnaires. The information collected includes location, staff, total number of clients served in 1990, age and sex of residents (Facility questionnaire only), and number of current and discharged clients. The NHPI has served as a sampling frame for other health care provider inventories as well as a source of data.

The NSAS was conducted in 1994, 1995, and 1996, to supplement information about inpatient medical and surgical care collected through the NHDS, largely in response to the dramatic growth of ambulatory surgery facilities in the 1980s. The data were abstracted from medical records and included patient demographic information, expected source of payment, time spent in different phases of care, type of anesthesia used, final diagnoses, and surgical and diagnostic procedures performed.

The NHHCS was conducted in 1992, 1993, 1994, 1996, 1998, and 2000 through interviews with administrators and staff at home health agencies (who provide care in an individual's place of residence, for the purpose of restoring or maintaining health or minimizing the effects of illness or disability) and hospices (who provide palliative and supportive care services, for a dying person and their families, in either the person's home or in a specialized facility); the staff member was directed to refer to the patient's medical record to answer questions about current and discharged patients. The data collected about agencies included ownership, certification, number of patients seen, and number and types of staff members. The data collected on current and discharged patients included demographic characteristics, functional status, health status, payment information, and services used.

The NEHIS was the first federal survey to collect data about employer-sponsored health insurance that represented all employers in the United States and all health plans offered by those employers. The NEHIS was conducted only once, in 1994: Due to confidentiality concerns, the data have not been released to the general public but may be accessed through application to the Research Data Center of the NCHS. The NEHIS conducted a probability sample of business establishments (e.g., a single General Motors plant in a specific geographic location), governments, and self-employed individuals; for employers who offered a large number of insurance plans to their employees, a random sample of plans offered was also taken. This methodology allows valid estimates to be computed at the level of the individual state as well as at the national level. Information collected by the NEHIS includes availability of employer-sponsored health insurance, characteristics of plans offered, benefits, and costs.

—Sarah Boslaugh

See also Health Care Delivery; Health Care Services Utilization; Health Economics; National Ambulatory Medical Care Survey; National Center for Health Statistics

Further Readings

Dennison, C., & Pokras, R. (2000). Design and operation of the National Hospital Discharge Survey: 1988 Redesign. National Center for Health Statistics. *Vital Health Statistics*, *1*(39), 1–42.

Gottfried, I. B., Bush, M. A., & Madans, J. H. (1993). Plan and operation: National Nursing Home Survey Followup, 1987, 1988, 1990. National Center for Health Statistics. *Vital Health Statistics*, *1*(30), 1–65.

Haupt, B. J. (1994). Development of the National Home and Hospice Care Survey. National Center for Health Statistics. *Vital Health Statistics*, *1*(33).

Kovar, M. G. (1989). Data systems of the National Center for Health Statistics. National Center for Health Statistics. *Vital Health Statistics*, *1*(23).

McCaig, L. F., & McLemore, T. (1994). Plan and operation of the National Hospital Ambulatory Medical Care Survey. National Center for Health Statistics. *Vital Health Statistics*, *1*(34).

McLemore, T., & Lawrence, L. (1997). Plan and operation of the National Survey of ambulatory surgery. National Center for Health Statistics. *Vital Health Statistics*, *1*(37).

Web Sites

National Health Care Survey: http://www.cdc.gov/nchs/nhcs.htm.

National Health Interview Survey

The National Health Interview Survey (NHIS) is the principal source of national information about the health of the civilian noninstitutionalized U.S. population. The NHIS began collecting data in 1957 and has been conducted by the National Center for Health Statistics (NCHS) since 1960, when the NCHS was formed by combining the National Vital Statistics Division and the National Health Survey. Content covered by the NHIS has been updated every 10 to 15 years since its inception and was substantially revised for the surveys beginning in 1997.

The NHIS was authorized by the National Health Survey Act of 1956, which provided for a continuing survey to collect accurate and up-to-date information on the health of the U.S. population. The topics covered by the 2005 NHIS include demographics, specific medical conditions, activity limitations, general mental and physical health, physical activity, alcohol use, access to health care, health services utilization, health-related knowledge, blood donation, and HIV testing. The data drawn from the NHIS are used to monitor trends in illness and disability, to track progress toward achieving national health objectives, for epidemiologic and policy analysis, and for evaluating the effectiveness of Federal health programs.

The NHIS gathers data through personal interviews conducted in the respondent's household by employees of the U.S. Census Bureau following procedures specified by the NCHS. Within a household (defined as the people living within an occupied living unit and often synonymous with "family"), one adult is selected to provide data on all household members. In addition, one randomly selected adult and one randomly selected child (if available) are selected for further data collection; the adult provides information for both himself or herself and the child. The NHIS data are collected using a sampling plan that allows for the creation of national but not state-level estimates of the topics covered: Detailed information about the sampling plans used in different years is available from the NHIS Web site. Oversampling of African Americans and Hispanics has been included in the NHIS sampling plan since 1995 to allow more accurate estimates for those subgroups. In 2004, the NHIS collected data from 94,460 individuals in 36,579 households, with a response rate of 86.9%. The NHIS data, questionnaires, and ancillary information including documentation and syntax files to process the data are available from the NHIS Web page, as are tables and reports drawn from the NHIS summarizing different aspects of the health of the U.S. population.

—Sarah Boslaugh

See also Centers for Disease Control and Prevention; Governmental Role in Public Health; Health Behavior; National Center for Health Statistics; Public Health Surveillance

Further Readings

Adams, P. F., & Schoenborn, C. A. (2006). Health behaviors of adults: United States, 2002–2004. *Vital Health Statistics*, *10*(230), 1–140.

Bloom, B., & Dey, A. N. (2006). Summary health statistics for U.S. children: National Health Interview Survey, 2004. *Vital Health Statistics*, *10*(227), 1–85.

Web Sites

National Health Interview Survey (NHIS): http://www.cdc.gov/nchs/nhis.htm.

National Immunization Survey

The National Immunization Survey (NIS) began collecting data in 1994 for the purpose of establishing estimates of up-to-date immunization levels in each state, the District of Columbia, and 27 large urban areas. The NIS is conducted by the National Opinion Research Center for the Centers for Disease Control and Prevention; it is jointly sponsored by the National Immunization Program and the National Center for Health Statistics. Children between the ages of 19 and 35 months living in the United States at the time of the survey are the target population for the NIS: Data about their immunizations are collected from a parent or other adult from the child's household who is knowledgeable about their immunization record. If possible, this information is confirmed with the child's immunization providers.

The vaccinations included in the NIS are those recommended by the Advisory Committee on Immunization Practices, which are currently 4 doses of diphtheria, tetanus, and acellular pertussus vaccine (DTaP); 3 doses of polio vaccine; 1 dose of measles/mumps/rubella (MMR) vaccine; Haemophilus influenzae Type b vaccine (Hib); hepatitis A vaccine (Hep A); 3 doses of hepatitis B vaccine (Hep B); 1 dose of varicella zoster vaccine (chicken pox); 4 doses of pneumococcal conjugate vaccine (PCV); and influenza vaccine. Hepatitis A is recommended only in selected states with a high incidence of the disease. All vaccines except varicella, influenza, and pneumococcal have been included in the NIS since its inception: Pneumococcal was added in 2002, and influenza and hepatitis A were added in 2003.

The NIS collects data in two ways: through telephone interviews with households selected through random-digit dialing and through a mail survey of physicians and other vaccination providers; the latter is called the Provider Record Check Study. The telephone interview collects information from parents of eligible children about the immunizations each child has received, the dates of the immunizations, and the demographic and socioeconomic information about the household. If the parent grants permission, the child's vaccination providers are contacted to verify the child's vaccination record. The state and local estimates of vaccination coverage are calculated every quarter using NIS data and are used to evaluate progress toward national and local immunization goals. The coverage for series of vaccines is also reported, including the 4:3:1:3:3 series (4+ DTaP, 3+ polio, 1+ MMR, 3+ Hib, and 3+ Hep B).

—*Sarah Boslaugh*

See also Centers for Disease Control and Prevention; Child and Adolescent Health; National Center for Health Statistics; Survey Research Methods; Vaccination

Further Readings

Smith, P. J., Hoaglin, D. C., Battaglia, M. P., Barker, L. E., & Khare, M. (2005). Statistical methodology of the National Immunization Survey, 1994–2002. National Center for Health Statistics. *Vital and Health Statistics*, *2*(138).

Web Sites

National Immunization Survey (NIS) data and documentation are available for download from the NIS Web site: http://www.cdc.gov/nis/datafiles.htm. (Currently data for the years 1995–2004 are available.)

National Institutes of Health

The National Institutes of Health (NIH) is a U.S. Federal government agency, located in a suburb of Washington, D.C., that disburses more than $28 billion annually to fund biomedical research. The extramural, or granting, program, distributes 80% of the funds to universities and foundations in the United States and across the world. The scientists in the intramural, or on-campus, program conduct basic and clinical research based in 27 institutes and centers, including a brand-new (2004) 242-bed research hospital.

The NIH traces its origins to 1887 and a one-room laboratory in Staten Island, New York. Originally conceived of as the research division of the Marine Hospital Service (MHS), the NIH, then known as the Hygienic Laboratory, was an experiment. The MHS had been charged with preventing people with cholera, yellow fever, and other infectious diseases

from entering the United States: Perhaps a laboratory could help public health officials understand these diseases better to prevent future epidemics.

The experiment worked. Within 5 years, the Congress deemed the laboratory worthy of expansion and moved its facilities to Washington, D.C., to be closer to the seat of the Federal government. Its first director, Joseph Kinyoun, was charged with tasks such as cleaning up the city's water supply and reducing its air pollution. In 1901, the Congress authorized $35,000. The next year, the divisions of the new laboratories were formalized. The Division of Pathology and Bacteriology was joined by the Divisions of Chemistry, Pharmacology, and Zoology. The professional staff was filled out with scientists with doctoral degrees rather than physicians, emphasizing the importance of basic research.

In its initial years, epidemiological studies made up much of the work of the agency. In 1906, Hygienic Laboratory workers pursued a landmark study of typhoid in Washington, D.C., in which they identified the milk supply as the culprit in spreading the disease. In the next decade, new diseases were elucidated, such as tularemia, and old diseases were explored. For example, groundbreaking use of epidemiological techniques allowed Joseph Goldberger to prove that pellagra was caused by a vitamin deficiency and was not an infectious disease as had been previously assumed. In the next few decades, NIH scientists pursued studies of diseases in varied communities, from endemic and epidemic typhus in the South to Rocky Mountain spotted fever in the West. They identified and confirmed the vectors (spreaders) of diseases, such as showing, for example, that the body louse was responsible for spreading epidemic typhus fever.

The early decades of the 20th century brought other scientific advances in the area of public health. The Hygienic Laboratory, in charge of regulating biologics before the 1971 creation of the Food and Drug Administration, established the standards for antitoxins. Working to prevent diseases as well as to understand them, scientists studied water pollution and sewage, with important results in the creation of pure water systems. Evans, studying undulant fever, helped officials decide to call for the pasteurization of milk to prevent this and other illnesses. And in the 1920s, Hygienic Laboratory scientists studied the relationship between canning and food poisoning. This type of research led to better public health guidelines.

In 1930, major changes came to NIH. The Congress renamed the Hygienic Laboratory the National Institute (singular) of Health and authorized the payout of fellowship money for basic research. These fellowships form the majority of the NIH's research program today. In 1937, with the founding of the National Cancer Institute, the Congress began a several-decade process of opening new institutes at NIH to deal with specific diseases. Cancer, a chronic disease, marked an important switch away from the agency's long focus on infectious diseases. Institutes would often be formed around a single disease, such as heart disease, mental health, and diabetes. In the late 1930s, the NIH campus moved to its current location in suburban Bethesda, Maryland, where the operation expanded into several buildings specially equipped with up-to-date facilities.

During World War II, the scientists at NIH worked with the military to analyze the reasons why so many potential inductees were unfit for general military service. The two common causes of rejection were defective teeth and venereal disease. These realizations led to new funding for research into these areas. Other divisions worked on issues such as dangers in war-related industries. Research into vaccines expanded during the war years, and scientists worked on vaccines for yellow fever and typhus, specifically for military forces. The researchers at the NIH campus in Bethesda teamed up to find an alternative to quinine for prevention and treatment of malaria, a major scourge for American troops overseas.

In the post–World War II era, the NIH became more recognizable as the agency that it is today. When the grants program was expanded to the entire agency, the total budget expanded from $8 million in 1947 to more than $1 billion in 1966. In these years, the Congress designated more institutes to focus on specific diseases: lifecycle research on childhood and aging and drug and alcohol abuse. By 1960, there were 10 components. By 1970, this number increased to 15, and in 2006 the NIH had 27 institutes and centers.

Epidemiological work in the mid-20th century done by the NIH was spread around the world. Using data obtained from local neighborhoods and from infants and children housed at an institution in Washington, D.C., Robert Huebner and his colleagues identified new viruses. An expanded vaccine research program helped find methods to combat them. The Nobel Laureate Carleton Gadjusek

identified the kuru virus prevalent among the South Fore people of New Guinea as stemming from a particular funerary practice. Kuru was later identified as a prion disease. Baruch Blumberg and others discovered the Australian antigen during their work in the 1950s and 1960s. This led to a test to screen donated blood for hepatitis B, greatly reducing the risk of transfusion hepatitis. One of the NIH's most famous long-term epidemiological studies was the Heart Disease Epidemiology study at Framingham, Massachusetts, which started in 1949 and followed subjects for many years while recording notes about their diet and lifestyles. After 1946, the Centers for Disease Control in Atlanta took over much of the NIH's epidemiology work, especially in terms of identifying the causes of epidemics.

A major asset was the opening of a hospital to the NIH main campus in Bethesda. The Clinical Center, which opened in 1953 and expanded into a state-of-the-art new building in 2004, was specially designed to bring research laboratories into close proximity with hospital wards to promote productive collaboration between laboratory scientists and clinicians.

The NIH budget slowed down considerably in the 1960s and 1970s. This was due in part to wariness on the part of the Congress and the public about the efficacy of basic research in solving major health crises. However, the HIV/AIDS crisis in the 1980s highlighted the need for basic research in immunology and other disciplines. The important NIH research in the late 20th century also included studies that helped demonstrate that recombinant DNA research did not pose great risk of unleashing deadly novel organisms, leading to a rapid increase in molecular studies of disease. In the late 1980s, the NIH and the Department of Energy launched the Human Genome Project with the goal of mapping and sequencing the entire collection of human genes. The 1990s saw an emphasis in research on women's health with the Women's Health Initiative. And there have been many other successes: More than 100 NIH-funded scientists have won the Nobel Prize, including 5 who did their prize-winning work in the intramural program.

The turn of the 21st century has been a period of regrowth. The NIH Director Elias Zerhouni started the Road Map Initiative to direct monies to research problems that needed the attention of the entire agency rather than just one institute. The doubling of the NIH budget between 1998 and 2003 was meant to jump-start research: With an FY06 budget of more than $28 billion, the NIH today would barely be recognizable to Joseph Kinyoun and his staff of the one-room laboratory more than a century earlier.

—Sarah A. Leavitt

See also Centers for Disease Control and Prevention; Food and Drug Administration; Framingham Heart Study; Goldberger, Joseph; Governmental Role in Public Health

Further Readings

Farreras, I. G., Hannaway, C., & Harden, V. A. (Eds.). (2004). *Mind, brain, body, and behavior: Foundations of neuroscience and behavioral research at the National Institutes of Health.* Amsterdam: IOS Press.

Hannaway, C., Harden, V. A., & Parascandola, J. (Eds.). (1995). *AIDS and the public debate: Historical and contemporary perspectives.* Amsterdam: IOS Press.

Harden, V. A. (1986). *Inventing the NIH: Federal Biomedical Research Policy, 1887–1937*. Baltimore: Johns Hopkins University Press.

Harden, V. A. (1990). *Rocky Mountain spotted fever: History of a twentieth-century disease.* Baltimore: Johns Hopkins University Press.

Kraut, A. M. (2003). *Goldberger's War: The life and work of a public health crusader.* Boston: Hill & Wang.

Web Sites

National Institutes of Health: www.nih.gov.
Office of NIH History: www.history.nih.gov.

National Maternal and Infant Health Survey

The National Maternal and Infant Health Survey (NMIHS) was a longitudinal study of factors related to adverse pregnancy outcomes. Conducted by the National Center for Health Statistics in 1988 and 1991, the latter is often referred to as the Longitudinal Follow-Up. The NMIHS was conducted to augment data available in vital statistics records, collecting information on maternal sociodemographic characteristics, pregnancy history, health status, and health care types and sources. The vital statistics records included birth, fetal death, and infant death records. The NMIHS was the first national survey conducted in the United States to collect data simultaneously on births, fetal deaths, and infant deaths.

The NMIHS data were collected by questionnaires mailed to a nationally representative survey sample of women who gave birth or had a fetal or infant death in 1988. The data were collected from 9,953 women who gave birth, 3,309 who had fetal deaths, and 5,532 who had infant deaths. The NMIHS data are weighted to be representative of all births, fetal deaths, and infant deaths in 1988. Additionally, 93% of mothers consented to have their health care providers contacted. Questionnaires were administered to physicians, hospitals, and other health care providers linked to the outcomes, and this information was added to that collected from the individual mothers.

The mother's questionnaire was 35 pages long and included detailed questions on prenatal care; health during pregnancy; use of birth control; breastfeeding; desire for the pregnancy; use of tobacco, alcohol, and drugs; and demographic and socioeconomic characteristics of the mother and father. The provider questionnaire included questions about prenatal and postpartum care, medication use, diagnostic and other procedures performed, and infant health status.

The data from the NMIHS have been used by researchers to study a range of issues related to maternal and child health and well-being. The findings include the following:

- High income inequality was associated with an increased risk of depression and poor physical health, especially for the poorest fifth of women with young children.
- Low birthweight and early-childhood asthma were strongly and independently linked, with an estimated 4,000 excess asthma cases ascribed to low birthweight.
- Women with depressive symptoms after delivery were significantly more likely to report child behavior problems, including temper tantrums and difficulties interacting with other children, than those who did not report depressive symptoms.

The data from the NMIHS are publicly available. The data from the 1988 and 1991 surveys are available separately; however, to protect the confidentiality of respondents, the data sets can be linked only at the Research Data Center of the National Center for Health Statistics in Hyattsville, Maryland.

—Anu Manchikanti

See also Child and Adolescent Health; Maternal and Child Health Epidemiology; Pregnancy Risk Assessment and Monitoring System; Preterm Birth

Further Readings

National Center for Health Statistics. (2005). *National Maternal and Infant Health Survey*. Retrieved August 15, 2005, from http://www.cdc.gov/nchs/about/major/nmihs/abnmihs.htm.

Sanderson, M., & Gonzalez, J. F. (1998). 1988 National Maternal and Infant Health Survey: Methods and response characteristics. *Vital and Health Statistics. Series 2, Data Evaluation and Methods Research*, *125*, 1–39.

National Mortality Followback Survey

The National Mortality Followback Survey (NMFS) has been conducted sporadically since 1961 by the National Center for Health Statistics (NCHS) to collect information on various specific topics related to mortality in the United States, including the events and circumstances that preceded death, and relevant characteristics of the decedent. The NMFS has been conducted six times: four times in the 1960s, in 1986, and most recently, in 1993.

The main national database for the United States mortality statistics (the National Death Index) is maintained by the NCHS and is compiled from the official death certificate registries maintained by individual states. The NMFS was created to enrich the national mortality database by collecting information not available from death certificates. Each of the six surveys has been unique; each has focused on different topics and has used a variety of survey instruments and sources. The objective has been to focus on specific topics of current interest to public health researchers and policymakers, using instruments designed to fit the purposes of the individual survey rather than using a uniform instrument to gather data on a consistent group of topics.

All the surveys have obtained information from persons identified on the death certificates as informants (next of kin or a person familiar with the decedent). The surveys have collected information by means of mail questionnaires, personal interviews, and/or telephone interviews. The sample for each survey has been

drawn from the Current Mortality Sample, a systematic 10% of the states' death certificates.

The topics covered have varied from survey to survey. The 1961 survey sought information on the use of hospital and institutional care during the last year of life. The 1962 to 1963 survey focused on socioeconomic factors. The 1964 to 1965 survey obtained data on health care expenditures during the last year of life, sources of payment, and health insurance coverage. The 1966 to 1968 survey focused on the link between smoking and cancer mortality.

The two most recent surveys have been more comprehensive (covering a larger sample and more information items and sources) than the surveys conducted in the 1960s, and they have attempted to provide comparability with the data obtained by previous surveys.

The 1986 survey covered three topics: socioeconomic factors, risk factors related to premature death, and health care provided in the last year of life. Brief questionnaires were mailed to all hospitals, nursing homes, and other health care facilities reportedly used by decedents in their last year of life. The sample (18,733) was an approximate 1% sample of all deaths of adults over age 25 in 1986.

The 1993 survey included the 1986 topics and added a fourth topic, disability in the last year of life. The 1993 survey sought information from medical examiner/coroner offices if death was due to homicide, suicide, or accidental injury. The 1993 sample included 22,957 deaths in 1993 of persons aged 15 and older.

Both the 1986 and 1993 surveys examined the reliability of information reported on the death certificate by comparing it with the information reported by the survey respondent. The comparable items included age, race, gender, veteran status, education, occupation, and industry.

The two most recent surveys each included all states except one. In 1986, Oregon was not represented, and in 1993, South Dakota was not sampled, due to state restrictions on the use of death certificate information. Public use data files from the 1986 and 1993 NMFS surveys are available for purchase through the National Technical Information Service. The reports based on the surveys have been published by the National Center for Health Statistics.

—Judith Marie Bezy

See also Death Certificate; National Center for Health Statistics; National Death Index

Further Readings

National Center for Health Statistics. (n.d.). *National Mortality Followback Survey (NMFS) Public-Use Data Files*. Retrieved September 7, 2006, from http://www.cdc.gov/nchs/products/elec_prods/subject/nmfs.htm.

Seeman, I. (1992). National Mortality Followback Survey: 1986 Summary, United States. *Vital and Health Statistics*, *20*(19).

Seeman, I., Poe, S., & Powell-Griner, E. (1993). Development, methods, and response characteristics of the 1986 National Mortality Followback Survey. Hyattsville, MD: National Center for Health Statistics. *Vital and Health Statistics*, *1*(29).

National Survey of Family Growth

The National Survey of Family Growth (NSFG) is a nationally representative survey of adults in the United States that collects data on topics related to sexual behavior, fertility, infant health, marriage, and family life. It has been conducted periodically since 1973, and the first five administrations of the NSFG (1973, 1976, 1982, 1988, and 1995) collected information from women aged between 15 and 44 years only (considered the age range most likely to give birth); the most recent NSFG, conducted in 2002, collected information from both men and women and included more questions about sexual behavior than the previous surveys. Although the NSFG sample is nationally representative, it is not sufficiently large to allow analyses at geographic levels smaller than the four census regions (Northeast, Midwest, South, and West); metropolitan area versus nonmetropolitan area analyses are also supported.

The NSFG data are collected through in-person home interviews. For the 2002 administration, responses to particularly sensitive questions on topics such as sexual orientation, number of sexual partners, safe sex practices, and pregnancy terminations were collected through Audio Computer-Assisted Self-Interviewing (ACASI). This is an interview technique in which the respondent enters answers to questions directly into the computer rather than responding verbally to questions posed by the interviewer; the purpose of this technique is to gather more honest responses to sensitive questions by freeing the respondent from discussing personal issues in front of another person. The survey sample for each of

the first five administrations of the NSFG included about 8,000 to 10,000 women, sampled from the civilian, noninstitutionalized population of women aged between 15 and 44 years living in the 48 contiguous United States; all 50 states were included beginning with the fourth administration. The 2002 NSFG survey sample included 7,643 women and 4,928 men.

For the first five administrations of the NSFG, data were collected into two types of files: an *interval*, or *pregnancy*, file and a *respondent* file. The data collected have varied somewhat in each administration, but in general the *pregnancy* file contains information about topics such as contraceptive use, prenatal care, pregnancies, and births, and the *respondent* file includes personal and demographic information about the women surveyed, such as education, race, employment, marital status, living arrangements, family size, number of pregnancies and adoptions, health insurance coverage, and child care arrangement. The 2002 NSFG collected data in three types of files: a *female respondent* file, a *male respondent* file, and a *female pregnancy* file. The *female respondent* and *male respondent* files contain information similar to the respondent file of the first five administrations, while the *female pregnancy* file includes the woman's pregnancy history.

Most NSFG data are available on CD-ROM from the National Center for Health Statistics Web site and on data tapes from the National Technical Information Service. The sensitive data collected by ACASI are not included in these sources: Due to confidentiality concerns, access to these data requires a special application to the NSFG.

—*Sarah Boslaugh*

See also Interview Techniques; Reproductive Epidemiology; Sexually Transmitted Diseases; Sexual Risk Behavior; Women's Health Issues

Further Readings

Lepkowski, J. M., Mosher, W. D., Davis, K. E., Groves, R. M., van Hoewyk, J., & Willem, J. (2006). National Survey of Family Growth, Cycle 6: Sample design, weighting, imputation, and variance estimation. *Vital Health Statistics*, *2*(142), 1–82.

Web Sites

National Survey of Family Growth: http://www.cdc.gov/nchs/nsfg.htm.

Native American Health Issues

See American Indian Health Issues

Natural Experiment

A natural experiment is an observational study that takes advantage of a naturally occurring event or situation that can be exploited by a researcher to answer a particular question. Natural experiments are often used to study situations in which a true experiment is not possible, for instance, if the exposure of interest cannot be practically or ethically assigned to research subjects. Situations that may create appropriate circumstances for a natural experiment include policy changes, weather events, or natural disasters. This entry describes natural experiments, examines the limitations to such experiments that exist as a result of confounding, and discusses the use of instrumental variables to control confounding.

The key features of experimental study designs are manipulation and control. Manipulation, in this context, means that the experimenter can control which research subjects receive which exposures: For instance, those randomized to the treatment arm of an experiment typically receive treatment from the drug or therapy that is the focus of the experiment, while those in the control group receive no treatment or a different treatment. Control is most readily accomplished through random assignment, which means that the procedures by which participants are assigned to a treatment and control condition ensure that each has equal probability of assignment to either group. Random assignment ensures that individual characteristics or experiences that might confound the treatment results are, on average, evenly distributed between the two groups. In summary, then, an experiment is a study in which at least one variable is manipulated and units are randomly assigned to the different levels or categories of the manipulated variables.

Although the gold standard for epidemiologic research is often considered to be the randomized control trial, this design can answer only certain types of epidemiologic questions, and it is not useful in the investigation of questions for which random assignment is either impracticable or unethical. The bulk of

epidemiologic research relies on observational data, which raises issues in drawing causal inferences from the results. A core assumption for drawing causal inference is that the average outcome of the group exposed to one treatment regimen represents the average outcome the other group would have had if they had been exposed to the same treatment regimen. If treatment is not randomly assigned, as in case of observational studies, the assumption that the two groups are exchangeable (on both known and unknown confounders) cannot simply be assumed to be true.

For instance, suppose an investigator is interested in the effect of poor housing on health. Because it is neither practical nor ethical to randomize people to variable housing conditions, this subject is difficult to study using an experimental approach. However, if a housing policy change such as a lottery for subsidized mortgages was enacted that enabled some people to move to more desirable housing while leaving other similar people in their previous substandard housing, it might be possible to use that policy change to study the effect of housing change on health outcomes. One well-known natural experiment occurred in Helena, Montana, where smoking was banned from all public places for a 6-month period. The investigators reported a 60% drop in heart attacks for study area during the time the ban was in effect.

Because natural experiments do not randomize participants into exposure groups, the assumptions and analytical techniques customarily applied to experimental designs are not valid for them. Rather, natural experiments are quasi experiments and need to be thought about and analyzed as such. The lack of random assignment means multiple threats to causal inference, including attrition, history, testing, regression, instrumentation, and maturation, may influence observed study outcomes. For this reason, natural experiments will never unequivocally determine causation in a given situation. Nevertheless, they are a useful method for researchers and if used with care can provide additional data that may help with a research question and that may not be obtainable in any other way.

Instrumental Variables

The major limitation in inferring causation from natural experiments is the presence of unmeasured confounding. One class of methods designed to control confounding and measurement error is based on instrumental variables (IV). Although these variables have been used in economics for decades, they are little known in epidemiology. While useful in a variety of applications, the validity and interpretation of IV estimates depend on strong assumptions, the plausibility of which must be considered with regard to the causal relation in question.

If we are interested in the causal effect of X (exposure) on Y (outcome), and we can observe their relation to a third variable Z (IV or instrument) that is associated with X but not with Y (except through its association with X), then under certain conditions we can write the $Z-Y$ association as the product of $Z-X$ and $X-Y$ associations as follows:

$$\text{Assoc } ZY = \text{Assoc } ZX \times \text{Assoc}XY$$

and solve this equation for the XY association.

This equation is particularly useful when (1) the XY relationship is confounded by unmeasured covariates (but the ZX and ZY relationships are not) or (2) the XY relationship cannot be directly observed, but Z is an observable surrogate, or instrument, for X.

IV analyses use data from researcher-randomized or natural experiments to estimate the effect of an exposure on those exposed. IV analyses depend on the assumption that subjects were effectively randomized, even if the randomization was accidental (in the case of an administrative policy change or exposure to a natural disaster) and/or adherence to random assignment was low. IV methods can be used to control for confounding in observational studies, control for confounding due to noncompliance, and correct for misclassification.

Confounding in Observational Studies

Administrative policies, government legislation, and other external forces often create natural or quasi experiments in which an individual's probability of exposure is affected by forces uncorrelated with individual-level health outcomes. Such policies could be used as an instrument in epidemiologic analysis. For instance, if we are interested in the causal effect of X (income) on Y (self-rated health), and we can observe their relation to a third variable Z (an IV or instrument, in this case, a state-level increase in the minimum wage), and we know the relation between X and Y is confounded by U (unobserved or unmeasured variables, such as race, wealth, area-level economic health, etc.), then we can use Z to estimate the

relationship between X and Y provided it meets both the following assumptions:

1. Z is associated with X.
2. Z is not associated with Y.

We cannot use Z to estimate the relationship between X and Y if any of the following is true:

1. Z is associated with Y.
2. Z is associated with U.
3. U is associated with Z.

Characteristics of a Good Instrument

To even consider a potential instrument for use in IV, it must meet the following three assumptions: (1) the instrument (Z) must be associated (in known and measurable ways) with the exposure (X), (2) the instrument (Z) cannot be associated with outcome (Y), and (3) the deviation of the instrument (Z) from the exposure (X) is independent of other variables or errors. Furthermore, to interpret IV effect estimates as causal parameters, we need to further assume that the direction of the effect of Z on X is the same for everyone in the sample (the monotonicity assumption). Any instrument not meeting these assumptions for the causal contrast in question is not appropriate for use in analysis.

Limitations of IV Analysis

IV analysis is limited, particularly in light of the assumptions imposed on the relationship between the instrument and the exposure and outcome. Even a small association between the instrument and the outcome, which is not solely mediated by the exposure of interest, can produce serious biases in IV effect estimates. It can also be difficult to identify the particular subpopulation to which the causal effect IV estimate applies (those whose exposure would be affected by the instrument if offered). If the relationship between the instrument and the exposure is weak, IV analysis can add considerable imprecision to causal effect estimates. The multiple instruments can be useful both to improve the statistical power of an IV analysis and, if power is adequate, to test the validity of instruments against one another. Some instruments may be valid only after conditioning on a measured covariate. A common prior cause of the instrument and the outcome renders the instrument invalid unless that confounder can be measured and statistically controlled. Small sample size poses an additional challenge in applying IV methods.

—Lynne C. Messer

See also Causal Diagrams; Causation and Causal inference; Quasi Experiments; Study Design

Further Readings

Glymour, M. (2006). Natural experiments and instrumental variable analysis in social epidemiology. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in social epidemiology* (pp. 429–460). San Francisco, CA: Jossey-Bass.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, *29*, 722–729.

Sargent, R. P., Shepard, R. M., & Glantz, S. A. (2004). Reduced incidence of admissions for myocardial infarction associated with public smoking ban: Before and after study. *British Medical Journal*, *328*(7446), 977–980.

Negative Binomial Regression

See Regression

Negative Predictive Value

See Clinical Epidemiology

Network Analysis

Over the past two decades, the epidemic of HIV has challenged the epidemiological community to rethink the framework for understanding the risk of infectious disease transmission, both at the individual level and at the level of population transmission dynamics. Research has rapidly converged on the central importance of partnership networks. Systematic patterns in social networks have always served to channel infectious diseases—from the sequence of plagues in

Europe and the introduction of European childhood infections into the Native American populations to the polio epidemics of the early 20th century and the contemporary outbreaks of cholera and typhoid that attend mass movements of refugees. The methodology for network data collection and analysis in epidemiology, however, is only now being developed. This entry examines the role of networks in disease transmission, as well as the origins of network analysis in social science and epidemiology. The entry focuses on the role of social networks in sexually transmitted infections (STIs), where networks determine the level of individual exposure, the population dynamics of spread, and the interactional context that constrains behavioral change. Network analysis has had the largest impact in this field, and it represents a paradigm shift in the study of STI.

Types of Transmission Via Social Networks

Like the movement of exchangeable goods, the diffusion of pathogens through a human population traces the structure of social networks. The pattern of spread is jointly determined by the biology of the pathogen and the social structure that can support it, so different kinds of diseases travel along different structural routes. The plague, for example, is spread by a mobile vector of rats and fleas that makes for an efficient, long-lasting infectious vehicle. The disease can travel via long-distance transportation and trade routes even when travel is slow paced, with macroeconomic relations helping to structure the diffusion path. For influenza and measles, in contrast, transmission requires casual or indirect personal contact in a relatively short period of time. The spread of these infections is structured by locations of frequent collective activity, such as schools and supermarkets today, with transportation networks serving as potential bridges between communities and sparsely settled or less traveled routes serving as buffers. Finally, there are infections spread only by intimate or prolonged contact; STIs are a classic example. These diseases travel along the most selective forms of social networks, operating on what is comparatively a very sparse microstructure, with a typically modest duration of infection. The structure of sexual networks varies within and between societies, governed by local norms, power differentials, and oppositional subcultures. Here, as with other infectious diseases, the transmission network determines the potential for epidemics and the opportunities for prevention.

Network Epidemiology and Sexually Transmitted Infections

Network epidemiology offers a comprehensive way of thinking about individual sexual behavior and its consequences for STI transmission. Unlike other health-related behaviors (e.g., smoking) and safety-related behavior (e.g., using seat belts), behaviors that transmit STIs directly involve at least two people, as well as other persons to whom they may be linked. Understanding this process requires moving beyond the standard, individual-centered research paradigm. This has important implications for the analytic framework, data collection, and intervention planning.

The analytic framework must take a relational approach, integrating individual behavior into partnership contexts, and aggregating partnership configurations into networks. This is a marked departure from the standard approach to behavioral research that seeks to link individual attributes to individual outcomes. The data collection and statistical analysis need to be revised accordingly, making the partnership—rather than the individual—the primary sampling unit. While we know a lot about sampling individuals, we know much less about sampling partnerships and networks. Finally, analyzing the network data that are collected requires different statistical methods, since the defining property of such data is that the units are not independent. The methods for analyzing dependent data are not unknown—spatial statistics, time series, and multilevel models provide a starting point—but the statistical tools needed to analyze networks have only recently been developed.

Given these difficulties, why bother taking a network approach? Why not simply focus on individual risk factors for acquisition of disease? The answer is that network epidemiology succeeds where more traditional epidemiological approaches have failed: explaining differentials in risk behavior, epidemic potential in low-risk populations, and the persistent and substantial prevalence differentials across populations.

In one sense, a network explanation is almost tautological: Individuals are infected by their partners, who are in turn infected by their partners—*networks* is just a term that describes this process. But the

concept also has explanatory power and prevention implications, as it changes the focus from "what you do" to "whom you do it with." This allows for behaviors to vary within as well as between persons, and for the same individual behavior to lead to different infection outcomes in different contexts.

As a result, the network perspective changes the way we think about targeting concepts such as "risk groups" and "risk behaviors." The inadequacy of these concepts became clear as HIV prevalence rose among groups that do not engage in individually risky behavior, for example, monogamous married women. By the same token, a group of persons with extremely "risky" individual behavior may have little actual risk of STI exposure if their partners are uninfected and are not linked to the rest of the partnership network. It is not only individuals' behavior that defines their risk but also the behavior of their partners and (ultimately) their position in a network.

The network perspective also changes the way we think about the population-level risk factors. The key issue is not simply the mean number of partners but the connectivity of the network, and connectivity can be established even in low-density networks. One of the primary ways in which this happens is through concurrent partnerships. Serial monogamy in sexual partnerships creates a highly segmented network with no links between each pair of persons at any moment in time. If this constraint is relaxed, allowing people to have more than one partner concurrently, the network can become much more connected. The result is a large increase in the potential spread of STIs, even at low levels of partnership formation.

Finally, the network perspective changes the way we think about behavior change. Because the relevant behavior occurs in the context of a partnership, individual knowledge, attitudes, and beliefs do not affect behavior directly. Instead, the impact of these individual-level variables is mediated by the relationship between the partners. A young woman who knows that condoms help prevent the sexual spread of HIV may be unable to convince her male partner to use one. It is not her knowledge that is deficient, but her control over joint behavior.

Origins of the Field

The analysis of network structures has a relatively long history in the social sciences, which is where the most comprehensive methodology has been developed, and the study of diffusion through structured populations in epidemiology is also long standing. In recent years, physicists have also begun to work on the epidemiology of diseases on networks.

Social Science Roots

Social network analysis is an established subfield in the social sciences, with a professional organization (the International Network for Social Network Analysis [INSNA]), an annual meeting (the "Sunbelt" social network conference, now in its 26th year), and several journals (*Social Networks*, *Connections*). It is an interdisciplinary field, and it has developed a unique set of methodological tools.

For the past 25 years, social network analysts have been developing quantitative tools for empirical studies. There are several distinct approaches, defined by the type of data collected. The first is based on a network census—data collected on every node and link for a (typically small) population. The current textbooks and most popular computer packages for social network analysis have these methods at their core. The second approach is based on sampled network data. The most well-known of these is the local network (or egocentric) sample design: a sample of the nodes (egos), with a "name generator" in the questionnaire to obtain a roster of their partners (alters), and "name interpreters" to collect information on these partners (for a good example and discussion). With this simple study design, no attempt is made to identify or enroll the partners. Local network data collection costs about the same as a standard survey, is relatively easy to implement, and is less intrusive than complete network data collection. In between these two approaches lies a range of link-tracing designs for collecting network data—snowball samples, random walks, and most recently, respondent-driven sampling. The absence of methods for analyzing such data previously limited the use of this approach, but new methods are now available and are becoming more common.

Rapid progress is being made now in the development of methodology for network analysis. Methods have been developed to handle networks sampled with egocentric and link-tracing designs. Statistical theory is being developed for estimation and inference, which is complicated for networks given the dependent data and nonlinear threshold effects. The class of models being developed not only can represent an arbitrarily complex network structure but can

also test the goodness of fit of simple parsimonious models to data. A computer package (statnet) has been released that allows researchers to use these methods for network analysis. The algorithm used for network estimation in this package can also be used for simulation. So, for the first time, researchers can simulate networks using models and parameters that have been derived from data and statistically evaluated for goodness of fit.

Roots in Epidemiology

Epidemiologists had a tradition of modeling infectious disease spread through "structured populations" well before the explicit connection was made to social network analysis. The spatial spread of infections was an early focus, with models for the dynamics of childhood disease transmission among families, neighborhoods, schools and playgroups, epidemics on islands, and pandemics spreading through the network of airline routes. The models for wildlife disease transmission, especially rabies, were built around small interacting subgroups connected by occasional long jump migrations, anticipating the "small world" models in the recent physics literature.

Simple network-like models for sexually transmitted pathogens began to be developed in the late 1970s and early 1980s when, despite the relative availability of penicillin, gonorrhea and syphilis rates rose precipitously in the United States. The surveys of STI clinic patients in the late 1970s found that repeat cases contributed disproportionately to the total caseload: A total of 3% to 7% of the infected persons accounted for about 30% of the cases. Simulation studies showed that this group can act as a "reservoir," allowing an infection to persist in a population where the average level of activity is otherwise too low to allow for sustained transmission. This research led to the "core group" theory: If endemic persistence is due to this small core group, then all cases are caused directly or indirectly by the core. The core group thus came to be seen as the primary driving force in STIs, and also as the locus for highly effective intervention targeting. But concerns began to be raised about the limitations of the core group concept with the emergence of generalized HIV epidemics in the countries of sub-Saharan Africa.

The first explicit link between social network methods and STI dynamics was made by Alden Klovdahl in 1985. His paper was written before the virus that causes AIDS had been identified, and the mechanism of transmission had not been conclusively demonstrated. A large number of theoretical and empirical studies have since followed that use network concepts and methods to help understand the different patterns of HIV spread in different countries, the disparities in infection prevalence within countries, the behaviors that increase exposure risk, and the opportunities for prevention. At their best, network models help us understand the population-level implications of individual behavior: How the choices that individuals make link together and aggregate up to create the partnership network that either inhibits or facilitates transmission.

What Have We Learned?

Network methods are beginning to be used in a number of communicable disease contexts, including studies of the most effective intervention strategies for containing the spread of pandemic influenza (H5N1, the "avian (bird) flu"), containing the impact of bioterrorist attacks with agents such as smallpox, and reducing the spread of bovine spongiform encephalitis across farms. The most intensive use of the methods, however, is in the field of STI, and that is where the most detailed lessons have been learned.

Using network analysis, researchers have identified two basic behavioral patterns that have a large impact on the STI transmission network: selective mixing and partnership timing. Both are guided by norms that influence individual behavior, which in turn create partnership network structures that leave distinctive signatures on transmission dynamics and prevalence. Selective mixing is about how we choose partners: How many partnerships form within and between groups defined by things such as age, race, and sexual orientation. Assortative mixing leads to segregated networks that channel infection and can sustain long-term prevalence differentials, such as the persistent racial differentials observed in the United States. Partnership timing is about the dynamics of relationships: Monogamy requires partnerships to be strictly sequential, concurrency allows a new partnership to begin while an existing partnership is still active. Long-term monogamous pair formation slows down the rate of disease transmission, as concordant pairs provide no opportunity for spread, and discordant pairs remain together after transmission has occurred. Concurrent partnerships, in contrast, can dramatically amplify the speed of transmission.

Partnership networks also have other structural features that can be important for STI spread, including closed cycles (e.g., the triangles and odd-numbered cycles that can emerge in same-sex networks, and larger even-numbered cycles for heterosexual networks) and highly skewed distributions for the number of sexual partners.

Small differences in the pattern of contacts can have huge effects on the transmission network structure. An average increase of only 0.2 concurrent partners can be enough to fundamentally change the connectivity of a network and create a robust connected network core. This is important to remember when evaluating the significance of empirical differences in sample data. Most samples are not large enough to detect a difference this small as statistically significant, but the difference may still be substantively important. The good news is that, just as small changes may be enough to push transmission above the epidemic threshold in some groups, small changes may be all that are needed to bring transmission down below the threshold.

Persistent prevalence disparities across populations for a wide range of infectious diseases are a signal that the underlying transmission network is probably the cause. A combination of processes may be at work: assortative mixing (which segregates populations) and small variations in concurrent partnerships (which differentially raises the spread in some groups). The disparities that can be sustained in such networks can be surprisingly large, even when the behaviors do not appear to differ much by group. For this reason, it is important that we have an accurate empirical picture of the key aspects of the transmission network. We need to know what behavior needs to change, who needs to change it, and the relevant cultural contexts, so that our intervention efforts can be properly targeted and maximally effective.

—Martina Morris

See also Outbreak Investigation; Partner Notification; Sexually Transmitted Diseases

Further Readings

Borgatti, S., Everett, M., & Freeman, L. (1999). *UCINET (Version 5)*. Harvard, MA: Analytic Technologies.

Fischer, C. S. (1982). *To dwell among friends: Personal networks in town and city.* Chicago: University of Chicago Press.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). *Statnet: Software tools for the representation, visualization, analysis and simulation of social network data (Version 1.11)*. Seattle, WA: Center for Studies in Demography and Ecology, University of Washington.

Hethcote, H., & Yorke, J. A. (1984). *Gonorrhea transmission dynamics and control.* Berlin, Germany: Springer Verlag.

Kaplan, E. H., Craft, D. L., & Wein, D. L. (2003). Emergency response to a smallpox attack: The case for mass vaccination. *Proceedings of the National Academy Sciences*, *99*, 10395–10440.

Klovdahl, A. S. (1985). Social networks and the spread of infectious diseases: The AIDS example. *Social Science & Medicine*, *21*(11), 1203–1216.

Koehly, L., Goodreau, S., & Morris, M. (2004). Exponential family models for sampled and census network data. *Sociological Methodology*, *34*, 241–270.

Longini, I. M., Jr., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D. A., et al. (2005). Containing pandemic influenza at the source. *Science*, *309*(5737), 1083–1087.

Morris, M. (Ed.). (2004). *Network epidemiology: A handbook for survey design and data collection.* Oxford, UK: Oxford University Press.

Snijders, T. A. B. (2002). Markov Chain Monte Carlo estimation of exponential random graph models [Electronic version]. *Journal of Social Structure*, *3*(2). Retrieved from http://citeseer.ist.psu.edu/snijders02markov.html.

St. John, R., & Curran, J. (1978). Epidemiology of gonorrhea. *Sexually transmitted diseases*, *5*, 81–82.

Thomas, J. C., & Tucker, M. J. (1996). The development and use of the concept of a sexually transmitted disease core. *Journal of Infectious Diseases*, *174*(Suppl. 2), S134–S143.

Thompson, S. K., & Frank, O. (2000). Model-based estimation with link-tracing sampling designs. *Survey Methodology*, *26*(1), 87–98.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge, UK: Cambridge University Press.

Neuroepidemiology

Neuroepidemiology is the application of the methods of epidemiology to the problems of clinical neurology, to study the frequency of neurologic disorders, their risk factors, and their treatments. In addressing the distribution and determinants of neurologic disease in the population, the end goal of neuroepidemiology is to prevent or improve the outcomes of neurologic disease. According to World Health Organization (WHO) data,

neuropsychiatric disorders account for more than 10% of the global burden of disease. This entry discusses the unique issues one must address in neuroepidemiological studies; considers neurologic disorders either infectious in nature or common enough to have implications for public health, as well as uncommon disorders that offer important lessons in the study of such conditions; outlines the process of events that must occur for case identification for inclusion in neuroepidemiological studies; and considers important outcome measures evolving in this field.

Special Aspects of Neurologic Conditions

The human nervous system comprises the brain, spinal cord, and peripheral nerves. The brain and spinal cord constitute the central nervous system (CNS) while the remainder of the nervous system, including the lumbosacral and brachial plexus, constitute the peripheral nervous system (PNS). Virtually all bodily functions are controlled or regulated by the nervous system, including motor function, sensation perception, memory, thought, consciousness, and basic survival mechanisms such as regulation of heart rate and rhythms and respirations. As such, dysfunction in the nervous system can present with a vast spectrum of physical signs and symptoms, and the potential to misattribute neurologic disorders to other organ system is considerable.

To understand the complexities of neurologic investigations and disorders one must recognize that CNS tissue does not, as a general rule, regenerate or repair particularly well after insult or injury. Recovery is more likely to be mediated by plasticity in the system that allows alternate pathways to assume functions previously held by dead or damaged regions, and such plasticity is most abundant in infants and children, declining substantially with age. Unlike other organ systems, the CNS relies almost exclusively on glucose for metabolic function. The metabolic rate of the brain is extremely high and very vulnerable to injury if nutrients, such as glucose or oxygen, provided through blood flow are disrupted. Anatomically, the nervous system exists in a separate compartment from the rest of the body, being protected from traumatic injury by bony encasement (skull and spinal column) and from exogenous exposures by the blood-brain barrier and the blood-nerve barrier. Anatomic localization of injury or dysfunction in the nervous system is the most critical element in determining symptomatology—"Where is the lesion?" is the key mantra physicians address when first assessing a patient with potential neurologic disease, lesion location being one of the most important aspects for the development of a differential diagnosis and ultimately a clinical diagnosis leading to treatment.

Given the nervous system's poor capacity for regeneration, limited plasticity, and anatomic isolation (i.e., encased in the skull and/or spinal column), access to CNS tissue for pathologic diagnosis is often not possible. Since a vast array of physical signs and clinical symptoms can occur as a result of nervous system dysfunction and there is limited opportunity for pathologic diagnosis, expert physician evaluation and/or careful application by trained personnel using validated diagnostic criteria are the most crucial tools for case identification. A worldwide survey of available resources for neurological diagnosis and care conducted by the WHO and the World Federation of Neurology clearly illustrate the devastating lack of health care providers with neurologic expertise in most of the developing world, although developing regions suffer disproportionately from such conditions. Hence, lack of experts either to make the diagnosis or to develop appropriate diagnostic tools for population-based assessment remains a major barrier to neuroepidemiological studies in many regions of the world.

Barriers to Neuroepidemiological Studies

In their excellent textbook on neuroepidemiology, Nelson, Tanner, Van Den Eeden, and McGuire (2003) carefully outline the particular challenges to studying neurologic disorders from an epidemiologic perspective. These include the following:

- The diagnostic criteria vary across studies and over time. The resources-limited settings, particularly those without access to imaging or neurophysiologic studies, will be limited in their application of diagnostic criteria using such technologies.
- The definitive diagnosis may require postmortem examination by a qualified neuropathologist. The proportion of deaths with associated autopsy completion is declining in developed countries, and neuropathologists are nonexistent or very limited in less developed countries.
- The diagnosis during life may require an expert (e.g., neurologist).

- Many neurologic diseases, including infectious disorders with important public health implications, are relatively rare.
- The precise time of disease onset may be uncertain given the insidious onset of some symptoms such as memory loss or weakness.
- There is a long latent period before diagnosis in some diseases, and the duration of this may vary based on the clinical expertise and diagnostic technologies accessible to a person with the condition. Under some circumstances, diagnosis never occurs.
- Intermittent symptoms and signs occur in some diseases.
- Most neurologic diseases are not reportable, even in developed countries, and there are very few neurologic disease registries.

Neurologic Disorders of Particular Relevance to Neuroepidemiology

Some neurologic disorders and their particular relevance to neuroepidemiology are listed in Tables 1, 2, and 3.

Case Identification

Neurologic Symptoms and Patient Interpretation

For case identification to occur during life, persons with the disorder and/or their family members must recognize the symptoms as abnormal and seek care. This may seem trivial, but cultural interpretation of the

Table 1 Neurologic Disorders That Contribute Substantially to the Global Burden of Disease

Disorder	*Relevance*
Cerebrovascular disease (stroke)	This is responsible for $> 10\%$ of the global burden of disease with incidence increasing.
Cerebral palsy	This is a common, chronic disorder originating in childhood with associated motor problems resulting in lifelong disability (Nelson, 2002).
Bacterial meningitis	This is infectious and a common killer of children, especially in Africa's meningitis belt. Many types are preventable with vaccination.
Epilepsy	This is the most common, chronic neurologic disorder in many regions of the world. Despite available treatments, the treatment gap remains $> 85\%$ in most of the developing countries. Stigma-mediated morbidity is particularly problematic.
Dementia	This increases substantially in prevalence in developed regions as the average age of the population increases. The improved survival results in longer duration of disability. This is an emerging economic crisis for the United States.
Cognitive impairment from malnutrition	Chronic micronutrient deficiency and protein malnutrition in infancy and early childhood appears to place children at risk for permanent cognitive impairment. The loss of human capital related to lack of basic goods is likely staggering, though poorly quantified.
Cerebral malaria	It kills more than 1 million children annually, most of whom are in Africa. The emerging data confirm that survivors are at risk of neuropsychiatric, neurologic, and cognitive impairments.
Traumatic brain/spinal cord injury	It represents preventable cause of neurologic morbidity and mortality, often among young adults, for most regions of the world.
Headache disorders	These are an important cause for short-term, recurrent disability resulting in loss of productivity and absenteeism from work and/or school. It is associated with substantial decline in health-related quality of life for many affected individuals.

Table 2 Potentially Infectious Disorders With Implications for Public Health

Disorder	*Relevance*
Bacterial meningitis	This is infectious and a common killer of children, especially in Africa's meningitis belt. Many types are preventable with vaccination.
Epilepsy	This is the most common, chronic neurologic disorder in many regions of the world. Despite available treatments, the treatment gap remains >85% in most of the developing countries. Stigma-mediated morbidity is particularly problematic.
Cerebral malaria	This kills more than 1 million children annually, most of whom are in Africa. The emerging data confirm that survivors are at risk of neuropsychiatric, neurologic, and cognitive impairments.
Tetanus	This is common in developing countries due to unvaccinated mothers and inappropriate care of the umbilical cord in the neonate. This is a potentially preventable cause of mortality for those under 5 years of age.
New variant Creutzfeldt Jacob Disease (nvCJD)	This is a fatal neurodegenerative disorder, primarily among young adults related to ingestion of infected meat—a human illness resulting from the bovine epidemic of "mad cow disease." This is a prion-related infection.
Amyolateral sclerosis Parkinsonism dementia (ALS/PD) complex of Guam	This is an epidemic of fatal neurodegenerative disease not previously described that occurred in the Chamorro population of Guam. The etiology remains unclear, and the epidemic appears to be resolving or evolving (Wiederholt, 1999).
Kuru	This is a prion-medicated infectious disorder identified among the tribes of Papua New Guinea. Intense epidemiologic and anthropologic investigations identified traditional burial preparations involving aspects of cannibalism as the cause of this epidemic. Reviewing the history of the kuru epidemic and the associated investigations offers many "lessons" for neuroepidemiologists today.
CJD	This is a rapidly progressive neurodegenerative dementia with variable other features that may be genetic and/or infectious in nature. Rare but with the potential to spread through inappropriate sterilization of biopsy equipment or tissue handling.

symptoms of events such as seizures may affect where care is sought. In developing regions, seizures may be interpreted as the result of witchcraft or spiritual possession and care may be sought from traditional healers and/or clerics. Partial seizures with psychic phenomena (e.g., intense fear) that are not accompanied by generalized tonic-clonic seizures may be misinterpreted as psychiatric symptoms. Individuals with recurrent severe headaches may self-treat with over-the-counter medications and never seek formal medical care, particularly if the patient has limited access to medical treatment. Care for intermittent symptoms from conditions such as multiple sclerosis (MS) may be deferred or delayed until disability occurs, and this will result in prevalence skewed toward individuals with better access to care and a prognosis that is apparently worse for those from a lower socioeconomic status or regions geographically distant from advanced diagnostic services. When reviewing neuroepidemiologic data for different subpopulations, one should consider how early symptoms might be interpreted (or misinterpreted) by the population under study.

Health Care Provider Expertise

Even if someone suffering from nervous system dysfunction seeks physician-level care, the neurologic knowledge and expertise of the physician may determine whether or how rapidly a diagnosis is made. In the United States, many medical schools do not require graduating students to complete a rotation in clinical neurology. In developing regions, there may be no neurologist available to train medical students and postgraduates in training. The level of health care

Table 3 Conditions Often Under Epidemiologic Study for Etiologic and Prognostic Data

Condition	*Relevance*
Traumatic brain/spinal cord injury	This represents preventable cause of neurologic morbidity and mortality, often among young adults, for most regions of the world.
Parkinson's complex of Guam	This is an epidemic of fatal neurodegenerative disease not previously described that occurred in the Chamorro population of Guam. The etiology remains unclear, and the epidemic appears to be resolving or evolving.
Kuru	This is a prion-medicated infectious disorder identified among the tribes of Papua New Guinea. Intense epidemiologic and anthropologic investigations identified traditional burial preparations involving aspects of cannibalism as the cause of this epidemic. Reviewing the history of the kuru epidemic and the associated investigations offers many "lessons" for neuroepidemiologists today.
CJD	This is a rapidly progressive neurodegenerative dementia with variable other features that may be genetic and/or infectious in nature. Rare but with the potential to spread through inappropriate sterilization of biopsy equipment or tissue handling.
Multiple sclerosis	This is a chronic inflammatory disorder involving demyelination of the CNS. It is common among neurologic disorders. Usually nonfatal, but may result in substantial disability among people during their biologically and economically productive lifetime. The etiology remains unclear.
Parkinson's disease	This is a chronic, progressive neurodegenerative disorder characterized by motor abnormalities, including tremor. It is relatively common among neurologic disorders. The etiology remains unclear.
Amyolateral sclerosis (ALS)	This results in progressive loss of motor neurons resulting in progressive weakness with bulbar and respiratory weakness, usually resulting in death within ~1 year unless ventilatory support is provided. The etiology remains unclear.
Primary brain tumors	Tumor type and location are generally age dependent. Data are available through cancer registries. The etiology usually is unknown.

provider expertise is likely especially important in diagnosing uncommon conditions (e.g., Creutzfeldt-Jakob disease [CJD]) or uncommon presentations of common conditions (e.g., partial seizure disorder without secondary generalization).

Diagnostic Testing

The array of diagnostic testing available to provide supportive and confirmatory data of neurologic disorders is substantial, but these cannot be used as a substitute for clinical expertise. The conduct and interpretation of such studies requires additional expertise and are subject to misapplication in the wrong hands. Common diagnostic tests used in more developed regions are noted below. Further details are available on the Web site listed at the end of this entry.

Neuroimaging

The anatomic details of abnormalities in the CNS can often be identified by computed axial tomography—the CT scan. The CT scans offer good clarity for acute blood and bone pathology, but imaging of the lower brain (i.e., the brainstem region) and acute ischemia cannot be seen. The CT scan has the advantage of being rapid and relatively inexpensive compared

with magnetic resonance imaging (MRI). MRI provides much more detailed information regarding the state of soft tissue and is the preferred imaging modality for cord lesions and acute stroke. Vascular anatomy, based on blood flow, can also be viewed. Vascular anatomy can also be assessed with Doppler technology. The WHO atlas provides details as to the availability of these imaging modalities in various regions of the world.

Electrophysiologic Tests

Brain function and the potential propensity for seizure activity as well as subtyping of seizure disorders into various syndromes may be facilitated with the use of electroencephalography (EEG). An EEG involves placement of recording electrodes on the scalp with amplifiers used to record brain activity in the cortical neurons accessible to this surface. This is a relatively cheap and noninvasive test, but requires substantial expertise for interpretation, especially among children. Inexperienced physicians have a propensity to report false-positive abnormalities in normal records.

The function of peripheral nerves, the spinal cord, and brain stem function can be assessed by various methods in which stimuli (electrical, visual, or auditory) are provided peripherally and a more central/proximal response to the stimuli is recorded. Examples include visual evoked responses, bilateral auditory evoked responses, and somatosensory evoked potentials. The peripheral nerve function is also assessed through nerve conduction velocities measured along the peripheral neuroaxis.

Muscle abnormalities may be investigated by electromyography, which involves the insertion of very fine needles into muscle and recording the background activity and activity changes that occur with activation. The clinical and epidemiological values of essentially all neurophysiologic studies heavily depend on the expertise of the technician and the interpreting clinician.

Cerebrospinal Fluid and Other Laboratory Analysis

The presence of the blood-brain and blood-nerve barriers limit the capacity of routine serum and blood work to provide information on the state of the nervous system. Cerebrospinal fluid (CSF) can be obtained, however, through a relatively noninvasive and generally safe procedure—the lumbar puncture (sometimes referred to as a spinal tap). Theoretically, most physicians should be able to perform a spinal tap but patient reluctance to undergo the procedure and physician inexperience can result in limited CSF data or delayed diagnoses. Basic laboratory tests, such as a Gram stain or cell count are almost universally available in laboratories, even in developing regions. More sophisticated tests, for example, the 14-3-3 protein used in the diagnosis of CJD, usually have to be sent to central research or university laboratories. The stability of assessments on CSF samples that have been stored and shipped is often unknown.

Complex Diagnostic Criteria

For research purposes, clear diagnostic criteria are necessary and many neurologic disorders use fairly complex clinical diagnostic criteria. The application of such complex criteria will depend on the expertise of the assessor as well as availability of diagnostic tests. For population-based studies, screening instruments are often used to identify potential cases that are then referred on to more advanced expertise for second-level assessment. In such cases, the screening tool should be designed to provide a low false negative, high false positive rate (i.e., more sensitive, less specific). The validity of such screens, and of the associated diagnostic criteria against the "gold standard," ideally should be available. Screens and assessments that include patient reports of symptoms (i.e., most screens and assessments) should be validated in the culture and language in use.

Other Outcome Measures

Mortality

For some neurologic disorders, simply "counting" bodies is probably sufficient for understanding the incidence of the disease. Rapidly progressive, fatal conditions such as CJD exemplify this. But if progression is very rapid, and/or autopsy data are limited, the diagnosis of rapidly progressive, fatal neurologic conditions is likely to be missed, especially in regions with limited resources and expertise.

Functional Status

For chronic and disabling conditions, such as MS, mortality data will provide little information about the epidemiology of the disease. The measures of

disability such as the Barthel Index or Kurtkze's Disability Scale are more useful for this purpose. An array of neurological scales are available for use within both pediatric and adult populations and have been detailed in Herndon's textbook of neurological rating scales.

Health-Related Quality of Life (HRQOL)

The trends in more patient-oriented outcome measures (i.e., outcomes that matter to the patient!) have led to the development of various instruments to assess HRQOL based on patient-reported data of symptoms and impact of symptoms on their lives. Generic measures applicable to most conditions include the SF-36® (short form 36), a 36-item questionnaire. Many neurologic disorders also have a disease-specific instrument already developed and validated in many populations. For example, the QOLIE-89 (quality of life in epilepsy) includes the SF-36® plus 53 additional, epilepsy-specific items. Although HRQOL measures are not often used as the primary outcome in clinical trials, these measures provide important additional data regarding disease severity, treatment effects, and prognosis among people with chronic neurologic disorders and should be of substantial interest to those working in neuroepidemiology.

—Gretchen L. Birbeck

See also Psychiatric Epidemiology; Quality of Life, Quantification of; Screening; SF-36® Health Survey

Further Readings

Batchelor, T., & Cudkowicz, M. (Eds.). (2001). *Principles of neuroepidemiology*. Woburn, MA: Butterworth-Heinemann.

Bergen, D. C. (1998). Preventable neurological diseases worldwide. *Neuroepidemiology*, *17*(2), 67–73.

Herndon, R. (2006). *Handbook of neurologic rating scales* (2nd ed.). New York: Demos Medical.

Nelson, K. B. (2002). The epidemiology of cerebral palsy in term infants. *Mental Retardation and Developmental Disabilities Research Reviews*, *8*(3), 146–150.

Nelson, L. M., Tanner, C. M., Van Den Eeden, S., & McGuire, V. M. (Eds.). (2003). *Neuroepidemiology: From principles to practice*. Oxford, UK: Oxford University Press.

World Health Organization. (2004). *Atlas: Country resources for neurological disorders 2004*. Geneva: Author.

Wiederholt, W. C. (1999). Neuroepidemiologic research initiatives on Guam: Past and present. *Neuroepidemiology*, *18*(6), 279–291.

Web Sites

Details on common diagnostic tests used in more developed regions can be accessed at http://www.merck.com/mrkshared/mmanual/section14/chapter165/165d.jsp.

Newborn Screening Programs

Newborn screening for disease is a highly effective public health effort to prevent the consequences of certain diseases in affected newborns. Through testing of blood samples from and administration of hearing tests to newborn infants, targeted diseases are detected very early, often before manifestations of diseases are evident, enabling rapid initiation of treatment of these diseases. This entry summarizes the mechanism of screening, the diseases screened, and the treatment of some of these diseases and highlights the potential of newborn screening for identification and control of other health problems.

Newborn screening comprises a system through which a laboratory, public or private, processes a newborn blood specimen to detect the possible presence of a disease in the infant. The newborn screen blood sample is usually obtained by a health care provider, typically a hospital nurse. The blood sample is placed on a special newborn screening card, the blood is dried, and the card is then transported to the testing laboratory. If the test is normal, the results are sent to the infant's health care provider and the testing is complete; if the test is abnormal, newborn screening programs follow-up measures ensure that the infant with a positive result enters into treatment for the disease. These steps include notification of the infant's physician and family of the positive screening result; obtaining a specimen for a second screening test; and, if the second screen is positive, a visit to a clinical specialist for diagnostic testing (the laboratory screening test typically detects an elevation in a substance that can occasionally be temporary and not indicative of actual disease). Finally, if the diagnostic test indicates the presence of a disease, the infant undergoes the therapeutic treatment recommended by existing clinical standards for the specific disease typically by a specialist trained to care for the specific disorder.

All 50 states and the territories perform screening tests of newborn blood specimens to detect diseases for which a treatment prevents the medical complications of untreated disease. With improvements in testing

technology, most newborn screening programs are now expanding the number of disorders for which screening is done. This entry discusses the development of newborn screening, the current expansion of the programs, and the potential for future newborn screening.

Historical Background

Following the rediscovery of Mendel's genetic principles at the beginning of the 20th century, medical practitioners began to recognize that many human diseases are genetic. Throughout the early part of the century, an understanding of the principles of genetics advanced. Subsequent advances included the identification of deoxyribonucleic acid (DNA) as the genetic material; the delineation of the molecular structure of DNA by Watson and Crick in 1953; and, at the end of the century, completion of the draft sequence of the human genome.

Hand in hand with these advances in genetics were advances in biochemistry. It became clear that while DNA contained the information for life, biochemical pathways and molecules produced from genetic material were the engine of this information. If there is an alteration in genetic information, this typically results in a biochemical disturbance.

In 1902, Sir Archibald Garrod noticed that patients with a disease he called alkaptonuria excreted excessive amounts of alkapton (a urinary chemical that turned the urine to a dark color that was later identified as homogentisic acid) into the urine. Based on the pattern of inheritance Garrod recognized as Mendelian (in this case, recessive), he correctly concluded that this disorder represented a genetic alteration in metabolism. The term *inborn errors of metabolism* eventually was coined to describe collectively the diseases of patients with genetic defects in biochemical pathways.

Enzymes perform most of the biochemical reactions in cells. They are proteins whose function is to perform a chemical reaction in which one chemical substance is converted into another. The original chemical is called a substrate, and the end chemical is the product. Research over the century has identified thousands of chemical reactions, and these reactions are mediated by thousands of enzymes. If an enzyme does not function, then the reaction does not occur and substrates for the reaction accumulate and products become deficient. As all enzymes are the products of genes, the presence of defective enzymes usually means an alteration in the genetic information present in the patient.

Following Garrod's initial description, additional inborn errors were identified based on analyses of patient samples. Typically, the substrate for a defective enzymatic reaction accumulates in tissues and blood and is excreted into urine and/or stool where the elevations can be detected by testing. Phenylketonuria (PKU) was recognized as an inborn error in 1934 and determined to be due to elevations of the amino acid phenylalanine due to defective function of the enzyme phenylalanine hydroxylase. Analysis of institutionalized, mentally retarded patients revealed that many of them had PKU. In the 1950s, Dr. Horst Bickel and associates showed that blood levels of phenylalanine could be reduced in PKU patients by a diet low in protein (and, thus, phenylalanine). With reduction of blood phenylalanine levels, many medical symptoms improved. These observations set the stage for newborn screening.

In the early 1960s, motivated in part by a family history of mental retardation, in a son, and phenylketonuria, in a niece, Dr. Robert Guthrie described a method for the detection of elevated blood phenylalanine in blood samples obtained from newborns. He deduced that placement of affected infants on infant formula low in protein would reduce their blood levels of phenylalanine and prevent development of mental retardation. The problem was to identify infants affected with PKU before the onset of symptoms. Guthrie approached public health officials, and policies to screen all newborn infants for PKU were implemented. This effort rapidly spread throughout the United States, and soon all states were screening infants for PKU. Dr. Guthrie's hypothesis regarding early treatment of PKU by a phenylalanine (protein) restricted diet was correct and highly successful in preventing the devastating complications of untreated disease.

Building on the PKU experience, it was soon recognized that other inborn errors could be detected by assays of accumulated compounds or of enzymes in newborn blood and that many of these additional diseases had effective treatments. From the 1960s to the present, the number of disorders identified through newborn screening programs has slowly increased.

Current Screening Procedures

Typically, a newborn screen is obtained from an infant at approximately 24 to 48 hr of age. The heel of the infant is warmed, and a lancet is used to puncture the skin and obtain capillary blood. The drops of blood are

placed onto a special filter paper card, and the blood spot is dried. Demographic information is recorded on the card, and it is sent to the screening laboratory.

Once at the laboratory, small circular punches of the dried blood are obtained and processed for analysis. The sample may be tested for chemicals that accumulate due to an enzymatic defect, the activity of a specific enzyme can be assayed, or a protein can be analyzed by biochemical means.

Full testing of the sample usually takes about 2 to 3 days. The results are then compared with laboratory-generated normal values and the result reported to the infant's physician. Typically, the result of the newborn screen is complete when the infant is 7 to 10 days of age. This rapid analysis is necessary as some of the disorders for which screening is done can cause critical illness in the first 2 weeks of life. If there is an abnormality, the physician may need to repeat the newborn screen or move to more definitive testing.

Disorders Detected in Newborn Screening

Within the past decade, the application of tandem mass spectrometry to newborn screening has enabled significant expansion of the number of disorders that can be detected. This has led organizations such as the American College of Medical Genetics and the March of Dimes to propose a panel of disorders in an attempt to expand and unify newborn screening programs in all states. The recommended panel includes 29 disorders, including congenital hearing loss. These 29 disorders are thought to represent disorders for which a favorable treatment exists. They can be broadly grouped into amino acid disorders, organic acid disorders, fatty acid oxidation defects, hormonal disorders, hemoglobinopathies, vitamin disorders, carbohydrate disorders, pulmonary disorders, and congenital hearing loss. Tandem mass spectrometry does enable testing for other disorders for which effective treatments do not yet exist and leaves the decision for testing of these additional disorders to individual states.

Amino Acid Disorders

These disorders include some of the first to be part of routine newborn screening programs. PKU is due to a functional defect in the enzyme phenylalanine hydroxylase. As a result, phenylalanine, which derives from dietary protein, accumulates to high levels and, with time, can cause neurologic damage and ultimately mental retardation. Treatment with a low-protein/phenylalanine diet prevents development of these symptoms.

Maple syrup urine disease is due to a functional defect in the enzyme branched chain α-ketoacid dehydrogenase. Accumulation of the branched chain amino acids leucine, isoleucine, and valine and their respective ketoacids is rapidly damaging to the nervous system. Rapid treatment with a low-protein diet reduces these levels and prevents neurologic damage.

Homocystinuria is due to defective function of the enzyme cystathionine-β-synthase. Elevation of methionine and homocysteine occur and, with time, can damage the eye and blood vessels. A low-protein/methionine diet reduces blood levels and the risk of these complications.

Tyrosinemia Type I is due to dysfunction of the enzyme fumarylacetoacetic acid hydrolase. Damage to the liver occurs within 4 to 6 months and can be prevented with medications and a low-tyrosine diet.

Citrullinemia and argininosuccinic acidemia are urea cycle disorders due to defective function of argininosuccinic acid synthase and lyase, respectively. Severe elevations in blood levels of ammonia result and can damage the nervous system. Institution of a low-protein diet helps lower blood ammonia levels and prevent damage.

Organic Acid Disorders

Organic acid disorders comprise the group providing the largest increase in the number of diseases included in expanded newborn screening programs. Included in the recommended 29 disorders are the following: isovaleric acidemia, glutaric acidemia Type I, 3-hydroxy-3-methylglutaric acidemia, multiple carboxylase deficiency, methymalonic acidemia due to mutase deficiency, cblA and cblB deficiency, 3-methylcrotonyl-CoA carboxylase deficiency, propionic acidemia, and β-ketothiolase deficiency. As a group, they typically present with severe acidosis and neurologic dysfunction. Treatment is effected through institution of a low-protein diet and disease-specific medications.

Fatty Acid Oxidation Defects

The fatty acid oxidation defects are due to defective functioning of enzymes involved in the breakdown of stored fat used for energy production. Typically, they cause symptoms during times of insufficient food

intake, but some also cause liver or heart damage without fasting. There are many enzymes in these metabolic processes, including medium chain acyl-CoA dehydrogenase, very long chain acyl-CoA dehydrogenase, long chain 3-hydroxy acyl-CoA dehydrogenase, trifunctional protein, and others. Treatment varies with the individual disorder but in general includes avoidance of fasting and limitation of fat intake.

Hormonal Disorders

Congenital hypothyroidism is one of the most common disorders detected by newborn screening and was the second disorder (following PKU) to be included on a routine basis in newborn screening programs. Insufficient thyroid hormone production by the thyroid gland, whether due to failure of formation of the gland or due to an enzyme defect in the synthesis of hormone, results in mental retardation and poor growth. Treatment with replacement of thyroid hormone is effective in preventing these symptoms.

Congenital adrenal hyperplasia, due to adrenal 21-hydroxylase deficiency, can cause loss of body salts and masculinization of female genitalia. The loss of body salt can be life threatening. Treatment by hormone replacement can reverse the loss of body salt. Treatment of masculinization of the female genitalia may require surgery.

Hemoglobinopathies

Hemoglobin is the oxygen-transporting protein present in red blood cells. Genetic alterations in the structure of hemoglobin may alter its function. One of the most common of these defects, that as a group are called hemoglobinopathies, is sickle-cell anemia. This disorder is common in populations of individuals of African American ancestry and causes anemia and a predisposition to bacterial infection that can be prevented with antibiotics. The newborn screen will also detect other clinically significant hemoglobinopathies such as thalassemia and hemoglobin E.

Vitamin Disorders

Biotinidase is an enzyme involved in preserving the body's levels of the important vitamin biotin. When biotinidase function is defective, the body gradually becomes deficient in biotin, and this deficiency disrupts function of biotin-requiring enzymes. The symptoms include skin rash, hair loss, seizures, and neurologic damage. Supplementation with biotin prevents these symptoms.

Carbohydrate Disorders

Classic galactosemia is due to a defect in the function of the enzyme galactose-1-phosphate uridyltransferase. Galactose is a sugar found in a variety of foods, especially in dairy foods containing the disaccharide lactose. Defective functioning of galactose-1-phosphate uridyltransferase causes accumulation of galactose, which can damage the liver and the eyes. Restriction of dietary lactose reduces blood levels and prevents this damage.

Pulmonary Disorders

Cystic fibrosis is one of the most common genetic diseases in populations of European ancestry. It is due to a defective function of the cystic fibrosis membrane transconductance regulator. Abnormal movement of water and salts within internal body ducts results in abnormally thick mucous. This thick mucous plugs the ducts of the respiratory, reproduction, and gastrointestinal tracts. This causes damage to the pancreas and the lungs. Identification of affected infants allows early treatment for nutritional and growth problems.

Hearing Loss

Congenital hearing loss is very common, and identification of infants enables interventions to improve speech development. There are many genetic and nongenetic causes of hearing loss in the newborn. Early treatment with speech therapy helps hearing impaired children improve communication skills.

The Future of Newborn Screening

The 29 disorders were recommended for screening because each has some therapeutic intervention that helps prevent development of medical complications. There are, however, many other disorders that could be detected in the newborn screen sample. It is highly likely that testing will be expanded beyond 29 disorders in the future.

The newborn blood sample can contain antibodies that indicate exposure to infectious diseases such as toxoplasmosis, cytomegalovirus, and human immunodeficiency virus. While not genetic diseases, these

infectious diseases have important public health considerations that make early identification important. The blood spot may contain substances such as methamphetamine, heroin, and cocaine that would indicate the use of these substances by the mother shortly before delivery. Importantly, the blood sample also contains DNA, the genetic material of the human body. Tests for genetic diseases by analysis of DNA continue to expand at an exponential pace. Potential diseases for testing include certain cancers, Huntington disease, fragile X syndrome, and many other inherited diseases.

Such testing does, however, have ethical and legal risks. Because of public health and legal considerations beyond the medical effects on the infant, such testing would likely require parental informed consent. Additionally, psychological harm may result from knowing in childhood that one is going to develop an untreatable disease in the future, and identification of individuals with a genetic disease may result in discrimination in obtaining health insurance. These are important issues that will need to be resolved by future debate and policy but highlight the testing potential offered by the newborn screen.

—Randall A. Heidenreich

See also Genetic Counseling; Genetic Disorders; Mutation; Screening

Further Readings

Guthrie, R. (1992). The origin of newborn screening. *Screening*, *1*, 5–15.

Marsden, D., Larson, C., & Levy, H. L. (2006). Newborn screening for metabolic disorders. *Journal of Pediatrics*, *148*, 577–584.

Therrell, B. L., Panny, S. R., Davidson, A., Eckman, J., Hannon, W. H., Henson, M. A., et al. (1992). U.S. newborn screening guidelines: Statement of the Council of Regional Networks for Genetic Services. *Screening*, *1*, 135–147.

Nightingale, Florence (1820–1910)

In the 19th century, Florence Nightingale played a key role in the areas of public health policy, medical statistics, hospital design, and patient care. Stepping over gender stereotypes, she reached beyond the typical role of a nurse by studying statistics, hospital management, philosophy, and sanitation. She was a fan of Edwin Chadwick, a sanitary reformer who was influential in passing England's Public Health Act of 1848. Chadwick's premise was that filth, poor ventilation, and unclean water were the causes of disease development. Each of these factors was present in Scutari (the Greek name for Istanbul, Turkey) during the Crimean War. Nightingale's statistical abilities were apparent in her work there; according to historical documents, approximately a year after her arrival at Scutari in 1854, survival rates improved significantly.

Nightingale was never mentioned in a voluminous post–Crimean War report written by the surgeons who served in the war. Historians attribute their refusal to grant her credit for her accomplishments to their resentment of her obtained power. They argued that her statement that providing hygiene, clean air, and nourishment had beneficial effects on mortality rates was inaccurate; instead, they claimed decreased mortality rates were a consequence of lower nurse-patient ratio after she arrived with more nurses. In spite of her critics, her successes were acknowledged by the creation of the Nightingale Fund as a "thank you" offering from the people of England. The first nonsectarian nursing school was established as a result of the fund. Furthermore, Nightingale is credited for using statistical graphs effectively to make changes in hospitals. She became a Fellow of the Royal Statistical Society in 1858 and an honorary member of the American Statistical Association in 1874. Nightingale's contributions to health statistics and epidemiology are clearly outlined in Table 1.

Nightingale's words reflect the challenge for health care researchers worldwide:

> You can see the power of careful, accurate, statistical information from the way that I used them in my pleas to Government to improve the conditions of ordinary soldiers and of ordinary people. I collected my figures with a purpose in mind, with the idea that they could be used to argue for change. Of what use are statistics if we do not know what to make of them? What we wanted at that time was not so much an accumulation of facts, as to teach the men who are to govern the country the use of statistical facts. (Maindonald & Richardson, 2004, unpaginated)

—Anne P. Odell

See also Epidemiology, History of; Public Health, History of; War

Table 1 Modern Hospital Epidemiology Versus 19th Century Hospital Epidemiology

Modern Hospital Epidemiology				
Observational Studies	*Analytical Studies*	*Interventional Epidemiology*	*Prevalence Studies*	*Results*
Frequency of infections in relation to person, time, and place	Cohort studies trials	Random clinical trials	Snapshot in time of infections clinically active	Prevention of catheter-related bloodstream infection
Hypothesis about source, reservoir, mode, and route of transmission	Case-control trials	Evidence-based trials	Efficacy of infection control practices—measured by repeated prevalence studies	Prevention of pneumonia
	Quantifies attributes, morbidity, mortality, and costs			Prevention of surgical site infections
	For future projections and prevention			Control of multiresistant microorganisms

Note: This chart demonstrates the influence of Nightingale's work on modern hospital statistics from the National Center for Nursing Research (1992). The highlighted areas are stemmed from Nightingale's work.

Further Readings

Gill, C. J., & Gill, G. C. (2005). Nightingale in Scutari: Her legacy reexamined. *Clinical Infectious Diseases*, *40*(12), 1799–1805.

Maindonald, J., & Richardson, A. M. (2004). This passionate study: A dialogue with Florence Nightingale. [Electronic version]. *Journal of Statistics Education*, *12*(1). Retrieved July 19, 2006, from http://www.amstat.org/publications/jse/v12n1/maindonald.html.

Nightingale, F. (with contributions by Mundinger, M. O.), et al. (1999). *Florence Nightingale : Measuring hospital care outcomes*. (Excerpts from the books Notes on matters affecting the health, efficiency, and hospital administration of the British Army founded chiefly on the experience of the late war, and Notes on hospitals.) Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations.

Nonparametric Statistics

Some of the most popular statistical inferential techniques in epidemiological research are those that focus on specific parameters of the population such as the mean and variance. These *parametric statistics* share a number of common assumptions:

- There is independence of observations except when data are paired.
- The set of observations for the outcome (i.e., dependent) variable of interest has been randomly drawn from a normally distributed or bell-shaped population of values.
- The dependent variable is measured on at least an interval-level scale of measurement (i.e., it is rank ordered and has equidistant numbers that share similar meaning).
- The data are drawn from populations having equal variances or spread of scores.
- Hypotheses are formulated about parameters in the population, especially the mean.
- Additional requirements include nominal- or interval-level independent variables, homoscedasticity, and equal cell sizes of at least 30 observations per group.

Examples of commonly used parametric statistical tests include the independent *t* test, the Pearson product-moment correlation, and analysis of variance (ANOVA). These techniques have frequently been used even when the data being analyzed do not adequately meet the assumptions of the given parametric test.

While some parametric tests (e.g., the *t* test) are *robust* in that they can withstand some violations of their assumptions, other tests (e.g., ANCOVA and Repeated Measures ANOVA) are not so flexible. It is extremely important, therefore, that the researchers carefully examine the extent to which their data meet the assumptions of the tests that they are considering. When those assumptions are not met, one option is to use nonparametric statistics instead.

Characteristics of Nonparametric Statistics

There are alternative statistical tests that make fewer assumptions concerning the data being examined. These techniques have been called *distribution free-er* (because many are not entirely free of distributional assumptions) or *nonparametric tests.* Common assumptions for nonparametric tests include the following:

- Like parametric tests, nonparametric tests assume independence of randomly selected observations except when the data are paired.
- Unlike parametric tests, the distribution of values for the dependent variable is not limited to the bell-shaped normal distribution; skewed and unusual distributions are easily accommodated with nonparametric tests.
- When comparing two or more groups using rank tests, the distribution of values within each group should have similar shapes except for their central tendency (e.g., medians).
- There are no restrictions as to the scale of measurement of the dependent variable. Categorical and rank-ordered (ordinal) outcome variables are acceptable.
- The major focus of analysis in nonparametric statistics is on either the rank ordering or frequencies of data; hypotheses, therefore, are most often posed regarding ranks, medians, or frequencies of data.
- The sample sizes are often smaller (e.g., $n \leq 20$).

Types of Nonparametric Tests

There are a wide variety of nonparametric statistical tests that are available in user-friendly computer packages for use in epidemiology. Table 1 summarizes the most commonly used nonparametric tests, their purposes, the type of data suitable for their use, and their parametric equivalents, if any. The following is a brief overview of these statistics. For more details on these and other nonparametric tests as well as instructions on how to generate these statistics in various statistical packages, the interested reader is

Table 1 Commonly Used Nonparametric Tests, Their Purpose, Types of Data Required, and Their Parametric Equivalents

Tests	*Nonparametric Test*	*Level of Measurement of Data*		*Parametric Equivalent*
		Independent Variable	*Dependent Variable*	
1. *Goodness-of-fit tests* (To determine if the distribution of a data set is similar to that of a hypothesized target population)	Binomial test	—	Nominal (dichotomous)	None
	Chi-square goodness-of–fit test	—	Nominal	None
	Kolmogorov-Smirnov one-sample test	—	Ordinal, interval, or ratio	One-sample *t* test
	Kolmogorov-Smirnov two-sample test	Nominal (dichotomous)	Ordinal, interval, or ratio	Independent *t* test
2. *Tests for two related samples, pretest-posttest measures for single samples* (To identify differences in paired data, e.g., pre-post data for same group of subjects, or subjects matched according to defined criteria)	McNemar test	Paired nominal (dichotomous) data		None
	Wilcoxon signed ranks test	Paired ordinal, interval, or ratio data		Paired *t* test
3. *Repeated measures for more than two time periods or matched conditions* (To evaluate differences in paired data repeated across more than two time periods or matched conditions)	Cochran's *Q* test	Paired nominal (dichotomous) data		None
	Friedman test	Paired ordinal, interval, or ratio data		Within-subjects repeated measures ANOVA

4. *Tests for differences between two independent groups* (To examine differences between two independent groups)	Fisher exact test	Nominal	Nominal	None
	Chi-square test of independence	Nominal	Nominal	None
	Mann-Whitney *U* test	Nominal	Ordinal, interval, or ratio	Independent *t* test
5. *Tests for differences among more than two independent groups* (To compare more than two independent groups)	Chi-square test for independent samples	Nominal	Nominal	None
	Mantel-Haenszel chi-square test for trends	Nominal	Nominal	None
	Kruskal-Wallis one-way ANOVA by ranks test	Nominal	Ordinal, interval, or ratio	One-way ANOVA
6. *Tests of association* (To examine the degree of association, or correlation, between two variables)	Phi coefficient	Nominal (dichotomous)	Nominal (dichotomous)	None
	Cramér's *V* coefficient	Nominal	Nominal	None
	Point biserial correlation	Ordinal, interval, or ratio	Nominal	None
	Spearman rho rank order correlation	Ordinal, interval, or ratio	Ordinal, interval, or ratio	Pearson *r*

referred to the texts on nonparametric statistics cited at the end of this entry.

Goodness-of-Fit Tests

Goodness-of-fit tests are used when a researcher has obtained a set of data from a sample and wants to know if this set of data is similar to that of a specified target population. For example, we might want to know whether the distribution of smokers and nonsmokers in a given sample is similar to previously published national norms. We might also be interested in comparing rates of specific cancers from one country with that of another. These types of tests are goodness-of-fit tests because they compare the results obtained from a given sample with a prespecified distribution. Three nonparametric goodness-of-fit tests that are frequently used in epidemiology are the binomial test, the chi-square goodness-of-fit test, and the Kolmogorov-Smirnov one- and two-sample tests.

Binomial Tests

The binomial test uses the binomial distribution to determine the probability that a sample of data with dichotomous outcomes (e.g., smokers vs. nonsmokers) could have come from a population with a prespecified binomial distribution. This test is especially useful when sample sizes are small. All that is required is (1) a dichotomous outcome variable whose values are frequencies, not scores, and (2) knowledge about the expected proportions in the population. There is no parametric alternative to this test.

Chi-Square Goodness-of-Fit Tests

Not all nominal-level variables are dichotomous. For example, the researcher may be interested in comparing the frequencies of different types of cancer in a given sample with what would have been expected given what is known or hypothesized about a target population. The chi-square goodness-of-fit test allows for comparison of actual frequencies of categorical data with those of a population of interest. This very flexible test of frequencies has few assumptions, and because it evaluates nominal-level data, there is no parametric alternative.

Kolmogorov-Smirnov One- and Two-Sample Tests

The Kolmogorov-Smirnov (K-S) one- and two-sample tests allow the researcher to examine whether a set of continuous outcome data (i.e., at least ordinal level of measurement) is similar to a hypothesized set of continuous data with a prespecified distribution (e.g., normal, Poisson, or uniform distributions). To do this, the K-S statistic compares the cumulative distribution of a sample variable with that which would have been expected to occur had the sample been obtained from a theoretical parent distribution (e.g., a Poisson distribution). The one-sample K-S test focuses on a continuous outcome variable (e.g., time to recovery from chemotherapy) and the two-sample test allows for a comparison of the continuous variable between two independent groups (e.g., gender). Technically, the parametric counterparts to the one- and two-sample K-S tests are the one- and two-sample *t* tests. However, the K-S statistics have the advantage of comparing the cumulative distribution of the sample data with that of a prespecified distribution, not just the measure of central tendency.

Tests for Two Related Samples

In epidemiology, the researcher is often interested in evaluating data that have been collected from a single sample that have been paired through using subjects as their own controls (e.g., pretest, posttest data) or as matched pairs (e.g., subjects matched on age and then randomly assigned to an intervention/control group). The two commonly used nonparametric tests for two related samples are the McNemar and Wilcoxon signed ranks tests.

McNemar Tests

The McNemar test is useful when the researcher has a pretest-posttest design in which subjects are used as their own controls, and the dependent variable is dichotomous. This versatile statistic can be used to determine whether the distribution of a dichotomous outcome variable (e.g., willingness to undergo a colonoscopy: yes or no) changes following an intervention (a colon cancer prevention program). Because the McNemar test is used with dichotomous data, there is no parametric counterpart to this test.

Wilcoxon Signed Ranks Tests

The McNemar test can only determine whether or not a change has occurred from one time period to another; it cannot be used to evaluate the extent of change in a variable. The Wilcoxon signed ranks test is a commonly used statistical test that enables the

researcher to assess the extent of change in a continuous variable (e.g., self-efficacy regarding ability to undergo a colonoscopy) across two time periods (e.g., prior to and following a colon cancer prevention program).

The assumptions of the Wilcoxon signed ranks test are fairly liberal. The major requirement is that the continuous data being examined be paired observations that are at least ordinal level of measurement both within and between pairs of observations. The parametric alternative to the Wilcoxon signed ranks test is the paired *t* test.

Repeated Measures for More Than Two Time Periods

In epidemiological research, we are often interested in repeated observations across more than two time periods, for example, preintervention, postintervention, and follow-up. Two nonparametric tests that can be used to evaluate differences in paired data repeated across more than two time periods or matched conditions are the Cochran's *Q* test and the Friedman test.

Cochran's Q Tests

Cochran's *Q* test extends the McNemar test to examine change in a dichotomous variable across more than two observation periods. It is especially appropriate when subjects are used as their own controls and the dichotomous outcome variable (e.g., smoking cessation, yes or no) is measured across multiple time periods or under several types of conditions. Like the McNemar test, Cochran's *Q* test can only detect whether or not a change has occurred across time, not the extent of that change. It also focuses on change in a single group across time (e.g., the intervention group). Group × time interactions cannot be evaluated using this statistic. There is no parametric alternative to this test.

Friedman Tests

The Friedman test is the preferred nonparametric statistic when the outcome data being evaluated across multiple time periods are continuous (e.g., body weight). This test can also be used to evaluate differences among matched sets of subjects who have been randomly assigned to one of three or more conditions. The Friedman test examines the ranks of the data generated during each time period or condition to determine whether the variables share the same underlying continuous distribution and median. When significant differences are found with this statistic, post hoc tests (e.g., the Wilcoxon signed ranks test) are needed to determine where the specific time differences lie. Because this is a within-subjects test, the Friedman test is not useful for evaluating between-group differences. The parametric equivalent to this test is the within-subjects repeated-measures ANOVA without a comparison group.

Tests for Differences Between Two Independent Groups

In epidemiology, we are often interested in comparing outcomes obtained among groups that are independent of one another, such as an intervention and control group, smokers and nonsmokers, persons with or without cancer. Three nonparametric tests that are available when the independent variable is nominal level of measurement with two mutually exclusive levels are the Fisher exact test, the chi-square test of independence, and the Mann-Whitney *U* test.

Fisher's Exact Tests

Fisher's exact test is used to evaluate the degree of association between a dichotomous independent and dependent variable (e.g., gender and smoking status). It is especially useful when sample sizes are small (e.g., $n \leq 15$). Because the Fisher exact test deals exclusively with variables that are measured at the nominal level, there is no parametric equivalent to this test. When the sample size is sufficiently large, the chi-square test of independence is typically used.

Chi-Square Test of Independence

The chi-square test of independence (χ^2) is one of the most commonly used nonparametric statistics in epidemiology. It is an easily understood statistic that is used to evaluate the association between two categorical variables that have two or more levels. Because the generated chi-square statistic is an overall test of association, additional tests (e.g., the phi and Cramér's *V* statistics) are used to evaluate the strength of the relationship between the two categorical variables. Since both the independent and the dependent variables are nominal level of measurement, there is no parametric equivalent to this chi-square test.

Mann-Whitney U Tests

The Mann-Whitney *U* test is useful when the independent variable is dichotomous, and the continuous

dependent variable is measured on at least an ordinal scale. Like its parametric counterpart, the independent *t* test, the Mann-Whitney test compares measures of central tendency between two independent groups (e.g., perceived quality of life in smokers and nonsmokers). Unlike the *t* test, the Mann-Whitney test uses medians for comparison, not means. The Mann-Whitney test is almost as powerful as the *t* test, especially when the sample size is small and the outcome data being analyzed are not normally distributed.

Tests for Differences Among More Than Two Independent Groups

In epidemiology, there are often more than two independent groups that are being compared. Subjects may be assigned randomly to more than two intervention conditions, or three or more patient groups having different diagnoses (e.g., type of cancer) may be compared with regard to an outcome variable (e.g., level of fatigue). Three nonparametric tests that may be used given these conditions will be briefly examined below: the chi-square test for *k* independent samples, the Mantel-Haenszel chi-square test, and the Kruskal-Wallis one-way ANOVA by ranks test.

Chi-Square Tests for k Independent Samples

The chi-square test for *k* independent samples (χ^2) is an extension of the chi-square test for two independent samples discussed above except that, with this chi-square statistic, both the independent and the dependent variables are categorical with more than two levels (e.g., the relationship between five levels of racial/ethnic status and four stages of cancer among women with breast cancer). Similar to the previous chi-square statistic, the data consist of frequencies, not scores, and there are no repeated observations or multiple response categories. Once a significant association is determined, the Cramér's *V* statistic is used to assess the strength of the relationship between the two categorical variables. Because it is insensitive to order, this test is not the one of choice when either of the categorical variables has ordered levels. There is no parametric alternative to this test.

Mantel-Haenszel Chi-Square Tests for Trends

It often happens in epidemiological research that the categorical variables being examined have a natural order (e.g., stages of disease, where 1 is the least serious and 4 the most serious) yet are not sufficiently ordinal to be considered continuous variables. Unlike the previous chi-square tests, the Mantel-Haenszel chi-square test for trends takes order into account. It has been used, for example, in case control studies in which there are stratified 2×2 tables, and the researcher is interested in comparing the likelihood of the occurrence of an event (e.g., contracting cancer) given two groups that have been exposed or not exposed to a risk factor (e.g., cigarette smoking) and who have been matched or paired with regard to certain characteristics (e.g., gender or age). Because the variables being examined with this test statistic are categorical, there is no parametric equivalent to this test.

Kruskal-Wallis One-Way ANOVA by Ranks Tests

The Kruskal-Wallis (K-W) one-way ANOVA by ranks is a test that can be used to determine whether *k* independent samples (e.g., stage of cancer) are similar to each other with regard to a continuous outcome variable (e.g., level of fatigue). The K-W test is an extension of the two-sample Mann-Whitney test and is used when the independent variable is categorical with more than two levels and the dependent variable is continuous. When a significant finding is obtained, post hoc tests (e.g., the Mann-Whitney *U* test) are used to determine group differences. The parametric equivalent to the K-W test is the one-way ANOVA. While the ANOVA is more powerful than the K-W test when the assumptions of the ANOVA are met, the K-W test is reported to be more powerful than the ANOVA when the data are skewed or when there are unequal variances.

Tests of Association Between Variables

It frequently occurs in epidemiological research that we are interested in measuring the degree of association or correlation between two variables. Depending on the level of measurement of the two variables being examined, there are a number of nonparametric tests that can provide information about the extent of their relationship. The four commonly used measures of association in epidemiology are the phi and Cramér's *V* coefficients for nominal-level variables, the point biserial correlation for examining the relationship between a dichotomous and a continuous

variable, and the Spearman rho rank order correlation coefficient for two continuous variables.

Phi and Cramér's V Coefficients

The phi and Cramér's *V* coefficients are used to evaluate the strength of relationship between two nominal-level variables when the chi-square statistic has been found to be significant. The phi is used when the two nominal-level variables are dichotomous; the Cramér's *V* coefficient is used with nominal-level variables with more than two levels.

Because both statistics take into account sample size, the researcher is able to compare strengths of association across studies. Both coefficients typically range in values between 0 and 1.00 with higher values indicating greater strength of association. Because both coefficients are calculated from the chi-square statistic, the requirements for these coefficients are similar to those for the chi-square statistic. If the contingency table being examined is 2×2 and the categories are coded "0" and "1," both coefficients have the same value and are similar to the absolute value of a Pearson correlation. If that is not the case, there is no parametric alternative to these useful tests.

Point Biserial Correlation

Sometimes the researcher is interested in assessing the strength of relationship between an independent variable that is continuous (e.g., number of cigarettes smoked) and a dependent variable that is dichotomous (disease state: lung cancer, no lung cancer). The point biserial correlation is a special case of the parametric Pearson product-moment correlation. It is considered to be nonparametric because one of the variables being assessed is dichotomous. Like the Pearson *r*, the point biserial correlation coefficient can range between -1.0 and $+1.0$ with higher absolute values indicating a greater strength of relationship between the two variables.

Spearman Rank-Order Correlation

The Spearman rank-order correlation coefficient (also known as *Spearman's rho* or r_s) is one of the best-known and most frequently used nonparametric statistics. It is used to examine the relationship between two continuous variables (e.g., age and level of depression). Its parametric alternative is the Pearson product-moment correlation coefficient.

Like the point biserial correlation, Spearman's rho is a special case of the Pearson *r* but is based on the ranking of observations, not their actual values. This test statistic can range in value between -1.0 and $+1.0$ with higher absolute values indicating a stronger relationship. The squared values of Spearman's rho offer a reasonable estimate of the strength of the relationship between the two continuous variables of interest.

Conclusion

When making a decision as to whether to use a parametric or nonparametric test, the researcher needs to be aware of the assumptions underlying each test being considered and assess the extent to which the data meet those assumptions. No statistical test is powerful if its assumptions have been seriously violated. Nonparametric statistics are extremely useful analytic tools given their ability to accommodate small sample sizes, categorical and ordinal level data, and unusual sampling distributions. As a result, they offer feasible and potentially powerful solutions to problematic situations.

—Marjorie A. Pett

See also Chi-Square Test; Fisher's Exact Test; Logistic Regression; Measures of Association

Further Readings

Conover, W. J. (2003). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Daniel, W. W. (2000). *Applied nonparametric statistics*. Boston: PWS-Kent.

Green, S. B., & Salkind, N. J. (2007). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Munro, B. H. (2005). *Statistical methods for health care research* (5th ed.). Philadelphia: Lippincott.

Pett, M. A. (1997). *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions*. Thousand Oaks, CA: Sage.

Sheskin, D. J. (2007). *Handbook of parametric and nonparametric statistical procedures* (4th ed.). New York: Chapman & Hall.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Normal Distribution

The normal distribution, also known as Gaussian distribution or "bell-shaped" distribution, is the most

widely used distribution in statistical work for both theoretical and practical reasons. It was first introduced by French mathematician Abraham de Moivre in an article in 1734. The name *Gaussian distribution* refers to the German mathematician and scientist Carl Friedrich Gauss, who rigorously applied the distribution to real-life data. This distribution was used in the analysis of errors of experiments during the early 19th century.

The normal distribution is the cornerstone of most statistical estimation and hypothesis testing procedures, and statistical methods used in epidemiology are no exception. Many important random variables in epidemiology and health sciences, such as distribution of birthweights, blood pressure, or cholesterol levels in the general population, tend to approximately follow a normal distribution. Moreover, the central limit theorem provides a theoretical basis for its wide applicability. Many random variables do not have a normal distribution themselves; however, the sample mean of the variable has an approximate normal distribution when the sample size is large enough, and the sampling distribution of the mean is centered at the population mean. The normal distribution is generally more convenient to work with than any other distribution, particularly in hypothesis testing and confidence interval estimation. For example, in linear and nonlinear regression, the error term is often assumed to follow a normal distribution.

Characterization of the Normal Distribution

The normal distribution is fully defined by two parameters, μ and σ^2, through its probability density function as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < +\infty, \quad [1]$$

where μ is the mean parameter and could take any real value and parameter σ^2 is the variance of the normal distribution (equivalently, σ is standard deviation) with $\sigma > 0$. For example, for diastolic blood pressure, the parameters might be $\mu = 80$ mmHg, $\sigma = 10$ mmHg; for birthweight, they might be $\mu = 120$ oz, $\sigma = 20$ oz. Figure 1 shows the plot of the probability density function for a normal distribution with $\mu = 80$ and $\sigma = 10$.

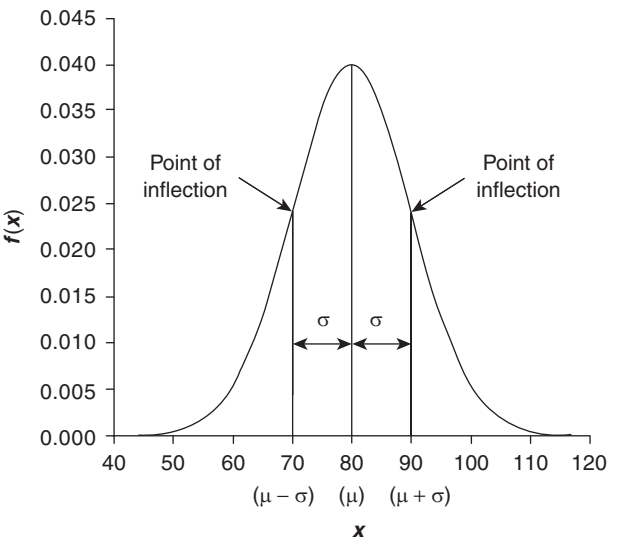


Figure 1 Graphical Illustration of Probability Density Function of Normal Distribution $N(80, 100)$

The density function of the normal distribution resembles a bell-shaped curve, with the mode at μ and the most frequently occurring values around μ. The curve is unimodal and symmetric around μ, and for normal distribution, the mean, median, and mode all equal to μ. The curve has an inflection point on each side of μ at $\mu - \sigma$ and $\mu + \sigma$, respectively. A point of inflection is a point where the slope of the curve changes direction. The distances from μ to points of inflection provide a good visual sense of the magnitude of the parameter σ.

To indicate that a random variable X is normally distributed with mean μ and variable σ^2, we write $X \sim N(\mu, \sigma^2)$. The symbol $\sim$ indicates "is distributed as." The entire shape of the normal distribution is determined by μ and σ^2. The mean μ is a measure of central tendency, while the standard deviation σ is a measure of spread of the distribution. The parameter μ is called the location parameter, and σ^2 is the scale parameter. To see how these parameters affect location and scale, density functions of normal distribution with different means or variances can be compared. For instance, if two normal distributions have different means μ_1 and μ_2 but the same variance σ^2, where $\mu_1 > \mu_2$, then the two density functions will have same shape but the curve with larger mean (μ_1) will be shifted to the right relative to the curve with the smaller mean (μ_2). Figure 2 shows the comparison of

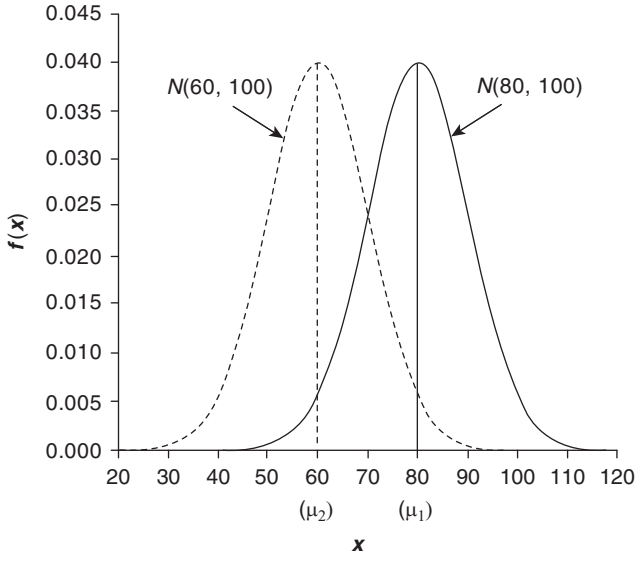


Figure 2 Comparison of Probability Density Function of Normal Distributions $N(80, 100)$ and $N(60, 100)$

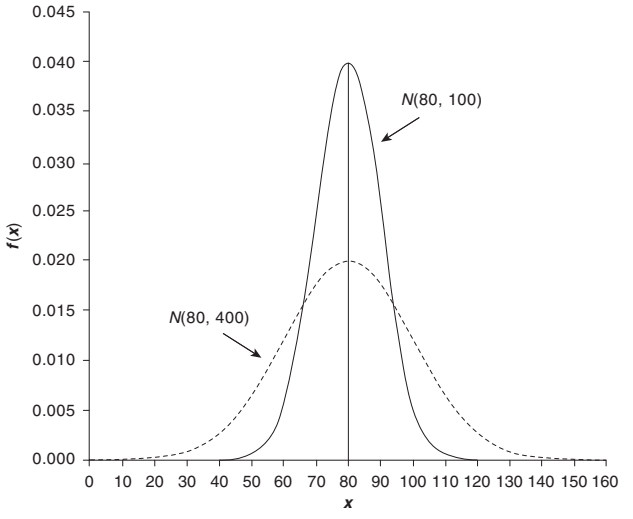


Figure 3 Comparison of Probability Density Function of Normal Distributions $N(80, 100)$ and $N(80, 400)$

density curves of $N(\mu_1 = 80, \ \sigma^2 = 100)$ and $N(\mu_2 = 60, \ \sigma^2 = 100)$. On the other hand, if two normal distributions with the same mean μ and different variance σ_1^2 and σ_2^2, where $\sigma_1^2 < \sigma_2^2$ are compared, then the two density functions will have same mode but the curve with larger variance (σ_2^2) will be more spread out compared with the other curve with the smaller variance (σ_1^2). Variance σ^2 determines the scale of the distribution. Figure 3 shows the comparison of density curves of $N(\mu = 80, \sigma_1^2 = 100)$ and $N(\mu = 80, \ \sigma_2^2 = 400)$.

Standard Normal Distribution

A normal distribution with mean 0 and variable 1 is called a standard normal distribution and denoted as $N(0, 1)$. The probability density function in this case reduces to

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \ -\infty < x < +\infty. \quad [2]$$

The standard normal distribution is symmetric around 0. The importance of the standard normal distribution is that any normal distribution can be transformed into a standard normal distribution, and tables for the standard normal distribution are widely available for calculating cumulative probabilities or p values for a Z test. Any normal distribution may be transformed into a standard normal distribution as given by

$$\text{If } X \sim N(\mu, \sigma^2) \text{ and } Z = (X - \mu)/\sigma, \text{ then } Z \sim N(0, 1). \quad [3]$$

The procedure in Equation 3 is known as standardization of a normal variable.

Cumulative Distribution Function

For a given normal distribution, it is often of interest to calculate the proportion of data falling within a certain range. For example, suppose that in a population, diastolic blood pressure follows a normal distribution with $\mu = 80$ mmHg and $\sigma = 10$ mmHg. People with diastolic blood pressure between 80 and 90 are categorized as prehypertensive; we may want to find the proportion of people in a given population with prehypertension based on diastolic blood pressure. Such proportion is calculated by using the cumulative distribution function (cdf).

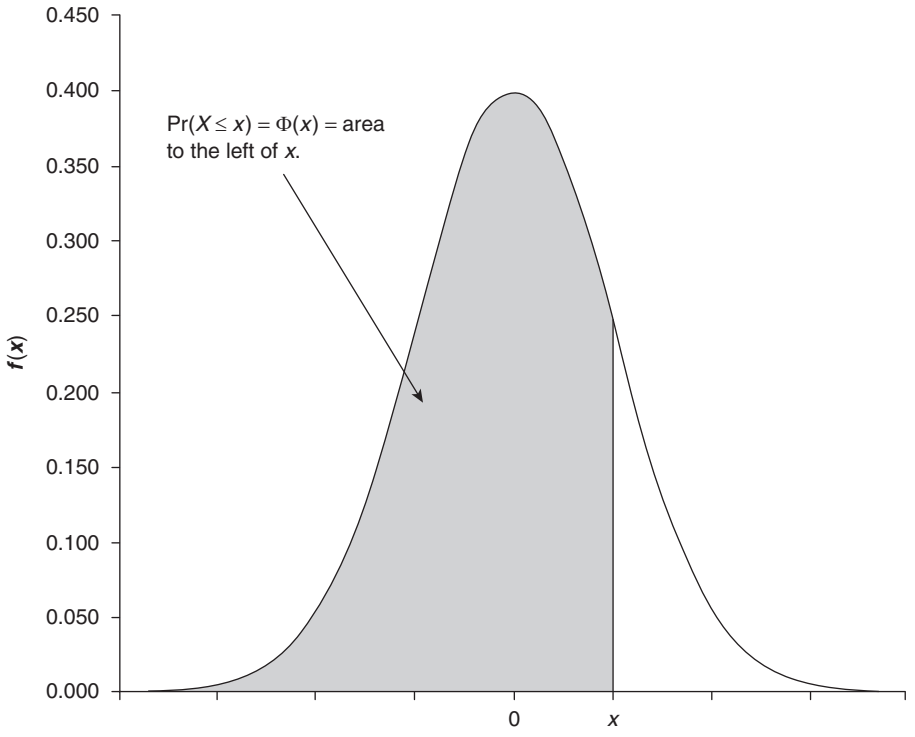


Figure 4 Graphical Illustration of Cumulative Distribution Function $\Phi(x)$ for a Standard Normal Distribution

The cdf for a standard normal distribution is denoted as

$$\Phi(x) = \Pr(X \leq x), \qquad [4]$$

where $X \sim N(0, 1)$, and this function is shown as in Figure 4. Φoxpgs the shaded area under the curve, to the left of x. The whole area under the curve is 1. For example, $\Phi(1, 96) = \Pr(X \leq 1.96) = 0.975$. That is, if X follows a standard normal distribution, then 97.5% of the values from this distribution will be less or equal than 1.96. A table for the standard normal distribution is provided in most statistics books, which allows calculation of this type of probability. Many software packages also provide a function to calculate the cdf. For example, $\Pr(X \leq 1.96)$ could be evaluated using function NORMSDIST in EXCEL 2003 as NORMSDIST(1.96) = 0. 975.

From the symmetric properties of standard normal distribution, the following relationship exists for cdf (Figure 5):

$$\begin{aligned}\Phi(-x) &= \Pr(X \leq -x) = \Pr(X \geq x) \\ &= 1 - \Pr(X \leq x) = 1 - \Phi(x).\end{aligned} \qquad [5]$$

Furthermore,

$$\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X \leq a) \quad \text{for any values } a < b. \qquad [6]$$

The above formulas provide means to evaluate the probabilities of the standard normal distribution. For any normal distribution $N(\mu, \sigma^2)$, the probabilities would be evaluated by first transforming $N(\mu, \sigma^2)$ into a standard normal distribution. Note that as stated in Equation 3 above, if $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$. Therefore, to find the probability of values between a and b with $a < b$ from the distribution $X \sim N(\mu, \sigma^2)$, use

$$\begin{aligned}\Pr(a \leq X \leq b) &= \Pr\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\ &= \Pr\left(Z \leq \frac{b-\mu}{\sigma}\right) - \Pr\left(Z \leq \frac{a-\mu}{\sigma}\right).\end{aligned} \qquad [7]$$

Recall that the area under the curve is 1 no matter what values μ and σ^2 take on for a given normal distribution. Now, let's go back to the blood pressure example. In a population, diastolic blood pressure

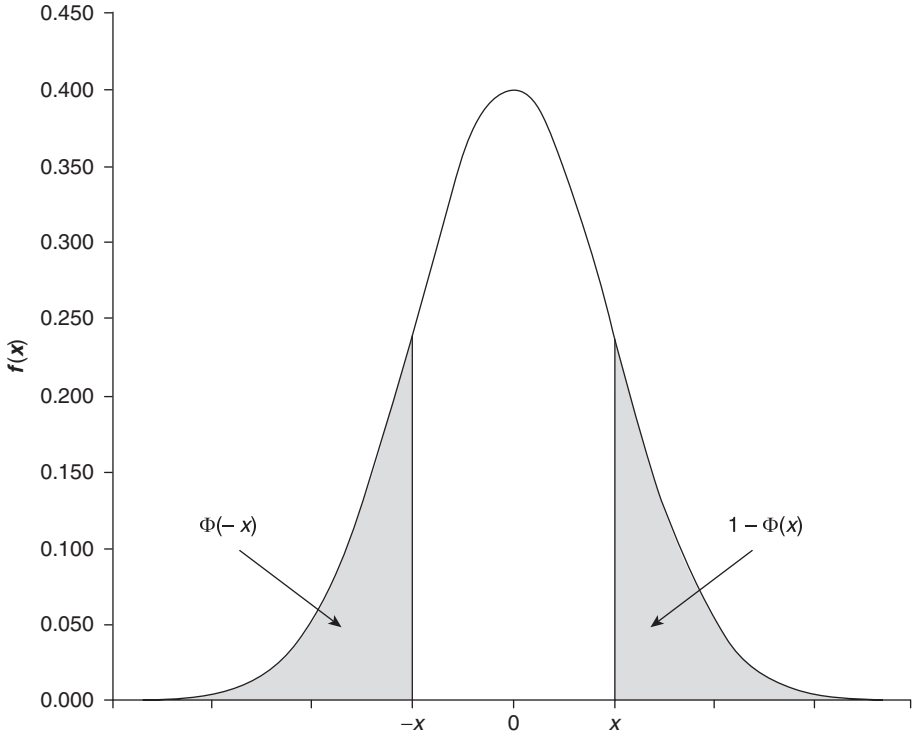


Figure 5 Graphical Illustration of the Symmetry in Cumulative Distribution Function $\Phi(x)$ for a Standard Normal Distribution

follows $N(80, 10)$. To calculate the proportion of prehypertensive individuals in this population, use the formula

$$\begin{aligned}\Pr(80 \le X < 90) &= \Pr\left(\frac{80-80}{10} \le Z < \frac{90-80}{10}\right)\\ &= \Pr(0 \le Z < 1)\\ &= \Pr(Z < 1) - \Pr(Z \le 0)\\ &= 0.50 - 0.16 = 0.34.\end{aligned}$$

So about 34% of people are prehypertensive based on diastolic blood pressure with a value between 80 and 90. Note that for any continuous distribution, $\Pr(X \le x) = \Pr(X < x)$ because the probability of any point is 0.

The cdf can be used to precisely calculate the probability of data falling into any range, given the mean and standard deviation of a data set that follows the normal distribution. But there is an empirical rule that is particularly useful in dealing with normal distributions:

- Approximately 68% of the data will fall within a standard deviation of the mean.
- Approximately 95% of the data will fall within 2 standard deviations of the mean.
- About 99% and 99.7% of the data will fall within 2.5 and 3 standard deviations of the mean.

The above empirical rule provides a handy quick estimate of the spread of the data and is quite useful in practice.

—Rongwei (Rochelle) Fu

See also Central Limit Theorem; Confidence Interval; Hypothesis Testing; Sampling Distribution

Further Readings

Casella, G., & Berger, R. L. (2001). *Statistical inference* (2nd ed.). Belmont, CA: Duxbury Press.

Krishnamoorthy, K. (2006). *Handbook of statistical distributions with applications.* Boca Raton, FL: Chapman & Hall.

Patel, J. K., & Read, C. B. (1996). *Handbook of the normal distribution* (2nd ed.). New York: Marcel Dekker.

Rosner, B. (2006). *Fundamentals of biostatistics* (6th ed.). Belmont, CA: Thomson Brooks/Cole.

Notifiable Disease

Notifiable diseases are those for which regular collection of case information is deemed necessary in preventing and controlling the spread of disease among the population. State and local officials have the authority to mandate diseases reporting within their jurisdictions. Nationally notifiable diseases are suggested by the Council of State and Territorial Epidemiologists (CSTE); participation by states and territories in the National Notifiable Disease Surveillance System (NNDSS) is voluntary.

History

To prevent the introduction and subsequent spread of cholera, smallpox, plague, and yellow fever in the United States, in 1887 the Congress authorized the U.S. Marine Hospital Service, now the Public Health Service (PHS), to collect case data from overseas consuls. The collection of information on these first four notifiable diseases was expanded to include cases in the Unites States in 1893. Until the 1950s, state and territorial health authorities worked with the PHS to designate additional notifiable diseases. The first annual summary of notifiable disease, published in 1912, included reports from 19 states, the District of Columbia, and Hawaii for 10 infectious diseases.

By 1928, 29 diseases were being reported by all states, the District of Columbia, Hawaii, and Puerto Rico. In 1951, the CSTE was formed and became responsible for designating diseases to be included in the NNDSS. That year, 41 infectious diseases were nationally notifiable.

Traditionally, notifiable diseases were infectious diseases. However, in 1995 the first noninfectious condition, elevated blood-lead levels, was added to the NNDSS. The following year the first risk factor, cigarette smoking, was added. In 2006, more than 60 diseases and conditions were nationally notifiable.

Local, National, and International Notifiable Diseases

While CSTE suggests nationally notifiable diseases, it does not have the authority to require states' participation in the NNDSS. Within its jurisdictions, states and territories have the authority to designate which diseases must be reported. Based on the needs and resources of each region, local notifiable diseases lists may exclude diseases included in the NNDSS and include diseases not surveilled nationally. Notifiable disease lists are not static; diseases are added or removed based on current public health needs. Notifiable diseases may be classified on the urgency of reporting and assigned varying time requirements. Generally, physicians and diagnostic laboratories are responsible for reporting cases to local health authorities who, in additional to immediate control and prevention activities, report cases to the state health departments.

All states and territories are required to report cases of cholera, plague, yellow fever, and other quarantinable diseases of international concern. Internationally reportable diseases are dictated by the International Health Regulations set forth by the World Health Organization.

—Michelle Kirian

See also Centers for Disease Control and Prevention; Council of State and Territorial Epidemiologists; Public Health Surveillance

Further Readings

Centers for Disease Control and Prevention. (2006, January). *National Notifiable Diseases Surveillance System.* Retrieved November 6, 2006, from http://www.cdc.gov/EPO/DPHSI/nndsshis.htm.

World Health Organization. (2006). *Epidemic and Pandemic Alert and Response (EPR): International Health Regulations (IHR).* Retrieved November 8, 2006, from http://www.who.int/csr/ihr/en.

Web Sites

Council of State and Territorial Epidemiologists: http://www.cste.org.

Null and Alternative Hypotheses

Stating null and alternative hypotheses has become one of the basic cornerstones of conducting epidemiological research. A hypothesis is typically defined as a tentative proposal or statement that explains certain observations or facts and is testable by further investigation. Testing

hypotheses allows researchers to assess scientifically whether the explanation in question can be falsified. Critical to this process is the idea that, in research, it can never be directly proven that a proposition is true. To do so would imply that the results of a single study would hold across all time, all persons, and all cultures. Therefore, falsification of the null hypothesis has become the basis of scientific investigation as currently practiced.

Researchers approach the idea of "truth" indirectly by developing and testing null and alternative hypotheses. Typically, null and alternative hypotheses are stated so that they are mutually exclusive and exhaustive. The null hypothesis, written as H_0, is the statement that the researcher hopes to reject. Specifically, it is a claim about a population parameter that is assumed to be true until it is declared false. Many times, but not always, the null hypothesis represents a null effect (i.e., there is no relationship between the independent and dependent variable). For example, in a cohort study examining tobacco use and lung cancer, the H_0 might be that smoking status is *not* significantly associated with the development of lung cancer. The alternative hypothesis, denoted as H_A or H_1, is the basic statement that is tested in the research; in the tobacco study example, the H_A might be that smoking status *is* significantly associated with the development of lung cancer. After stating the null and alternative hypotheses, researchers aim to find evidence to reject the null hypothesis; otherwise, they would state that they "failed to reject" the null. If researchers do find enough evidence to reject the null hypothesis, they still might not be able to theoretically "accept" the alternative hypothesis. This is because, in theory, the methods of hypothesis testing are probabilistic, and by definition, probability includes some level of uncertainty. This idea is similar to the guilty/not guilty decision in our judicial system. In finding a defendant not guilty, the jury determines that there is insufficient evidence to find the person guilty; this is not the same as claiming that he or she is innocent. However, in practice, when researchers reject the null hypothesis in their study, they generally do "accept" the alternative at least under the conditions of their specific experiment.

Because hypotheses are developed to be testable, they must be stated in a clear, unambiguous manner. The alternative hypothesis might describe a relationship with a specific direction (e.g., $\mu \geq 100$ or "smoking status is positively associated with the development of lung cancer") or could relate to the statistical test being employed (e.g., the odds ratio $\neq 1$ or $\beta \neq 0$). Conversely, the null hypothesis could indicate no effect (e.g., smoking status is not associated with the development of lung cancer or $\beta = 0$) or could describe the opposite direction of what researchers are investigating (e.g., $\mu < 100$). It is important to note that when writing null or alternative hypotheses, the population parameters (e.g., μ or p) are used instead of the sample statistics (e.g., $\bar{x}$ or $\hat{p}$) since researchers aim to make inferences to the population level.

History of Hypothesis Testing

The concept of hypothesis testing became institutionalized in research in the mid-1950s. In addition to being influenced by philosophers of science such as Karl Popper, the current hypothesis testing process has incorporated two approaches, the Fisher approach to null hypothesis testing and Neyman-Pearson decision theory. The original Fisher approach calls for the researcher to set a statistical null hypothesis and to report the exact level of significance (e.g., $p = .05$). This approach does not state a specific alternative hypothesis and was suggested only when researchers know little about the problem at hand. Building off Fisher's approach, the Neyman-Pearson theory indicates that two hypotheses are created and a choice must be made. In this approach, two statistical hypotheses are developed and α, β, and sample size are decided before the study begins. If study findings fall into the rejection region of the first hypothesis, then there appears to be more solid evidence to support the second hypothesis. The problem is that this approach puts the researcher at risk for making two types of error solely on the basis of the data. The first—and most egregious—is a Type I or an α (alpha) error, which occurs when the researcher decides to reject the null hypothesis when it is actually true. The value of α represents the probability of committing this type of error. The other type of error—a Type II or β (beta) error—occurs when the null hypothesis should be rejected but is not.

Critiques of Hypothesis Testing

While hypothesis testing remains an essential component of epidemiologic research, there are several criticisms of the method. Depending on how the hypotheses are stated, hypothesis testing may not provide any information on the magnitude or direction of the relationship. Evidence for whether to reject or not

reject the null hypothesis may also be highly sensitive to sample size and significance level (α) that was set. Most important, researchers should never use statistics mechanically. Past literature, clinical meaningfulness, and study rigor should all be carefully considered when conducting studies to test hypotheses. Additionally, to minimize problems in hypothesis testing, researchers should develop clear, testable hypotheses and report effect size and confidence intervals when discussing study results.

—Lisa S. Wolff

See also Significance Testing; Study Design; Type I and Type II Errors

Further Readings

Bohrnstedt, G. W., & Knoke, D. (1988). *Statistics for social data analysis.* Itasca, IL: F. E. Peacock.

Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: The *P* value fallacy. *Annals of Internal Medicine*, *130*(12), 995–1004.

Mann, P. S. (2001). *Introductory statistics*. New York: Wiley.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241–301.

Rothman, K. J. (2002). *Epidemiology: An introduction.* New York: Oxford University Press.

Nutritional Epidemiology

Although observations on relationships between diet and health have always been recognized, the systematic science of nutritional epidemiology in populations is relatively recent. Important observations propelling the field of nutrition forward were numerous in the 18th and 19th centuries, as it was recognized that deficiencies in certain classes of foods led to important diseases such as scurvy, rickets, pellagra, and beriberi. This was followed by the rapid sequential discovery of the vitamins in the early 20th century. Since then, the focus of diet and health has shifted to include the problems of obesity and the role of diet on chronic disease risk. Just as the parent field of epidemiology is defined as the investigation of the frequency, distribution, and risk factors influencing disease in populations, nutritional epidemiology may be defined as the frequency and distribution of nutrition-related diseases as well as the relation of nutritional intake and status to disease outcomes.

Nutrition Monitoring and Surveillance

The systematic quantification of nutritional status with population surveys has been done only relatively recently. Early surveys in the United States, such as the Ten-State Nutrition Survey released by the U.S. Centers for Disease Control in 1971, identified stunted growth associated with insufficient food in low-income groups. Such results led President Johnson to declare a "War on Hunger" in the United States in 1966. In the relatively short time since then, food availability has changed rapidly with associated changes in nutritional risk, such that we have gone from substantial undernutrition to a current epidemic of obesity in all income categories. At the same time, heart disease and cancer have escalated as the major diseases contributing to mortality as we have made progress against deficiency and infectious disease. Fortunately, these changes have been well documented at the national level with the representative U.S. Department of Agriculture Continuous Survey of Food Intake of Individuals and the U.S. Department of Health and Human Services National Health and Nutrition Examination Surveys (NHANES) that have been conducted periodically since 1970. These have now been combined into the Continuous NHANES, which is released for public use in 2-year cycles.

With the encouragement of the World Health Organization, nutritional surveillance activities are now also implemented in many developing countries. These include organized community weighing programs, school entry height measurements, and the addition of nutrition modules to existing national surveys to provide data that document trends in nutritional status and support policy decisions to target and improve identified nutrition problems.

Diet and Health

A central focus of nutritional epidemiology is to inform dietary recommendations by clarifying the role of food and nutrient intake to the risk of health outcomes. Early work in this area began from ecological observations that there were large differences in the prevalence of differing chronic diseases across countries with

differing food patterns. For example, death rates from heart disease or colon cancer have been shown to be positively associated with higher-than-average fat intakes across countries. However, such an association does not necessarily imply a causal relationship—particularly in such ecologic observations, as numerous other exposures differ across countries. Migration studies, documenting change in risk associated with the movement of groups from their native country to a new environment, provide further evidence of the importance of environmental exposures, as opposed to genetic variation, in disease risk. Central among these are changes in diets, adding more evidence for several diet and health relationships.

However, the relationship between behavior and nutritional status or nutrition and health is complex—chronic diseases, in particular, are due to a constellation of factors. The omnipresent possibility of alternative explanations precludes the ability to prove causality in observational nutritional epidemiology. To infer likelihood of causality, the criteria developed by Austin Bradford Hill (1897–1991), a British statistician, are followed. They include temporal relationship (a risk factor must precede the onset of disease), strength of statistical association, evidence of dose-response, consistency of the relationship across studies and population groups, biological plausibility, and careful consideration of possible alternative explanations.

Several of these criteria are challenging to fulfill in studies of diet and disease. For rare diseases, such as most cancers, the case-control design is usually used. Dietary intake and other exposures must be recalled by patients and/or their proxies as well as by matched controls. Although these studies have provided considerable information on likely dietary risk factors, concern about the likelihood of recall bias must always be considered. Longitudinal cohort studies offer several benefits in that the exposure—diet and nutritional status—may be measured prior to the development of the disease, providing evidence of temporality and avoiding much of the reporting bias. Such studies, including the Harvard University Nurses and Male Health Professionals Health Studies and the Framingham Heart Study, among others, have provided a tremendous amount of information that has improved our understanding of the central role that dietary intake plays in protecting health and preventing chronic disease. Although such large studies are few because of their high cost, they continue to provide exciting results.

Collinearity across foods and nutrients is another limitation in identifying causal associations between specific nutrients and disease outcomes. For example, diets that are high in fat are often also low in fruit and vegetables and associated nutrients. Therefore, it becomes difficult to tease out the effects of any single nutrient. Furthermore, the discovery of the importance of numerous chemical constituents in food and of interactions across nutrients, in combination with the failure of several large vitamin supplementation trials to show benefit, has increased the understanding that whole foods rather than single nutrients appear to be most protective. In response to this, dietary patterns research, looking at the total diet defined by computer algorithms based either on the correlation matrix of individual food group intakes (factor analysis) or the spatial distance across food group intakes (cluster analysis) has emerged as a complementary way of confirming the importance of diet on health.

All dietary methods suffer from random error due to day-to-day variation and imprecision in reporting. This has the tendency to attenuate correlations and relative risk estimates in relation to the true associations. Advances in assessment methodology and in statistical approach, including improved understanding of measurement error, have improved the ability to estimate associations and to adjust for alternative explanations.

Although nutritional epidemiology holds much in common with other branches of epidemiology, it poses unique challenges. In particular, the measurement and interpretation of dietary exposures require careful consideration in the design, implementation, and analysis. Much work during recent decades has been devoted to the measurement, validation, and interpretation of error in dietary assessment.

Dietary Exposures

Dietary intake is a complex exposure that includes a nearly infinite mix of foods, portion sizes, and preparations. Exposures of interest may include foods themselves, or more commonly, nutrients. These include total energy; macronutrients such as total fat, saturated fat, and carbohydrate; micronutrients such as vitamins and minerals; and more recently, a variety of phytochemicals in foods. The validity of any of these measures depends on the quality and availability of a precise nutrition database of food composition. The U.S. Department of Agriculture provides this information based on chemical analysis of foods.

However, constant changes in the food supply and demand for additional nutrients mean that this must be continually updated, and database inadequacies remain an important limitation for work in nutritional epidemiology in many countries.

The methods of dietary assessment in common use each have advantages and disadvantages. Short-term methods include weighed dietary records, where individuals are either observed or asked to record their own food intakes, using dietary scales to weigh portions before and any waste after consumption to get accurate quantitative measures of actual intake. This method provides good quantitative data, but it requires either that researchers observe intakes or that participants themselves record detailed intake data, and this may lead to selection bias. In addition, the attention given to food intake during the implementation of this method has shown that individuals tend to consume less than they usually do—either because of the work involved in the measurement itself or because of their heightened awareness of their intake. A second popular method is the 24-hr dietary recall. Either in person or by telephone, a trained interviewer asks individuals to recall what they ate and drank the previous day. The studies comparing reported with observed intakes have shown good validity for this method, although a tendency to underreporting persists, due to some forgotten foods and underestimation of portion sizes. Improvements in the method include the use of portion-size visuals to improve quantity estimation, and a multiple pass approach to improve completeness. Using the multiple pass approach, individuals are asked to recall their intake in a series of steps that begins with a quick list of foods consumed, followed by reviews that probe for forgotten foods, define eating occasion and time, and complete detailed information on preparation and portion size.

These methods provide valuable data for population intakes, including detailed information on specific foods used, preparation methods, meal patterns, the times of day that foods are consumed, and how foods are consumed together. As such, they continue to be used in national surveys for the purpose of nutrition monitoring and comparison of intakes by population subgroups. However, they have the distinct disadvantage of misclassifying individuals in relation to their usual intake, thereby attenuating the ability to relate dietary intake to health outcomes at the individual level. Because individuals consume differing types and quantities of foods on any given day relative to another, a single day, or even a few days, of intake may misrepresent usual exposure. The likelihood of misclassification varies by nutrient and by the complexity in the diet. The random error associated with this day-to-day variation has been quantified in numerous studies. Based on the ratio of within- (day-to-day) to between-person variation in intake, the number of days of intake needed for a stable estimate can be calculated. For energy and macronutrients, where habit and individual regulation of intake are more direct, a few days may be sufficient. The same is true for micronutrients or food substances present in specific foods that tend to be consumed (or not) regularly in the dietary pattern of the population, such as calcium from milk or caffeine from coffee. On the other hand, obtaining information on consumption of micronutrients that are irregularly distributed in foods that are consumed less frequently, such as vitamin A (very high in liver, and green leafy vegetables, but low in most foods) may require so large a number of days as to be practically infeasible.

For this reason, longer-term methods of usual intake tend to be favored in nutritional epidemiology. Food frequency questionnaires consist of a detailed food list. Subjects are asked to respond to a set of frequency options for each food-line item, which may include groups of foods that are nutritionally similar (such as "green leafy vegetables such as spinach, kale, and collard greens"). Frequency options generally range from never or less than once per month to more than once per day. By linking these responses to a specially designed nutrient database, nutrient intakes can be calculated. Some food frequency questionnaires rely only on frequency responses to calculate nutrients, while others add questions on usual portion size to further quantify exposures. Additional refinements—such as added questions on low fat versus regular versions of specific foods or specific type of breakfast cereal used—are often added to further improve estimation. It has been recognized that certain categories of foods, such as fruits and vegetables, may be overestimated when a long list of items is presented. Therefore, most questionnaires calibrate total fruit and vegetable intake with a summary question on the number of servings of these foods typically consumed per day. Statistical adjustment for total energy intake is generally conducted when examining nutrient intake to outcome measures. This realigns the ranked data on individual foods or nutrients to partially adjust for differences in total food intake due to differing individual energy

requirements, and also tends to correct some of the distortion that may otherwise occur by assuming standard portion sizes.

Validation work comparing such questionnaires with multiple dietary recalls and, more recently, to biomarkers, has generally supported their utility in nutritional epidemiologic research. They have been shown to rank individuals reasonably well with respect to comparison measures of intake—either multiple dietary records or recalls, or biomarkers. Their great advantage is their relatively low cost. However, they also have important limitations. Because of the grouping of foods and lack of information on specific preparations, as well as on actual portion size, the resulting estimation of nutrient intake is only semiquantitative and depends on assumptions included in the associated nutrient database. Recent comparisons with recovery biomarkers—which accurately capture the body's processing of energy or specific nutrients, and thus serve as objective measures of intake—have questioned the validity of food frequency questionnaires for estimating total energy and protein intakes. Furthermore, the food list and assumed preparation of foods will be valid only for the general population for which it was developed. As such, subgroups in the population or other populations with differing dietary patterns may be seriously misrepresented by standard instruments.

Anthropometric Measures

Nutritional status itself may be either an outcome or an exposure in nutritional epidemiology. Stunting and wasting in children and wasting in adults under conditions of severe food shortages remain important outcomes in nutritional epidemiology in underdeveloped countries, and these continue to be studied in relation to the proximal risk factors, food intake and infectious disease, as well as more distal economic and social factors. Measures of undernutrition have been defined for children by the World Health Organization based on growth standards developed with a healthy U.S. population. In addition to measures of relative weight for age, weight for height, and height for age, measures of upper arm circumference and head circumference are frequently used as indicators of inadequate growth and nutritional status.

On the other hand, the prevalence of obesity is increasing rapidly throughout the world. Accepted measures of obesity are based on body mass index (BMI; height in meters/square of weight in kilograms; m/kg^2) of >30 for adults. BMI cutoff values for children have been published by the U.S. Centers for Disease Control and Prevention, based on data from U.S. children, by age and sex group, as measured in national surveys prior to the recent increase in childhood obesity. Obesity is used both as an outcome measure of concern on its own and as a risk factor for many additional health outcomes. Numerous studies have identified obesity as a risk factor for type 2 diabetes, heart disease, and several cancers. There has also been an increasing understanding that fat deposited in the abdomen may pose greater health risks than fat deposited elsewhere in the body, because these fat cells are more metabolically active, releasing inflammatory hormones and fatty acids that negatively affect glucose metabolism, blood pressure regulation, and triglyceride production. Waist circumference has, therefore, emerged as an important measure of disease risk, with current cutoff values for risk as >88 cm for women and >102 cm for men.

Biochemical Indicators

Nutritional status may also be measured with biochemical indicators, or biomarkers, from body tissues—most frequently from plasma, serum, or urine. Iron status, for example, is a common nutrient for which laboratory tests provide accurate assessments. Biomarkers may also be used as indicators of dietary intake and, as such, are used to validate dietary methods. However, the translation of dietary intake into biomarkers is affected by several factors unique to each nutrient, including absorption and conditions of metabolism and homeostatic regulation, and it may be further affected by other factors such as medications, nutrient interactions, and smoking. For some nutrients, such as folate, there is usually a clear relationship between dietary intake and blood concentration, suggesting that either measure may be used to assess status. Others, such as calcium, are tightly regulated by homeostatic mechanisms that maintain blood concentrations within a narrow band. Therefore, usual dietary intake will be a better measure than blood measures. Some nutrients, such as plasma vitamins K, are metabolized rapidly in the blood and represent only recent intake, while others, such as serum ferritin, represent long-term stores.

Unfortunately, reliable biomarkers of dietary intake do not currently exist for most nutrients. Recovery biomarkers that have made important contributions to

advancing the field include doubly labeled water for energy intake and urinary nitrogen for protein intake. Several blood nutrient measures correlate well with intake. These include folate, vitamin B_6, K, C, E, and carotenoids. The blood concentrations of others, such as vitamin A, are regulated and therefore show insufficient variation to relate to diet except at the extremes. Another, vitamin B_{12}, often does not relate well to dietary intake, due to large individual variation in absorption. For some nutrients, functional indicators rather than direct indicators may be more useful. For example, methylmalonic acid—a product of a metabolic reaction that requires vitamin B_{12}—provides a useful measure of vitamin B_{12} status.

The wealth of information that has resulted from nutritional epidemiology studies over recent decades has made major contributions to our understanding of the central importance of good dietary intake patterns to our health—influencing national dietary guidelines and policy, food industry regulation and product formulation, and individual behavior. The field continues to evolve, with improved methodology. Emerging work highlights the important role of genetic variation in defining effects of dietary exposures on individual disease risk with the promise of exciting new findings and more specific individual dietary prescriptions for health—nutritgenomics—in the relatively near future.

—Katherine L. Tucker

See also Gene-Environment Interaction; Hill's Considerations for Causal Inference; Malnutrition, Measurement of; Obesity; Study Design

Further Readings

Bingham, S. A. (2002). Biomarkers in nutritional epidemiology. *Public Health Nutrition*, *5*, 821–827.

Briefel, R. R. (1994). Assessment of the U.S. diet in national nutrition surveys: National collaborative efforts and NHANES. *American Journal of Clinical Nutrition*, *59*, 164S–167S.

Flegal, K. M., Wei, R., & Ogden, C. (2002). Weight-for-stature compared with body mass index-for-age growth charts for the United States from the Centers for Disease Control and Prevention. *American Journal of Clinical Nutrition*, *75*, 761–766.

Frasier, G. E. (2003). A search for truth in dietary epidemiology. *American Journal of Clinical Nutrition*, *78*, 521S–525S.

Gibson, R. (2005). *Principles of nutritional assessment.* New York: Oxford University Press.

Jacques, P. F., & Tucker, K. L. (2001). Are dietary patterns useful for understanding the role of diet in chronic disease? *American Journal of Clinical Nutrition*, *73*, 1–2.

Kaaks, R., Ferrari, P., Ciampi, A., Plummer, M., & Riboli, E. (2002). Uses and limitations of statistical accounting for random error correlations in the validation of dietary questionnaire assessments. *Public Health Nutrition*, *5*, 969–976.

Margetts, B. M., & Nelson, M. (Eds.). (1991). *Design concepts in nutritional epidemiology.* New York: Oxford University Press.

Mason, J. B., & Mitchel, J. T. (1983). Nutritional surveillance. *Bulletin of the World Health Organization*, *61*, 745–755.

Ordovas, J. (2006). The quest for cardiovascular health in the genomic era: Nutrigenetics and plasma lipoproteins. *Proceedings of the Nutrition Society*, *63*, 145–152.

Willett, W. (1998). *Nutritional epidemiology.* New York: Oxford University Press.

O

Obesity

During the final two decades of the 20th century, a dramatic increase in the prevalence of obesity occurred in the United States and in many developed and developing countries throughout the world. By the end of the 20th century, there were more than 1 billion overweight adults worldwide, among whom more than 300 million were obese. Obesity had become a significant contributor to the global burden of chronic disease and disability and a major public health issue. The rising rates among children were a particular concern. This entry considers the definition and epidemiology of obesity, its health consequences, the causes of obesity, and efforts to address it through public policy.

Definition

Obesity is defined as increased body weight related to excess accumulation of body fat. The term commonly refers to a range of weight above that which is considered healthy for a given height. A simple and widely used method for estimating body fat, the body mass index (BMI), or Quetelet index, is calculated by dividing weight in kilograms by the square of height in meters (BMI = kg/m^2). BMI does not measure body fat directly but correlates with direct measures of body fat, such as underwater weighing and dual-energy X-ray absorptiometry (DXA).

The current BMI categories for adults 20 years and older published by the U.S. National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health is similar to the classification used by the World Health Organization (WHO) (see Table 1). The primary BMI cutoff points for excess weight occur at 25, 30, and 40 kg/m^2. Obesity is defined as a BMI of 30 or more.

For children and adolescents, BMI is plotted on national growth charts developed by the U.S. Centers for Disease Control and Prevention (CDC) to obtain percentile rankings among children of the same sex and age in the United States. The CDC and the American Academy of Pediatrics (AAP) recommend the use of BMI to screen for overweight in children and adolescents aged 2 to 19 years. Health advocates prefer the term *overweight* over *obese* to avoid the potential stigma of the label in this age group. The CDC BMI-for-age weight status categories for children and adolescents are provided in Table 2.

As a simple and inexpensive method for measuring relative weight, BMI cannot distinguish between

Table 1 Weight Status Categories for Adults Aged 20 Years and Older

Weight Status	*Obesity Class*	*BMI (kg/m^2)*
Underweight		<18.5
Healthy weight		18.5–24.9
Overweight		25.0–29.9
Obesity	I	30.0–34.9
	II	35.0–39.9
Extreme obesity	III	≥40.0

Source: U.S. National Institutes of Health.

Table 2 Weight Status Categories for Children and Adolescents Aged 2 to 19

Weight Status	*Percentile Range*
Underweight	Less than the 5th percentile
Healthy weight	5th percentile to less than the 85th percentile
At risk of overweight	85th to less than the 95th percentile
Overweight	Equal to or greater than the 95th percentile

Source: U.S. Centers for Disease Control and Prevention.

increased weight for height due to body fat from that attributable to fat-free mass (muscle, bone, and fluids). Thus, BMI may lead to overestimates of adiposity in athletes, for example. More direct methods of estimating body fat include skinfold thickness, ultrasound, computed tomography, and magnetic resonance imaging (MRI).

Among adults, the NHLBI guidelines recommend assessing two additional predictors of potential health risks associated with overweight and obesity, in addition to BMI. These are waist circumference, as abdominal fat is associated with greater health risk, and also other risk factors and comorbidities associated with obesity, such as high blood pressure and physical inactivity. Both absolute waist circumference and waist-to-hip ratio are used to assess central obesity, also known as ''apple-shaped'' or ''masculine'' obesity, in which the main fat deposits are stored around the abdomen and upper body.

Epidemiology

Obesity became a leading global public health concern by the end of the 20th century. In the United States, obesity prevalence among adults aged 20 years or older doubled between 1980 and 2002. By 2003 to 2004, an estimated 66.3% of U.S. adults were either overweight or obese, including 32.2% who were classified as obese. During the final two decades of the 20th century, obesity rates among adults increased threefold or more in parts of the United Kingdom, Eastern Europe, the Middle East, the Pacific Islands, Australasia, and China. In developing countries, excess body fat may often paradoxically coexist with undernutrition. While global recognition of the growing public health problem did not occur until the end of the 20th century, by 2002, obesity was considered the sixth most important risk factor contributing to the overall burden of disease worldwide. At least 1.1 billion adults worldwide were classified as overweight or obese, including 312 million who were obese.

The burden of obesity is not evenly distributed throughout populations. Obesity tends to be higher among women, the poor, and racial/ethnic minorities. Regional differences may also be observed. For example, some Asian countries have low overall rates of obesity but relatively high rates of central obesity, with the accompanying excess risks of diabetes and cardiovascular disease.

Among children and adolescents aged 6 to 19 years in the United States, obesity prevalence tripled between 1980 and 2002. By 2003 to 2004, an estimated 33.6% of U.S. children and adolescents aged 2 to 19 years were at risk of overweight or overweight, including 17.1% who were overweight. Globally, 10% of children were at risk of overweight or overweight by 2002.

Health Risks and Consequences

The growing rates of overweight and obesity became a major public health concern because of the elevated health risks associated with excess body weight. Obesity is related to a higher prevalence of intermediate metabolic consequences and risk factors, such as high blood pressure, elevated triglycerides (blood fat), decreased HDL cholesterol (''good cholesterol''), and the so-called metabolic syndrome, defined in the United States as three of five features: large waist circumference, abnormal concentrations of triglycerides, HDL cholesterol, fasting glucose, and hypertension. Obesity is also associated with several health outcomes, including type 2 diabetes, coronary heart disease, stroke, osteoarthritis, sleep apnea and respiratory problems, and some cancers, including endometrial, colon, gall bladder, prostate, kidney, and postmenopausal breast cancer.

Overweight and obesity are also associated with premature deaths. In the United States, an estimated 300,000 deaths per year can reportedly be attributed to obesity, which would make it second only to smoking as the main preventable cause of illness and premature death. Mortality appears to increase on a continuum with increasing body weight, with

a small increase in risk in those with a low BMI (below 20 or 22 kg/m^2) and a greater increase in those with a BMI above 30 kg/m^2.

Among children and adolescents, weight-related health consequences are expected to continue to increase in the coming years. Cardiovascular risk factors, such as high blood pressure and high cholesterol, are becoming increasingly common in this group, while ''adult'' diseases, such as type 2 diabetes, have increased dramatically among overweight adolescents. Other health conditions related to overweight among children and adolescents include asthma, sleep apnea, and nonalcoholic fatty liver disease. The most immediate consequences for overweight children are probably psychosocial, including social discrimination, decreased self-esteem, and decreased quality of life. Overweight children have an increased chance of becoming overweight or obese adults, with the accompanying health risks.

Causes and Solutions

Obesity is a consequence of excess energy (caloric) intake over total daily energy expenditure. While genes are important determinants of individuals' susceptibility to weight gain, energy balance results from the complex interactions between genetic, biological, psychological, behavioral, sociocultural, and environmental factors that affect both energy intake and expenditure.

While the exact causes remain uncertain for the increase in obesity rates among U.S. children, adolescents, and adults, several societal trends may be implicated. Nutritional patterns changed in recent decades as both parents entered the workforce, more meals were eaten outside the home, and marketing by the food industry increased the consumption of take-out foods and larger portion sizes. Levels of physical activity declined as the use of automobiles, television, computers, and mechanical aids increased. Changes to the physical design of communities, such as the increase in urban sprawl, created further barriers to physical activity. In poor communities and nonwhite communities, neighborhood safety and the lack of recreational facilities, supermarkets, and healthy food outlets present additional barriers to physical activity and healthy eating.

Likewise, the increase in obesity in other parts of the world may have been fueled by modernization, urbanization, and the globalization of food markets. Many developing countries are undergoing the nutrition transition toward increased consumption of energy-dense foods high in saturated fats and sugars, coupled with reduced physical activity. Fetal undernutrition followed by rapid childhood weight gain may contribute to the development of insulin resistance and the metabolic syndrome, a pattern observable in developing countries such as India and China.

Growing concern over the rising prevalence of obesity in the population overall and in children, in particular, has prompted increasing programmatic and policy responses at the local, national, and international levels. Government initiatives have included funding programs for after-school activities, removal of vending machines in schools, nutritional labeling of food products, and walking initiatives. Further efforts that engage multiple sectors, including government, industry, media, health systems, workplaces, schools, communities, and families will be needed to stem the tide of the obesity epidemic.

—Helen L. Kwon

See also Body Mass Index (BMI); Cardiovascular Disease; Chronic Disease Epidemiology; Diabetes; Urban Sprawl

Further Readings

Hedley, A. A., Ogden, C. L., Johnson, C. J., Carroll, M. D., Curtin, L. R., Flegal, K. M. (2004). Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. *Journal of the American Medical Association*, *291*(23), 2847–2850.

Institute of Medicine. (2005). *Preventing childhood obesity: Health in the balance*. Washington, DC: National Academies Press.

U.S. Centers for Disease Control and Prevention. (n.d.). *Overweight and obesity*. Retrieved September 22, 2005, from http://www.cdc.gov/nccdphp/dnpa/obesity/index.htm.

U.S. Department of Health and Human Services. (2001). *The Surgeon General's call to action to prevent and decrease overweight and obesity*. Rockville, MD: Author.

World Health Organization. (2003). Diet, nutrition and the prevention of chronic diseases. In *Technical report series 916.* Geneva, Switzerland: Author.

Observational Studies

Observational epidemiology refers to the branch of epidemiology devoted to using nonexperimental

studies to describe the health status of populations and generate evidence about determinants of health outcomes. Experimental designs in epidemiology, generally referred to as clinical trials, involve assignment of the principal independent variable, the "treatment," to subjects. Often this assignment is randomly allocated, which offers profound advantages for making causal inferences, but it can also be nonrandom. In observational studies, the investigators do not assign treatment to subjects. The principal independent variable is some endogenous or exogenous exposure observed as it naturally occurred. When observational epidemiologic studies are designed to draw inferences about health outcome across different exposure groups, they are considered "analytic." When they are intended only to describe the frequency of a risk factor or disease in a population, they are considered "descriptive." However, the line between descriptive and analytic observational epidemiologic studies is often blurred as descriptive studies typically contrast disease frequency endpoints across population subgroups, and analytic studies can also report on the absolute frequency of disease or exposure in the population sample under study. This entry describes cohort studies, case-control studies, cross-sectional studies, and ecologic studies and discusses the problem of confounding and ways in which it can be addressed.

There are fundamental challenges in making causal inferences about associations between exposures and diseases based on evidence from observational epidemiologic studies. In any analytic epidemiologic study, the goal is to estimate average causal effects in groups (i.e., we do not scrutinize individual participants, subject by subject, to see if in each case there is biological evidence linking exposure and disease). Because there is no way to observe the average disease experience of the exposed group under conditions of no-exposure (the "counterfactual" condition), estimating the average causal effect of exposure is done by comparing the disease experience of the exposed group with that of a different group of individuals without exposure. Consequently, the fundamental challenge to valid estimation of these effects is the similarity of these groups (or their "exchangeability"). Here, randomization is a great help because study groups that are randomly assigned to exposed versus unexposed status will, on average, be similar. However, when exposed and unexposed groups are merely observed as they occur in nature, it is extremely unlikely that they will be similar. If the groups differ on factors other than exposure that are associated with disease risk, estimated associations between exposure and disease will be confounded. Minimization of the influence of confounding is, consequently, critical to success in observational epidemiologic studies.

Types of Observational Studies

Analytic observational studies with individuals as the unit of analysis are generally categorized into three types: cohort studies, case-control studies, and cross-sectional studies. Observational studies correlating only group-level information on risk factors and outcomes are ecologic studies.

Cohort Studies

Cohort studies entail the follow-up of populations or population samples for incident disease endpoints. Exposure status is characterized at baseline and is commonly also tracked for change over time. Follow-up can occur coincident with calendar time (prospective or concurrent cohort studies) or, retrospectively, where disease status on the cohort members is known at the time the study is initiated and available data sources are used to "assemble" and characterize exposure in the cohort at baseline and is followed forward to the present (retrospective or nonconcurrent cohort studies). In cohort studies, analyses are based on contrasting disease experience in exposure groups. The approach of contrasting disease experience can be based on cumulative incidence, incidence rates, or time-to-event. Methods have been developed and are widely available to account for censoring—the loss of the ability to follow study subjects either at the start (late-entry or left censoring) or end (right censoring) of follow-up. Similarly, analyses of cohort data can be based on current or cumulative exposure at baseline and can consider time-varying exposure over the course of follow-up.

Case-Control Studies

Case-control studies are retrospective studies that involve sampling conditional on disease status. A sample of cases, typically incident cases, is selected, as is a sample of noncases (controls) drawn from the same population giving rise to cases. Matching

of controls to cases on potential confounders is a commonly used design feature in case-control studies. The measure of association between exposure and disease is based on a comparison of the odds of exposure in cases and controls. Because the case-control design is retrospective and involves sampling of controls, it tends to be less resource intensive than the cohort design that involves follow-up of large number of subjects. However, the process of selection of the case and control groups can introduce bias, as can any misclassification arising from the retrospective nature of exposure assessment. Cohort studies are by no means immune to selection and information biases but, because of their design, case-control studies typically require extra efforts to protect against bias.

The case-control design has been adapted for implementation within prospective cohort studies. Here, cases that arise during follow-up of the cohort make up the case group, and controls are either a sample of noncases matched on time of case diagnosis (the nested case-control study) or a sample of the entire cohort enrolled at the start of the study (the case-cohort design). These approaches take advantage of the prospective data collection and follow-up of the cohort design and economize by requiring exposure assessment only on cases and controls as opposed to the entire cohort. This is particularly useful in investigations involving costly biomarkers of exposure. Nested case-control studies are analyzed with conventional matched methods, and a range of more specialized techniques has been developed for case-cohort studies. Other adaptations of the case-control design also exist, with the most notable being the case-crossover study. This design is particularly well-suited for the study of trigger exposures—that is, those exposures occurring proximate to disease such as car phone use prior to automobile accidents or physical exertion prior to myocardial infarction. In the case-crossover study, exposure in cases is assessed during a short time window preceding disease onset. Rather than comparing this exposure experience with that of an independent control group as one would do in a conventional case-control study, it is contrasted with the exposure experience during a different time window within each case. Under this design, each case essentially serves as its own control, thus limiting the possibility of confounding by factors that are fixed or change slowly over time.

Cross-Sectional Studies

A cross-sectional study is based on a study sample taken at a particular point in time, with outcome and exposure data assessed simultaneously. This design is distinct from a case-control design in that there is no sampling conditional on disease status and no focus on incident disease, and there tends to be less effort to develop retrospective data on exposures. The cross-sectional study is most appropriately used in descriptive epidemiology to generate estimates of prevalence of health outcomes and risk factors in the population. Causal inference based on correlation of these cross-sectional measures is vulnerable not only to confounding but to a number of other major biases (e.g., incidence prevalence bias, reverse causation bias).

Ecologic Studies

Finally, observational studies that correlate data on risk factors and outcomes measured only at the group level are referred to as ecologic studies. Because the actual outcome status of individuals with particular exposure status is not known in these designs, they have been criticized as being subject to bias (referred to as ecologic fallacy) in making inferences about individual-level associations between risk factors and outcomes. However, associations between group-level variables can be important tools in evaluating community health. For example, there can be a true negative association between socioeconomic status and a disease occurrence at the individual level within communities, but across communities the group-level association between average socioeconomic status and average disease risk can be positive. If the individual-level data were used to anticipate the direction of the group-level association, inference would be biased (the atomistic or individualistic fallacy). Understanding the mechanisms behind individual and group-level associations can be quite informative in the planning of public health interventions. Multilevel analysis is the approach taken in studies designed to simultaneously assess the effect of both individual and group-level variables on individual outcomes.

The Challenge of Confounding

As mentioned, the principal threat to validity of causal inference in analytic observational epidemiology studies conducted at the individual level is confounding.

When designing observational studies, considerable attention needs to be placed on strategies for confounder control. This process should begin with careful consideration of the causal model underlying the outcome under study. Directed acyclic graphs, diagrams linking variables by arrows representing direct causal effects that also illustrate the underlying causal determinants of observed statistical associations, are being used increasingly by epidemiologists as a tool to help understand the fairly complex confounding structures that can arise even when as few as two or three potential confounders are under consideration. Options available for confounder control in the design phase of a study include restriction and matching, while stratification and multivariable methods are options during the data analysis phase of the study. Propensity scores are being considered more and more as an efficient means of controlling for multiple measured confounders through both matching and adjustment. A better understanding of the potential impact of unmeasured confounders can be gained through formal sensitivity analyses. Periodically, epidemiologists also debate the merits and applicability of instrumental variable model approaches, commonly used in econometrics, for control of unmeasured confounders, but the assumptions involved in these models have, to this point, limited their use in observational epidemiology.

Given the substantive challenges posed by confounding in observational epidemiology and the ability of randomization to address these, the randomized controlled trial (RCT) is viewed as the superior study design. In instances where conflicting evidence is generated by RCTs and observational epidemiologic studies, the temptation is to presume that the RCT results are more valid. However, there are a number of points to consider when comparing observational epidemiology with RCTs. First, there are a large number of risk factors that can never be evaluated in RCTs because it would be unethical or impossible to randomize. Next, RCTs are often conducted on select population subgroups, for example, those of a certain age or background risk level. While results from these trials may have superior internal validity for the subgroup studied, findings from observational studies on broader population groups can have more generalizability. Finally, all observational studies are not by definition of equal quality. In a number of research areas, findings from well-designed prospective cohort studies generate findings that agree with RCTs. In these situations, the discrepancies across the bodies of observational and experimental evidence have been driven by findings from case-control, cross-sectional, or ecologic studies.

—Craig Newschaffer

See also Bias; Causation and Causal Inference; Confounding; Descriptive and Analytic Epidemiology; Study Design

Further Readings

Concato, J. (2004). Observational versus experimental studies: What's the evidence for a hierarchy? *NeuroRx*, *1*, 341–347.

D'Agostino, R. B., Jr., & D'Agostino, R. B., Sr. (2007). Estimating treatment effects using observational data. *Journal of the American Medical Association*, *297*, 314–316.

Hernan, M. A. (2004). A definition of causal effect for epidemiological research. *Journal of Epidemiology and Community Health*, *58*, 265–271.

Lash, T. L. (2007). Heuristic thinking and inference from observational epidemiology. *Epidemiology*, *18*, 67–72.

Lawler, D. A., Smith, G. D., & Ebrahim, S. (2004). Commentary: The hormone replacement-coronary heart disease conundrum: Is this the death of observational epidemiology? *International Journal of Epidemiology*, *33*, 464–467.

Oral Contraceptives

Oral contraceptives, commonly referred to as "the pill," provide a hormonal method for women to prevent pregnancy. The two basic types are combination pills, which contain both estrogen and progestin (a synthetic version of progesterone), and progestin-only pills. Since their introduction in 1960, oral contraceptives have become the most popular form of reversible birth control in the United States. More than 100 million women use this method worldwide, although its use varies substantially by country.

History

In 1951, activist Margaret Sanger began to work with Dr. Gregory Pincus to develop a birth control pill. In less than a year, Pincus was able to demonstrate that progesterone inhibits ovulation in rabbits and rats, but he lacked the funding necessary to continue his research. Simultaneously, an orally effective form of

synthetic progesterone was created by Carl Djerassi, a chemist working in Mexico City, and Frank Colton, the chief chemist at the pharmaceutical company G.D. Searle. In 1953, philanthropist Katharine McCormick agreed to provide Pincus with funding for further research. Pincus collaborated with Dr. John Rock for the first human trials in 1954 and submitted the formulation developed by G.D. Searle and Company, called Enovid, for FDA approval in 1956. The FDA approved Enovid for treatment of severe menstrual disorders the following year. Searle received FDA permission in 1960 to sell a lower dose formula of Enovid as a contraceptive. In 1962, the drug company Syntex released another oral contraceptive, Ortho Novum, using the formula developed by Djerassi. The progestin-only pill was developed in the early 1970s in response to concerns about the relationship between estrogen and thrombo-embolic disease. In the 1980s, multiphasic oral contraceptives were developed that contain varying levels of progestin and estrogen throughout the standard 21-day cycle. Emergency contraception, a high dose of an oral contraceptive used by women after intercourse to prevent unwanted pregnancy, became more accessible beginning in the mid-1990s. In 2006, the FDA approved one option, Plan B, for use without a prescription for women aged 18 and older. Current oral contraceptives typically contain less than 1/10 the amount of progestin and 1/4 of the estrogen as found in the early versions.

Mechanism of Action

The hormones in combined oral contraceptives suppress both follicular development and ovulation. They also alter cervical mucus to make it more hostile toward sperm in case ovulation does occur. Progestin-only pills work by reducing and thickening cervical mucus to prevent sperm from reaching an egg. This type also inhibits the thickening of the uterine lining, which prevents a fertilized egg from implantation in the uterus.

The exact mechanism of action with emergency contraceptives is uncertain and may depend on the time in a woman's cycle that they are used. If taken at the beginning of a cycle, they may prevent or delay ovulation. If ovulation has already occurred, they may interfere with fertilization or implantation. The effectiveness of emergency contraceptives decreases as the length of time after unprotected intercourse increases.

Benefits

The most significant benefit of oral contraceptives is the decreased risk of pregnancy and pregnancy-related complications, including ectopic pregnancy. Among "perfect" users who do not miss any pills, approximately 1 woman in 1,000 become pregnant during the first year of use. Typical users have a pregnancy rate of 60 to 80 per 1,000 women during the first year. A World Health Organization (WHO) study found no significant difference in effectiveness when comparing six brands of combined pills. Progestin-only pills have slightly lower effectiveness.

Non-pregnancy-related benefits include less iron deficiency anemia due to lighter menstrual bleeding, more regular menstrual cycles, less severe premenstrual symptoms, and less dysmenorrhea. Oral contraceptives reduce the risk of epithelial ovarian cancer and endometrial cancer and prevent ovarian cysts. They may also protect against bone density loss, benign breast disease, pelvic inflammatory disease, and colorectal cancer.

Health Risks

Older women who use oral contraceptives and have hypertension or smoke have an increased risk of heart attack and stroke. A multicountry study by the WHO found that the relative risk for myocardial infarction increases substantially among users who also have other risk factors for heart disease such as smoking, hypertension, and diabetes. Among women without these additional risk factors, the study found no increased risk among current or past users.

The risk for other circulatory diseases, particularly venous thromboembolism, is also increased. Several studies have found a small increase in blood pressure among users, and one cohort study of 68,000 nurses in the United States reported that users were twice as likely to develop hypertension compared with nonusers, though this risk decreased after discontinuation. Additional health risks include the accelerated development of gallbladder disease among already susceptible women and rare, noncancerous liver tumors.

The relationship between oral contraceptives and the development of cervical cancer remains uncertain. While recent studies have found an increased association of cervical cancer and preinvasive lesions, the epidemiologic evidence cannot conclusively determine if this is due to biological or behavioral factors. Women

also have an elevated risk of diagnosis of early occurring breast cancer, which may be partially attributable to more frequent breast exams. The risks of both cervical neoplasia and breast cancer disappear within 10 years after discontinuation of use.

Common side effects of oral contraceptives include nausea, breast tenderness, irregular bleeding, and depression. These often subside after several months of use.

—Martha Decker

See also Reproductive Epidemiology; Women's Health Issues

Further Readings

ACOG Practice Bulletin. (2005). Clinical Management Guidelines for Obstetrician-Gynecologists, Number 69, December 2005 (replaces Practice Bulletin Number 25, March 2001). Emergency contraception. *Obstetrics and Gynecology*, *106*(6), 1443–1452.

Asbell, B. (1995). *The pill: A biography of the drug that changed the world.* New York: Random House.

Beral, V., Hermon, C., Kay, C., Hannaford, P., Darby, S., & Reeves, G. (1999). Mortality associated with oral contraceptive use: 25 year follow up of cohort of 46,000 women from Royal College of General Practitioners' oral contraception study. *British Medical Journal*, *318*(7176), 96–100.

Blackburn, R. D., Cunkelman, J. A., & Zlidar, V. M. (2000). *Oral contraceptives: An update. Population Reports*, Series A, No. 9. Baltimore: Johns Hopkins University School of Public Health, Population Information Program.

Chasan-Taber, L., Willett, W. C., Manson, J. E., Spiegelman, D., Hunter, D. J., Curhan, G., et al. (1996). Prospective study of oral contraceptives and hypertension among women in the United States. *Circulation*, *94*(3), 483–489.

Connell, E. B. (2002). *The contraception sourcebook.* New York: McGraw-Hill.

Gazit, C. (2002). *The pill: American experience.* Boston: WGBH. Retrieved October 1, 2006, from http://www.pbs.org/wgbh/amex/pill.

World Health Organization. (1997). Acute myocardial infarction and combined oral contraceptives: Results of an international multicentre case-control study. *Lancet*, *349*(9060), 1202–1209.

Oral Health

Oral health conditions include a number of congenital or developmental anomalies, such as clefts and tumors. But the most common oral health problem is tooth decay, known medically as dental caries. Although dental caries is largely preventable, more than half of all adults above age 18 present early signs of this disease and, at some point in life, about three out of four adults will develop the disease. In older adults, tooth decay and periodontal disease are the leading causes of tooth loss. Tooth decay is also common among children as young as 5 years and remains the most common chronic disease of children aged 5 to 17 years. It is estimated that tooth decay is four times more prevalent than asthma in childhood. Poor oral health has been related to decreased school performance, poor social relationships, and less success later in life. It is estimated that about 51 million school hours per year are lost in the United States alone because of dental-related illness. In adults, dental caries and, eventually, tooth loss can reduce chewing ability that leads to detrimental changes in food selection. This, in turn, may increase the risk of particular systemic diseases such as cardiovascular diseases.

Despite numerous epidemiologic studies currently available to assess the pattern of dental caries in oral health, there are still some fundamental questions that remain unanswered. As an example, many dentists believe that there are spatial symmetries in the mouth with respect to caries development, but this has never been demonstrated statistically. Although many dental studies provide detailed tooth-level data on caries activity, most analyses still rely on the decayed, missing, and filled (DMF) index introduced in the 1930s. This approach operates at the mouth level, that is, it counts the number of decayed teeth within a person's mouth without including information on which specific teeth are decayed; therefore, it may not be relevant in assessing the spatial distribution of the dental caries in a mouth. A more useful way to study oral health may be to collect data at the tooth level. Models that summarize tooth-level data can be used to answer questions related to tooth surface susceptibility to caries experience, symmetries in the mouth with respect to dental caries, and differences in surface susceptibility according to caries risk groups. Use of tooth-level models should improve our ability to analyze and interpret complex dependent data generated from clinical trials and epidemiologic studies in dental research.

Tooth Decay and Measurement

Tooth decay is ubiquitous and is one of the most prevalent oral diseases. It is a localized, progressive

demineralization of the hard tissues of the crown (coronal enamel, dentine) and root (cementum, dentine) surfaces of teeth. The demineralization is caused by acids produced by bacteria, particularly *mutans Streptococci* and possibly *Lactobacilli* that ferment dietary carbohydrates. This occurs within a bacteria-laden gelatinous material called dental plaque that adheres to tooth surfaces and becomes colonized by bacteria. Thus, caries results from the interplay of three main factors over time: dietary carbohydrates, cariogenic bacteria within dental plaque, and susceptible hard tooth surfaces. Dental caries is a dynamic process since periods of demineralization alternate with periods of remineralization through the action of fluoride, calcium, and phosphorous contained in oral fluids.

According to the World Health Organization, both the shape and the depth of a carious lesion at the tooth level can be scored on a 4-point scale, D_1 to D_4. Level D_1 refers to clinically detectable enamel lesions with noncavitated surfaces; D_2 to clinically detectable cavities limited to the enamel; D_3 to clinically detectable lesions in dentin; and, finally, D_4 to lesions into the pulp. The threshold traditionally used in epidemiologic studies is based on D_3, which ignores all signs of lesions less severe than clinically detectable lesions. In the past, it has been argued that enamel lesions could not be included in epidemiologic studies because the reproducibility achieved was lower and insufficient. However, as reported by Marisol Tellez, studies have shown excellent reproducibility with kappa coefficients around .80 using the D_1 threshold in survey settings as well as clinical trials.

Despite these detailed tooth-level data, the methods for analyzing dental caries outcomes in human populations still rely on the calculation of the DMF index, developed in the 1930s. This index is applied to all the teeth (DMFT) or to all surfaces (DMFS). These scores are typically analyzed as counts using Poisson models or negative binomial models to account for overdispersion as a result of mixtures in the data. Other approaches consist of dichotomizing or constructing ordered levels according to a graded scale using some threshold values. Despite an extensive use of these indexes and related models, they have some recognized limitations. These scores are not very informative in studying tooth-specific problems. Different types of teeth (incisors, canine, and molars in primary dentition) and tooth surfaces (facial, lingual, occlusal, mesial, distal, and incisal surfaces) are not equally susceptible to dental caries. For example, the different morphology of the pit-and-fissure surfaces of teeth makes them more susceptible to decay than the smooth surfaces. Thus, it is no surprise to find that the posterior molar and premolar teeth that have pit-and-fissure surfaces are more susceptible than the anterior teeth. The application of the traditional DMF index to the skewed data on caries that frequently emerge today is one of the factors contributing to the underestimation of the prevalence of caries and the overestimation of the temporal change. Thus, it places limitations on the population strategy to be used in caries prevention, and it contributes to a lack of discrimination between individuals with differences in caries activities. A tooth- and surface-level analysis can give a better understanding of the pattern of dental caries over time.

In addition, most research on dental caries is based on cross-sectional studies or surveillance data. For example, G. D. Slade and colleagues reported a cross-sectional dental study involving 9,690 Australian schoolchildren aged 5 to 15 years. However, like most caries surveys, this study is limited in its ability to examine questions involving duration of time, for example, the period between tooth eruption and caries development, and to investigate the interrelationship among teeth and tooth surfaces within the mouth in terms of caries development. Yet most interesting scientific questions in dentistry deal with the development of caries over time, determinants of oral hygiene in general, and the intra-oral distribution of the disease in the mouth across time. Therefore, in addition to cross-sectional studies and surveillance data, these questions beg for high-quality longitudinal studies. Most of the inconsistencies in the dental oral literature may be the result of the inability of cross-sectional studies to accurately describe the etiology of dental caries.

Statistical Models for Highly Correlated Dental Data

A tooth-surface and tooth-level analysis in dental research poses a number of difficulties due to inherent spatial association of tooth surfaces and teeth in the mouth. This then necessitates the use of multivariate methods for correlated data. These multivariate models consist of (1) a parameter vector that accounts for the effects the independent variables have on the location parameter of each tooth-surface and tooth-level

outcome and (2) an association component that corrects these location parameters for potential correlation among all outcomes from the same mouth. When the emergence times of dental caries of different teeth are to be compared, a multivariate model is needed to accommodate the multiplicity of outcomes from the same mouth. A popular model choice would be frailty models, which are basically random effects survival models. These models are enormously popular in the analysis of clustered time-to-event data. Such models account for correlations between survival rates in neighboring spatial regions such as tooth surfaces. In longitudinal studies that aim at studying the dynamics of caries across time, another important issue is the serial association induced by measurements taken on the same tooth or tooth surface over time. A good model for the covariance function of a stationary process in space and time should then accurately describe the variances and correlations of all linear combinations of the processes. In particular, it does not suffice to find a model that describes the purely temporal covariances and the purely spatial covariances accurately. Rather, it is critical to capture the spatiotemporal interactions as well. A great alternative to this class of models is the generalized estimating equations methodology where independent models are fitted to each tooth or tooth-surface outcome, and a sandwich estimator is used to adjust the model estimates' standard errors for association in the data.

In analyzing survival data in dental research, one is typically faced with the problem that the exact time of surface and tooth decay is typically unknown. This then imposes constraints on the model, rendering the approach based on the events (surface decay) themselves somewhat problematic. A simplistic approach to address these issues is to consider a survival approach where the outcome of interest is the emergence time of surface or tooth decay, which is an interval-censored observation. Compared with right-censored data, techniques for interval-censored data are less well developed, and their statistical properties are much more complex. For example, the convergence rate of nonparametric estimates of the survival curve for interval-censored data is known to be smaller than that of right-censored data. This then invalidates the use of the Wald approach to compute confidence intervals of the survival curve for interval-censored data. There is a common consensus from the current literature to use parametric models such as accelerated failure time models for interval-censored data. The widespread use of nonparametric methods for right-censored survival data has limited the use of parametric models with interval-censored data.

Future Research

Methods of identifying early carious lesions accurately and of identifying children at a high risk of dental caries are required. Also needed are studies to confirm the relation between the vulnerability of occlusal surfaces to caries and the time since tooth eruption. Prospective studies to examine all possible factors associated with nursing caries are also needed. For this, it is essential that we have effective and rigorous statistical methods to understand the etiology of dental caries. For this, it is recommended that regression models for spatial-temporal change data and models for multivariate failure time data with applications to caries on tooth surfaces be developed.

The use and acceptance of statistical models in dental research requires reliable and user-friendly software, readily available to perform regression analysis routinely. The software should be time-efficient, well-documented, and, most important, must have a friendly interface.

—David Todem

See also Child and Adolescent Health; Longitudinal Research Design; Regression; Study Design; Survival Analysis

Further Readings

Bogaerts, K., Leroy, R., Lesaffre, E., & Declerck, D. (2002). Modelling tooth emergence data based on multivariate interval-censored data. *Statistics in Medicine*, *21*, 3775–3787.

Hougaard, P. (2000). *Analysis of multivariate survival data.* New York: Springer.

Slade, G. D., Spencer, A. J, Davies, M. J., & Stewart, J. F. (1996). Caries experience among children in fluoridated Townsville and unfluoridated Brisbane. *Australian and New Zealand Journal of Public Health Dentistry*, *20*, 623–629.

Tellez, M., Sohn, W., Burt, B. A., & Ismail, A. I. (2006). Assessment of the relationship between neighborhood characteristics and dental caries severity among low-income African-Americans: A multilevel approach. *Journal of Public Health Dentistry*, *66*, 30–36.

Organ Donation

Organ donation has become the treatment of choice for end-stage renal disease (ESRD) in addition to other types of organ failure, notably heart, liver, and lung. By all accounts, the shortage of transplantable organs is a public health crisis with one person on the United Network of Organ sharing (UNOS) transplant waiting list dying approximately every 17 min. In 2005, there were more than 90,000 individuals awaiting transplantation.

There are two possible sources of organs for transplant: deceased organ donation that has provided the major source of transplantable organs and living donation usually, but not always, from the families of waiting recipients. Deceased donors are the only feasible source of heart donation and are by far the single most important source of livers, lungs, intestinal organs, and pancreata. Most living donation involves kidneys (92%) or liver segments (8%).

The number of deceased and living donors for all organs was 14,491 in 2004, with 7,593 deceased and 6,898 living donors. The number of all organ donors has increased at an average rate of 7% per year. Although the increase in living donors has been a major contributor in helping ameliorate the organ donor shortage in the United States, the numbers of persons on the waiting list is growing faster, with a net increase in the waiting list of 11% per year.

The other major change in the donor pool has been an inclusion of older donors and a change in the cause of death of donors. The average age of deceased donors rose by 2.1 years between 1996 and 2001 and is now in the mid-30s. Living donors are, on average, a year older than deceased donors. Living donors are more likely to be female (approximately 58%), while deceased donors are more frequently male (60%). A total of 79% to 82% of living donors are white, and deceased donors are also predominantly white (85%). Total minority donations increased by 56% from 1992 to 2001, while the number of white organ donors increased by 32% over the same period.

These trends in the organ donor profile reflect the continued shift away from the young adult who dies from a traumatic head injury to the older adult who dies from a cerebrovascular event. The progressive increase in the median age of deceased donors over the past 10 years has exceeded that of the general population since 1996.

In 2001, there were 695 donations resulting from anoxic brain deaths, up 12% from 2000 and up by 32% since 1995—the fastest rise among the causes of death for deceased donors. The rise in anoxic deaths resulted primarily from the increased frequency of drowning, drug intoxication, and cardiovascular events. Cerebrovascular deaths continue to lead as the primary cause for deceased donations (43% of all deceased donors in 2001).

Consent to organ donation by families of brain-dead patients has been a major barrier to maximizing the numbers of solid organs available for transplant in the United States. Despite public opinion polls reporting that more than 85% of the American public is willing to donate, fewer than half choose to donate a family member's organs when asked.

The Uniform Anatomical Gift Act (UAGA), drafted by the National Conference of Commissioners on Uniform State Laws in 1968 and modified in 1987, regulates organ donation in the United States. By 1973, it had been passed by all 50 states. Aimed at enabling individuals or their families to donate organs, UAGA also served to establish altruism and voluntariness as the bedrock of organ donation and procurement in the United States, while outlawing the sale of organs. This law recognizes the rights of individuals to donate by means of an organ donor card and gives the immediate family of a deceased person the option to donate. In 1973, the End-Stage Renal Disease Program provided federal financial support for organ transplantation by funding 100% of organ procurement costs through Medicare. Federal organization and oversight of organ procurement were further developed in 1984, when Congress passed the National Organ Transplantation Act (NOTA). This law created the Organ Procurement Transplant Network (OPTN), which has the responsibility for setting standards and rules regarding the distribution of human organs procured in this country and also outlaws the sale of organs.

Two key factors are responsible for the critical shortage of transplantable solid organs in the United States. First, it has been estimated that no more than 15,000 deceased brain-dead donors are available each year in the United States. In 2005, more than 90,000 individuals were waiting to receive a transplant. Second, the rate of consent for organ donation by next of kin has limited the number of organs available for transplant. On average, no more than 50% of those families from whom donation is requested agree to

donate. Increases in the total number of organs procured have resulted largely from an expansion of the donor pool (e.g., accepting older patients as donors) and from improvements in procedures for referring and requesting organ donation from families of potential donor patients (living donation).

Major legislative efforts to encourage the donation of organs have been undertaken since the 1970s. In 1986, Health Care Financing Administration (HCFA) made it mandatory that hospitals request organ donation from donor-eligible families, and the Joint Commission on Accreditation of Health Care Organizations (JCAHO) made it a requirement for hospital accreditation. These laws and regulations were not effective in improving consent rates. Whereas surveys show that 99% of Americans are aware of transplantation, and at least 85% say they would donate their organs if asked, rates of consent to requests made for deceased patients' organs continue to hover at 50%.

In 1998, HCFA required that hospitals notify their local organ procurement organization (OPO) about all deaths and imminent deaths, and that families must be approached about donation in collaboration with the local OPO. Underlying this regulation (known as "required referral" or "routine notification") was the premise that health professionals alone were not effectively communicating with families about donation. This regulation, too, has had little impact on actual rates of consent to donation, although some regions have seen an increase in numbers of organs procured. A new legislative effort, termed *donor designation* or *first-person consent*, now makes it possible for donation to occur without family permission if the deceased has a valid donor card or driver's license designation.

Recent studies have emphasized the importance of the process of asking for organ donation. This process entails the identification of donation-eligible patients and then making a request. It is first necessary to identify that someone is a potential organ donor. Until recently, this process was almost completely in the hands of hospital health care providers. Data showed that the ability of health care providers to recognize a donor was variable, ranging from 70% to 100%. To address this problem, the 1998 HCFA regulations required that the local OPO be called about each hospital death. Data indicate that referral and request rates also vary widely, ranging from 65% to 99%. Referral rates average 80%, and requests are made in 84% of cases.

Different practices of discussing and obtaining consent from families have been widely debated and are the subject of some controversy. Factors such as when the request should be made, who should request organ donation, what should be discussed with the family, and how (or if) families who initially refuse organ donation should be reapproached have all received attention. Some strategies, however, have not proven fruitful or have not been confirmed. For example, studies of timing of the donation request conducted in the early 1990s suggested that "decoupling" the request from pronouncement of death would create a significant rise in consent rates. However, studies that are more recent have revealed that the issue is more complex, and that raising the issue of organ donation with families earlier in the course of the patient's hospitalization—especially once the futility of treatment has been determined—may be the most useful practice.

Families often refuse to consent to organ donation because they are concerned about mutilation of the body. A recent study found that families were more likely to donate when this issue was discussed openly rather than avoided. Additionally, spending more time with families and discussing specific issues about organ donation, such as the patient's wishes concerning donation, the choice concerning what organs to donate, and the fact that there are no costs associated with donation, are significantly associated with consent to donation. Families who spent more time and discussed more donation-related issues are five times more likely to donate.

The 1998 regulations also sought to guarantee that experienced requesters speak with families. Again, recent data indicate that this will be a fruitful strategy if it can be successfully implemented. For example, an earlier study found that health care providers who rated themselves as more uncomfortable speaking with families about organ donation were less likely to obtain consent than those who reported themselves as comfortable with discussing the topic and answering the family's questions.

Legislative efforts have yet to close the gap between donor potential and organs procured. Studies now indicate that the process itself is of critical importance. Appropriate training and hospital donation development are needed to improve performance in the procurement of organs from deceased donors. Rates of living donation continue to rise in the United States, contributing to the availability of

transplants for patients in need of kidney and liver transplantation.

—Laura A. Siminoff

See also Ethics in Health Care; Governmental Role in Public Health; Health Care Delivery; Health Communication; Informed Consent

Further Readings

Boulware, L. E., Ratner, L. E., Cooper, L. A., Sosa, J. A., LaVeist, T. A., Powe, N. R. (2002). Understanding disparities in donor behavior: Race and gender differences in willingness to donate blood and cadaveric organs. *Medical Care*, *40*(2), 85–95.

Gallup Organization. (1993). *The American public's attitudes toward organ donation and transplantation.* Boston: Partnership for Organ Donation.

Institute of Medicine. (2006). *Organ donation: Opportunities for Action.* Washington, DC: Author.

Sheehy, E., Conrad, S. L., Brigham, L. E., Luskin, R., Weber, P., Eakin, M., et al. (2003. Estimating the number of potential organ donors in the United States. *New England Journal of Medicine*, *349*, 667–674.

Siminoff, L. A., Gordon, N., Hewlett, J., & Arnold, R. M. (2001). Factors influencing families' consent for donation of solid organs for transplantation. *Journal of the American Medical Association*, *286*(1), 71–77.

Siminoff, L. A., Lawrence, R. H., & Arnold, R. M. (2003). Comparison of black and white families' experiences and perceptions regarding organ donation requests. *Critical Care Medicine*, *31*(1), 146–151.

Siminoff, L. A., & Mercer, M. B. (2001). Public policy, public opinion, and consent for organ donation. *Cambridge Quarterly of Healthcare Ethics*, *10*, 377–386.

Sque, M., Long, T., & Payne, S. (2005). Organ donation: Key factors influencing families' decision-making. *Transplant Proceedings*, *37*, 543–546.

Osteoporosis

Osteoporosis is a skeletal disorder characterized by low bone mass and deterioration of bone quality leading to bone fragility and increased risk of fracture. More than 10 million Americans above the age of 50 have osteoporosis. Women are two to three times as likely to be affected as men. Osteoporotic fractures lead to significant morbidity and excess mortality; of those experiencing a hip fracture, one out of five will die in the following year and less than one out of three will regain prefracture physical ability. The direct and indirect costs associated with all osteoporotic fractures in the United States exceed $20 billion dollars annually.

History

The term *osteoporosis*—Greek for porous bone—was first coined in the 1830s by French pathologist Georges Chretien Frederic Martin Lobstein in describing the larger than normal holes he noticed in some bones. At the time of the observation, no further significance was associated with the condition. A century later, in 1934, Yale anatomists discovered estrogen's role in bone formation while studying factors causing increased bone mass in female pigeons as compared with males. With the pigeon study in mind, six years later, Fuller Albright, a Massachusetts General Hospital physician, proposed estrogen to be important in controlling calcium concentrations in human bone and suggested that estrogen-related bone loss was responsible for the fractures afflicting his older women patients.

Etiology

Primary, *age-related osteoporosis* is the most common form of the disease, resulting largely from age- and hormone-related decreases in bone quality. Risk factors for osteoporosis and related fractures include physical inactivity, previous fractures, smoking, low body weight, low exposure to sunlight (in people above 50), tendency to fall, alcohol consumption, and impaired vision. Preliminary research also suggests an important role for genetic predisposition in the development of osteoporosis. Reduced bone quality due to diseases, disorders, medication use, and/or toxin exposure is referred to as secondary osteoporosis. Many who are diagnosed with osteoporosis have had exposures to secondary causes occurring earlier in life.

Pathogenesis

Bone is a living matrix comprised largely of crystals of calcium and phosphate, or hydroxyapatite, and the protein collagen. Throughout life, bones change in shape, size, and position through the processes of modeling and remodeling. Modeling allows bones to grow and shift in space by forming new bone at one site while removing it from another within the same

bone. Remodeling is the removal and replacement of bone at the same site. Remodeling, which becomes the dominant process in both sexes around 20 years of age, is important in the repair of microdamage resulting from stresses on the skeleton, in the replacement of older bone tissue, and in the bioavailability of calcium and phosphate stored in bone for use in other bodily functions. Estrogen influences bone quality by directly and indirectly affecting the activity of osteoclasts, cells that break bone down, and osteoblasts, cells that build bone.

During menopause, a drastic reduction in estrogen production by the ovaries leads to remodeling imbalance and subsequent loss of bone mineral density (BMD). The sudden decrease in estrogen triggers an increase in osteoclast formation and recruitment, inhibits osteoclast apoptosis, and promotes osteoblast apoptosis. This acute phase of rapid bone loss lasts for 4 to 8 years.

Following the acute phase, a period of slow, continuous bone loss progresses throughout the rest of life. Age-related retarded bone formation, decreased calcium and vitamin D intake, decreased physical activity, and the loss of estrogen's effects on calcium absorption in the intestines and calcium conservation in the kidneys contribute to slow bone loss. Reduced dietary calcium intake and absorption further the disease process by increasing parathyroid hormone levels leading to calcium removal from the bones for use in other systems. Thinning of bone in itself is not a significant cause of morbidity and mortality; weakened bones are, however, more likely to suffer damage following trauma.

With aging, men suffer only a slow reduction in sex hormones and therefore osteoporosis develops through a slow, continuous process of bone loss similar to that in women. Up to 50% of elderly men are deficient in active sex steroids.

The disease processes associated with secondary osteoporosis are varied and include genetic disorders (e.g., cystic fibrosis), conditions leading to estrogen or testosterone deficiencies in adolescence or following development of peak bone density (e.g., anorexia nervosa, athletic amenorrhea, Turner's syndrome), excess production of or treatment with thyroid hormone or glucocorticoids (e.g., thyrotoxosis, Cushing's syndrome), diseases or conditions leading to reduced intestinal absorption of calcium or phosphate (e.g., celiac disease), conditions that disrupt vitamin D metabolism (e.g., cirrhosis due to hepatitis B or C), and rheumatic disorders (e.g., Lupus, rheumatoid arthritis).

Impact

While osteoporosis is associated with an increased risk for all types of fractures, the most common sites are the hip, spine, and wrist. Fractures, especially of the hip, can be painful and disfiguring, and often require hospitalization; most never regain prefracture physical functioning. Almost 10% of those experiencing any fracture and more than a quarter of those who suffer a hip fracture are physically impaired and unable to live without assistance. Fractures also have indirect consequences such as disrupted abdominal anatomy leading to constipation, distention, and reduced appetite. Psychological effects, including depression, anxiety, fear, and strained interpersonal relationships are common sequelae. Both hip and spine fractures are associated with excess mortality; one in five persons experiencing a hip fracture dies within the subsequent year due to comorbidities.

Epidemiology

Rates of osteoporosis and osteoporosis-related fractures vary by ethnicity, gender, and geography. Men, with higher peak bone densities than women, are less likely to suffer a primary osteoporotic fracture. In the United States, among Caucasians, 50% of women and 20% of men older than 50 will experience a fragility fracture in their remaining lifetime. Women are more likely to suffer hip and wrist fractures. Rates of vertebral fractures, which are more often associated with bending and lifting objects than with falls, are more similar between men and women.

Because of the lack of uniform availability of testing procedures, the incidence of hip fracture is often used as an international index for osteoporosis. Possibly due to genetic predisposition and/or environmental factors, fractures, especially those of the hip, are more likely to occur in Caucasian than in non-Caucasian populations. Generally, osteoporosis rates are highest in North American and European countries and lowest in African countries. By the year 2050, Asian populations are expected to account for more than half of all hip fractures. Fracture rates also vary within ethnic groups; for example, controlling for race/ethnicity, hip fracture rates are higher in urban areas. Secondary osteoporosis strikes both young and old and men and women equally.

Diagnosis

Osteoporosis is diagnosed by assessing BMD. While other bone parameters such as shape, geometry, type of bone (trabecular or cortical), degree of mineralization, microdamage accumulation, and rate of bone turnover are important in determining bone strength, these measures have yet to be incorporated into standard testing techniques. The World Health Organization has identified the following risk factors that, when measured in addition to BMD, give a more accurate prediction of future fracture risk: age, existence of previous fractures, glucocorticoid use, cigarette smoking, heavy alcohol consumption, and female low body weight.

Dual-energy X-ray absorptiometry is used to measure overall BMD and is the gold standard for osteoporosis diagnosis. According to the World Health Organization, someone with a BMD 2.5 or more standard deviations below that of the young adult mean is considered to have osteoporosis. Osteopenia, a precursor to osteoporosis, is diagnosed when BMD is between 1 and 2.5 standard deviations below the mean. Any fracture resulting from low trauma or due to a fall from standing height is also diagnostic of osteoporosis, regardless of BMD.

Prevention

Prevention of osteoporosis generally depends on increasing bone density and reducing the risk of falling, through exercises including walking, aerobics, weight bearing and resistance exercises, and balance training. Studies have shown reduced risk of fracture with smoking cessation and calcium supplementation. Pharmacologic treatments, such as bisphosphonates and hormone replacement therapy, have proven effective in increasing BMD and decreasing the risk of fractures, though recently characterized health risks associated with the latter have resulted in the discouragement of its use.

—Michelle Kirian

See also Aging, Epidemiology of; Injury Epidemiology; Vitamin Deficiency Diseases; Women's Health Issues

Further Readings

Fitzpatrick, L. A. (2002). Secondary causes of osteoporosis. *Mayo Clinic Proceedings*, *77*, 453–468.

Johnell, O., & Hertzman, P. (2006). *What evidence is there for the prevention and screening of osteoporosis?* Health Evidence Network report. Copenhagen, Denmark: WHO Regional Office for Europe.

Patlak, M. (2001). Bone builders: The discoveries behind preventing and treating osteoporosis. *FASEB Journal*, *10*, 1677E.

Sambrooke, P., & Cooper, C. (2006). Osteoporosis. *Lancet*, *367*, 2010–2018.

U.S. Department of Health and Human Services, Office of the Surgeon General. (2004). *Bone health and osteoporosis: A report of the Surgeon General.* Rockville, MD: Author.

Weitzmann, M. N., & Pacifici, R. (2006). Estrogen deficiency and bone loss: An inflamatory tale. *Journal of Clinical Investigation*, *116*(5), 1186–1194.

World Health Organization. (2003). *Prevention and management of osteoporosis.* WHO Technical Report Series 921. Geneva, Switzerland: Author.

Outbreak Investigation

Outbreak investigations are a subgroup of epidemiologic studies called "field investigations." When the numbers of persons affected by a particular disease, usually infectious, exceeds the number of cases expected in a given place during a given time period, it may be said that there is an outbreak of that disease. Epidemiologists may then conduct targeted investigations to (a) determine the cause and etiology of the disease, (b) to limit the spread and severity of illness of the disease, and (c) to prevent future outbreaks. In addition, investigations of this sort can serve to identify new modes of transmission of illnesses, identify new pathogens, and monitor the effectiveness of prevention activities. Collectively, these activities make up an outbreak investigation. Investigations of this type require epidemiologists to seek out and collect information (via interviews, lab studies, etc.) from persons affected by the disease.

Though the precise order may vary, most outbreak investigations share many steps and tasks in common. First, investigators must determine that an outbreak actually exists. The presence of a potential outbreak can be detected by any of several sources, including health care workers, laboratory workers, the general public, formal disease surveillance systems, or other health data. Investigators then compare the numbers of currently observed cases with historical data for the similar time period in previous

years to determine that the observed cases represent an actual outbreak. Several sources of data adequate for this sort of comparison exist and may include disease surveillance records, birth certificates, death certificates, hospital discharge information, and so on. Changes in observed numbers of cases may be due to reasons other than the presence of an outbreak, for instance if there was an underlying change in case ascertainment or a change in the population at risk for the disease.

In addition to confirming the presence of an outbreak, investigators must also confirm the diagnosis and generate a case definition. The case definition, ideally consisting of the simplest, most concrete criteria possible, will help investigators and health care workers identify persons to be included in the investigation, and may be refined as new information about the illness and at-risk populations come to hand. Interviews with or surveys of cases will help place each case within the epidemiologic triad of person, place, and time, and can help refine ideas and initial hypotheses regarding who is at risk as well as beginning to address the issues of how and why the outbreak began. In an iterative process, investigators will continue to refine their case definition and explanatory hypotheses as new information comes to hand via interviews and surveys.

Once some of this preliminary descriptive work has been completed, investigators will plan and conduct an analytic study to further identify the source of the infection. Most commonly, case-control study designs are used, though cohort studies can also be used. As in more controlled study settings, both study designs have their advantages and disadvantages. Case-control studies are often used in the context of a large outbreak, where relative efficiency in both time and cost is important. In addition, in many outbreaks, the entire cohort is often not clearly defined, making a case-control study approach more appropriate. Case-control studies may also be nested within larger cohort studies, where testing a specific hypothesis on the entire larger cohort is not feasible. Cohort study designs have the advantages that investigators can evaluate multiple disease outcomes and can directly measure attack rates.

In an outbreak investigation, an investigator's work is not completed with the identification of the source of the outbreak. Investigators must also prepare a written report summarizing what they have learned about the outbreak, to guide future control and prevention efforts. Investigators must also implement those control and prevention methods, usually with some combination of (a) eliminating the source of the pathogen, (b) interrupting the spread from person to person, and/or (c) protecting individuals from consequences of exposure with methods such as vaccination, prophylactic medication, and so on.

Outbreak investigations differ from other epidemiologic studies in several ways, and they present unique challenges as well. The problem necessitating an outbreak investigation is usually unexpected and, thus, rapid response and immediate epidemiologist presence are required. The need for timely intervention also often means that outbreak investigations are more limited than other investigations. The aim is to conduct studies as scientifically rigorous as possible within these constraints. As with other studies, outbreak investigations are vulnerable to various kinds of bias, though perhaps in ways unique to these investigations. For example, most epidemiologic studies are vulnerable to sampling biases, but in an outbreak investigation, some or many of the people involved may be highly motivated to not cooperate with investigation for reasons such as protecting their own financial interests or reputation. In addition, cases may be distributed over a wide geographic area, making full characterization of the cases difficult. Outbreaks may also occur in areas with poor infrastructure, making optimal investigation difficult. In the instance of a relatively small outbreak, sample size may not be sufficient to detect small effects. And publicity about the outbreak or investigation may serve to communicate preconceived notions about the outbreak, resulting in biased information and potentially erroneous conclusions.

—Annette L. Adams

See also Epidemic; Field Epidemiology; Natural Experiment

Further Readings

Brownson, R. C. (1998). Outbreak and cluster investigations. In R. C. Brownson & D. B. Petitti (Eds.), *Applied epidemiology: Theory to practice* (pp. 71–104). New York: Oxford University Press.

Koehler, J., & Duchin, J. (2003). Outbreak investigation. In T. D. Koepsell & N. S. Weiss (Eds.), *Epidemiologic methods: Studying the occurrence of illness* (pp. 464–489). New York: Oxford University Press.

Overmatching

To control for potential confounders or to enhance stratified analysis in observational studies, researchers may choose to match cases and controls or exposed and unexposed subjects on characteristics of interest. If matching is superfluous or erroneous, overmatching may occur. The three main effects of overmatching are a loss of statistical efficiency, introduction of bias, and loss of financial efficiency.

Background

To reduce confounding or to enhance stratified analysis, unexposed subjects in cohort studies or controls in case-control studies may be chosen to be identical or similar to exposed subjects or cases with respect to the distribution of one or more variables. Overmatching, sometimes referred to as overmatching bias, occurs when matching is done incorrectly or unnecessarily leading to reduced efficiency and biased results. Overmatching generally affects case-control studies.

Effects of Overmatching

Loss of Statistical Efficiency

In case-control studies, if cases and controls are matched on a variable that is associated to the exposure but not the disease, chosen controls are more similar to cases than the base population in respect to the exposure. The forced similarity between cases and controls in respect to the exposure obscures the relationship between the exposure and the disease. Matching on an exposure-associated variable will cause the crude odds ratio to be closer to 1—that is, to the null value. However, when stratified by the matching variable, stratum-specific odds ratios will be unbiased. If confounding is present, bias due to matching on an exposure-associated variable will cause the odds ratio to go toward the null regardless of the direction of the confounding. The degree of information loss due to overmatching depends on the absolute correlation between the matching variable and the exposure of interest. Matching on a nonconfounder necessitates stratified analysis that would otherwise not be necessary, and it reduces study efficiency.

Introduction of Bias

If controls are matched to cases on a variable that is affected by both the exposure and the disease or is an intermediate between exposure and disease, both the crude and adjusted odds ratios will be biased. Like matching on exposure-only-associated variables, matching on an intermediate or variable affected by exposure and disease will force the odds ratios toward the null. However, unlike matching on an exposure-only-associated variable, it is not possible to get unbiased stratified measures of effect.

Loss of Financial Efficiency

Matching can lead to greater statistical efficiency by ensuring that cases will have one or more matched controls for comparison in stratified analysis. Also, matching may offer a cost benefit if the collection of exposure data from many people is very expensive. However, if the matching process is complicated and involves many matching variables, it may be difficult and costly to identify and recruit potential controls. Also, if matching is done unnecessarily, additional costs associated with recruiting further controls may incur. Potential statistical benefits and costs should be assessed prior to matching.

—Michelle Kirian

See also Bias; Confounding; Matching; Study Design

Further Readings

Agudo, C., & Gonzales, C. A. (1999). Secondary matching: A method for selecting controls in case control studies on environmental risk factors. *International Journal of Epidemiology*, *28*, 1130–1133.

Rothman, K. J., & Greenland, S. (1998). Matching. In K. J. Rothman & S. Greenland (Eds.). *Modern epidemiology* (pp. 147–161). Philadelphia: Lippincott Williams & Wilkins.

Szklo, M., & Nieto, F. J. (2004). *Epidemiology: Beyond the basics.* Boston: Jones & Bartlett.

P

Pain

Pain is a complex biopsychosocial phenomenon that most human beings experience at different times throughout their life span. This phenomenon has perplexed man for centuries, and for centuries pain has been poorly managed. Today, health care professionals are placed in a position where patients rely on them to provide pain control, and effective pain management is important in enabling patients to progress in their rehabilitation and to have an improved quality of life. This entry reviews definitions of pain, summarizes the demographics and epidemiology of pain, and describes the physiology of pain and its categorization. It also considers the interventions available, as well as some of the ethical issues that arise with respect to pain treatment.

Ancient civilizations, including early Mesopotamia, Egypt, China, Greece, and Rome, used various primitive approaches to treat pain. Via writings, carvings, and other documents, anthropologists have found evidence of pain interventions, including ancient pharmacopoeia such as the use of opium, scopolamine, ephedrine, ginseng, Siberian wort, snake venom, and various other treatments. Nonpharmacological interventions included prayer, dance rituals, music, bloodletting, hydrotherapy, and other cultural remedies.

The universally accepted definition of pain, according to the International Association for the Study of Pain (IASP) and the American Pain Society (APS), is "an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage" (Mersky, 1986, p. S217). This definition does not ask for organic proof that pain exists. A second definition coined by McCaffery (1968) says, "pain is whatever the experiencing person says it is, whenever the experiencing person says it is (p. 95)." This definition points to the subjectivity of pain. Pain is experienced differently by individuals based on their cultural background, upbringing, personal values, genetics, and the meaning they attribute to pain.

Pain is a protective mechanism that provides warning of assaults or damage to or within the body; however, if left untreated, this protective mechanism can become a chronic and destructive condition. Unrelieved pain remains a critical problem in all areas of health care. The most undertreated populations include children, the elderly, minorities, and women mostly due to myths and misconceptions related to the pain mechanism. Toward the end of the 20th century and the beginning of the current century, the undertreatment of pain became increasingly evident as interested health care professionals studied this problem. Several organizations, most notably the Joint Commission on Accreditation of Health Care Organizations (JCAHO), have developed standards, publications, and guidelines for practice related to the relief of pain and care at the end of life.

Demographics and Epidemiology

Algology is the study and science of pain phenomena, and an algologist is a student, investigator, or practitioner of algology. Pain studies conducted by pain

organizations and algologists are tracked by the American Pain Foundation. According to APS, 50 million people are disabled by pain in the United States, and an estimated 9% of the U.S. adult population suffers from moderate to severe pain. Findings from studies conducted between 1996 and 2004 indicate that work to improve the relief of pain must continue.

Key findings include the following:

- More than two thirds of full-time employees (68%), the equivalent of more than 80 million full-time employees, suffer from pain-related conditions. Fourteen percent of all full-time employees—more than 17 million—took sick days in 1995 due to pain conditions, resulting in more than 50 million work days. Sixty-nine percent said they were compensated 100% of their salaries for sick days, which translates to a cost of more than $3 billion in wages for lost sick days. Furthermore, 80% of all pain sufferers have gone on short-term disability because of pain.

- One in five Americans above 60 years of age takes pain medication to control pain that lasts for 6 months or more; this represents 18% of Americans in this age group. Two out of three older Americans who take pain medications said pain still prevents them from performing routine tasks, engaging in hobbies, or doing things they enjoy.

- Despite people's use of medications, two thirds of chronic pain sufferers (13.6 million Americans) cannot perform routine tasks because of chronic pain. Portenoy breaks down the prevalence of cancer pain as follows: (a) at time of diagnosis, 30%, (b) during active treatment, 30% to 50%, and (c) with advanced disease, 70% to 90%. The APS reemphasizes the significance of undertreatment of pain as a public health problem with consequences that include increases in health care expenditures and worker absenteeism, as well as a decrease in quality of life.

Research findings by scientists and practitioners such as Bonica, Ferrell, Libeskind, Melzack, and Wall indicate that unrelieved pain can produce serious adverse immunological, psychological, and physical effects. The personal cost of unrelieved pain includes the eight Ds: depression, disease, distraction, drugs, doctor shopping, drinking, disability, and death. In recent years, there has been an increase of liability in relation to the undertreatment of pain, with successful lawsuits being waged under elder abuse laws on behalf of elderly clients whose pain was insufficiently treated.

The Agency for Health Care Research and Quality (AHRQ) affirms that "pain is a complex, physiological and subjective response with several quantifiable features, including intensity, time course, quality, impact, and personal meaning; and the single most reliable indicator of the existence and nature of pain is the patient's self-report" (p. 4). Several reasons are responsible for the undertreatment of pain, including myths and misconceptions, health professionals' lack of knowledge, patients' fears of addiction and overtreatment, and practitioners' fear of regulatory agencies.

In 2000, California recognized pain as a fifth vital sign. As a result, all health care agencies are required to include comfort assessment, and all medical schools are required to include pain management in their curriculum. It is believed that other states will soon follow with similar laws. Although there has been increased interest in the study of pain in recent years, there is still much that is not understood. In 2000, Congress declared 2000 to 2010 as the Decade of Pain control and Research, a mandate supported by the APS as the leading professional society in the field of pain management. In addition, the APS has developed and implemented the following core programs: (a) Public Awareness, (b) Professional Awareness, (c) Public Policy Agenda, and (d) Research Agenda.

Physiology

The oldest theory of pain that is still used today is the gate theory, which describes the mechanisms of pain. The flow of pain impulses from the peripheral nervous system and descending messages from the central nervous system can be increased or decreased by a neural "gating" mechanism in the dorsal horn (substantia gelatinosa). Endogenous opioids "dose" the gate and reduce transmission of pain. This mechanism, termed *nociception*, takes place in four stages.

1. *Transduction* begins in the periphery at the site of injury. Cell damage releases sensitizing substances: prostaglandin, bradykinin, serotonin, substance P, and histamine. An action potential results from the release of these substances (nociceptive pain) plus a change in the charge along the neuronal membrane or abnormal processing of stimuli by the nervous system (neuropathic pain). The change in the charge occurs when sodium moves into the cell and other ion transfers occur.

2. *Transmission* occurs in three phases. (a) The first phase is from the injury site to the spinal cord. Nociceptors terminate in the spinal cord. (b) The second phase is from the spinal cord to brain stem and thalamus. Release of substance P and other neurotransmitters continue the impulse across the synaptic cleft between the nociceptors and the dorsal horn neurons. From the dorsal horn of the spinal cord, neurons such as the spinothalamic tract ascend to the thalamus. Other tracts carry the message to different centers in the brain. (c) The third phase is from the thalamus to cortex. The thalamus acts as a relay station sending impulses to central structures for processing.
3. *Perception or the conscious experience of pain.*
4. *Modulation or the inhibition of nociceptive impulses.* Neurons originating in the brain stem descend to the spinal cord and release substances such as endogenous opioids, serotonin, and norepinephrine that inhibit the transmission of nociceptive impulses.

Pain Categories

Pain is broken down into categories based on the duration of pain (acute and chronic) and by physiology (cancer, nociceptive, and neuropathic).

Duration

Acute pain refers to pain that has a short life span. Once the cause, for instance, an injury, illness, or surgery, has resolved, the pain resolves. Postsurgical pain and procedural pain are subcategories of acute pain. Acute pain (also known as physiological pain) is usually somatic and/or visceral or nociceptive (e.g., surgical pain from traumatized skin, muscle, and visceral organs). Acute pain may be related to trauma, surgery, the inflammatory process, injury, or procedures. It is a short-term pain experience that demonstrates progressive resolution as the tissue heals, characterized by common physiologic responses (elevated heart rate, blood pressure, and respiratory rate; diaphoresis) and common behavioral responses (grimacing, crying, moaning, guarding). The cause is usually known and has a beginning and an end.

Chronic pain (also known as pathophysiological pain) persists longer than the usual course of an acute disease or a reasonable time for an injury to heal and may be associated with a chronic pathological process that causes continuous or intermittent pain for months or years. The cause of the pain may not be evident. Physiological and behavioral responses are not always evident in this group due to their return to baseline. Chronic pain serves no purpose and may have severe, intermittent exacerbations. The chronic pain patient population is increasing as a result of an increase in patient population of aging baby boomers and increased side effects from medical treatments, such as from antifungals or chemotherapies.

Physiology

Cancer pain results from three primary causes. It may be (1) related to tumor involvement, encroachment of surrounding tissues and organs by the tumor in 65% to 85% of cancer patients with pain; (2) a result of cancer treatment such as radiation or chemotherapy in 15% to 25% of cancer patients with pain; or (3) unrelated to cancer or its treatments in 3% to 10% of cancer patients with pain, resulting in structural and chemical alterations that affect the nature of the impulses. In some cases, these nerve endings remain hyperexcited, and normal touch is experienced as a severe painful burning sensation.

Neuropathic pain refers to an abnormal pain that outlasts the injury and is associated with nerve and/or central nervous system changes. Examples of this type of pain include neuropathies, postherpetic neuralgia, trigeminal neuralgia, complex regional pain syndrome, and others. Complex regional pain syndrome was previously termed *reflex sympathetic dystrophy*. During the Civil War, it was termed *causalgia*. Neuropathic pain is difficult to treat, but it can be done. This pain is described by patients as "hot, burning, prickly, needles and pins, electric-shock-like, shooting, lancinating etc."

Nociceptive pain emanates from damaged tissues, unlike neuropathic pain, which emanates from nerve damage. Examples of this include surgical pain, pain from fractures, burns, infections, lacerations, arthritis, organ (visceral) pain related to illness or organ damage, and so on. This pain is usually described using terms such as dull, deep, hard, bruised, sharp, aching, throbbing, or sore.

Pain Interventions

The most crucial part of managing pain is an appropriate pain assessment. This assessment consists of

onset, location, duration, characteristics, aggravating/alleviating factors, radiating, temporal factors, and severity. Several tools have been developed that are valid and reliable to measure pain. Examples of these tools include the 0 to 10 or 0 to 5 numerical scales, word descriptors, and Wong-Baker faces. The most commonly used scale is the 0 to 10 numerical scale. A behavior checklist has been designed for nonverbal or cognitively impaired patients. The list includes behaviors such as grimacing, moaning, guarding, and so on, and if the patient exhibits any or more of the behaviors on the list, it can be concluded that the patient is experiencing some form of discomfort.

Pharmacological interventions for pain include opioids, nonsteroidal antiinflammatory drugs (NSAIDS), and adjuvant analgesics. Nonpharmacological interventions include music, prayer, massage, acupuncture, cold, heat, relaxation exercises, and invasive procedures such as blocks, spinal cord stimulator implants, intrathecal pump implants, and others.

Research affirms that the best approach for the treatment of pain is multimodal. Treatment may include analgesics, physical therapy, cognitive approaches, mechanical interventions, complementary alternative approaches, and invasive procedures. Under the category of analgesics fall several medications that may be used independently of each other or in conjunction with each other. These analgesics include opioids, adjuvants, and NSAIDS.

Opioids are still the most important kind of analgesic. In the opioid category, there are two types: agonists and agonist-antagonist. Agonists are those opioids that bind to the mu receptors to produce analgesia. Agonist-antagonists are those opioids that presumably bind to the mu receptor but either exert no action (competitive agonist at the mu receptor) or exert only a limited action (partial agonist). Some medications such as nalorphine, cyclazosine, and nalbuphine block the mu receptor (competitive antagonists) but have a partial analgesic effect at other receptors. Analgesics mixed with other analgesic compounds—such as codeine and acetaminophen (Tylenol {#}3), hydrocodone and acetaminophen (Vicodin), or oxycodone and aspirin (Percodan)—are sometimes called weak opioids. Opioids do not have ceiling doses, meaning that the dose can be increased as much as needed until the patient achieves analgesia or experiences intolerable side effects. Mixed analgesics have ceilings based on the nonopioid. The maximum acetaminophen dose per 24 hr is 4 g. Adjuvants comprise medications that are indicated for other conditions and are used to treat pain. Adjuvants include anticonvulsants, antidepressants, steroids, local anesthetics, antihypertensives, and muscle relaxants. The final group of analgesics is the NSAIDS and acetaminophen.

Ethics

Practitioners are at times confronted with ethical issues related to treatments. One such treatment is the placebo. Placebo use is considered unethical if it is used for reasons other than research and the patient has not signed an informed consent. End-of-life care includes managing pain and symptoms such as terminal agitation. Opioids, sedatives, and other analgesics are used to provide comfort to the dying. Other dilemmas arise when patients are suspected of "drug seeking" and practitioners refuse to provide analgesic interventions. Definitions have been provided to help guide practice. The problem with opioids is that to patients with addictive personalities, the step to drug abuse is very short. This causes clinicians to feel uncomfortable when prescribing opioids—thus the tendency to underprescribe.

In 2001, the American Academy of Pain Medicine, the American Pain Society, and the American Society of Addiction Medicine recognized the following definitions and recommended their use:

1. *Addiction* is a primary, chronic, neurobiological disease, with genetic, psychosocial, and environmental factors influencing its development and manifestations. It is characterized by behaviors that include one or more of the following: impaired control over drug use, compulsive use, continued use despite harm, and craving.
2. *Physical dependence* is a state of adaptation that is manifested by a drug class specific withdrawal syndrome that can be produced by abrupt cessation, rapid dose reduction, decreasing blood level of the drug, and/or administration of an antagonist.
3. *Tolerance* is a state of adaptation in which exposure to a drug induces changes that result in a diminution of one or more of the drug's effects over time.
4. *Pseudoaddiction* describes patient behaviors that may occur when pain is undertreated. Patients with unrelieved pain may become focused on obtaining medications, may "clock watch," and may otherwise seem inappropriately "drug seeking." Even

behaviors such as illicit drug use and deception can occur in the patient's efforts to obtain pain relief. Pseudoaddiction can be distinguished from true addiction in that the behaviors resolve when pain is effectively treated.

—Patti Shakhshir

See also Cancer; Drug Abuse and Dependence, Epidemiology of; Ethics in Health Care; Quality of Life, Quantification of

Further Readings

Agency for Health Care Research and Quality. (1994). *Clinical practice guideline: Management of cancer pain.* Rockville, MD: U.S. Department of Health and Human Services.

American Cancer Society. (2001). *Guide to pain control: Powerful methods to overcome cancer pain.* Atlanta, GA: Author.

American Pain Society. (2003). *Principles of analgesic use in the treatment of acute pain and cancer pain* (5th ed.). Glenview, IL: Author.

Doka, K. J. (2006). *Pain management at the end of life: Bridging the gap between knowledge and practice.* Washington, DC: Hospice Foundation of America.

Kuebler, K. K., Berry, P. H., & Heidrich, D. E. (Eds.). (2002). *End of life care: Clinical practice guidelines*. Philadelphia: W. B. Saunders.

McCaffery, M. (1968). *Nursing practice theories related to cognition, bodily pain, and man-environment interactions.* Los Angeles: University of California at Los Angeles Student's Store.

McCaffery, M., & Pasero, C. (1999). *Pain: Clinical manual* (2nd ed.). St. Louis, MO: Mosby.

Melzack, R., & Wall, P. D. (2003). *Handbook of pain management.* New York: Churchill Livingstone.

Mersky, H. (Ed.). (1986). Classification of chronic pain: Descriptions of chronic pain syndromes and definitions of pain terms. *Pain*, *24*(Suppl. 1), S215–S221.

Portenoy, R. K., & Lesage, P. (1999). Management of cancer pain. *Lancet*, *353*(9165), 1695–1700.

Raj, P. P. (2003). *Pain medicine: A comprehensive review* (2nd ed.). St. Louis, MO: Mosby.

Ross, E. L. (2004). *Hot topics: Pain management*. Philadelphia: Hanley & Belfus.

Schechter, N. L., Berde, C. B., & Yaster, M. (Eds.). (2003). *Pain in infants, children, and adolescents* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

Smith, H. S. (2003). *Drugs for pain.* Philadelphia: Hanley & Belfus.

St. Marie, B. (2002). *American Society of Pain Management Nurses: Core curriculum for pain management nursing.* Philadelphia: W. B. Saunders.

Pan American Health Organization

The Pan American Health Organization (PAHO) is the oldest international public health agency in the world. Since 1902, it has worked to improve the health and living standards of the people of the Americas. PAHO is part of the United Nations system, serves as the Regional Office for the Americas of the World Health Organization, and is the health organization of the Inter-American System. The organization is not a financing agency but rather a technical cooperation agency, helping countries share technical information and mobilize health resources. From its headquarters in Washington, D.C., PAHO directs the scientific and technical efforts of experts in 27 country offices and in nine scientific centers. PAHO's annual budget for its core programs totals about $130 million, largely from assessed contributions paid by Member Governments. In addition, PAHO receives some support for health programs from other contributing countries, foundations, and the private sector.

PAHO refers to the regional cooperation among countries on health issues as Pan-Americanism, and the organization has helped countries work cooperatively, initiating multicountry health ventures in Central America, the Caribbean, the Andean Region, and the Southern Cone. Its field office on the U.S.-Mexico border works with states and counties on both sides of the border to solve common health problems.

PAHO's Mission, Vision, and Values

As stated on its Web site, PAHO's mission is to lead strategic collaborative efforts to promote equity in health, to combat disease, and to improve the quality and lengthen the lives of the peoples of the Americas. It works with the ministries of health of member nations and many other groups to improve the health of the peoples of the Americas and to strengthen health systems. PAHO promotes primary health care strategies to reach people in their communities and to extend health services equitably to all individuals, especially those who are vulnerable and impoverished. It supports programs to reduce the toll of chronic diseases and prevent transmission of communicable diseases, including old diseases that have reemerged—such as cholera, dengue, and

tuberculosis—and new diseases such as HIV/AIDS, West Nile virus, and SARS.

PAHO's vision is to be the major catalyst for ensuring that all the peoples of the Americas enjoy optimal health and contribute to the well-being of their families and communities.

PAHO's values, as stated on its Web site ("About PAHO"), are the following:

- *Equity:* Striving for fairness and justice by eliminating differences in health that are unnecessary and avoidable
- *Excellence:* Achieving the highest quality in what [the organization] does
- *Solidarity:* Promoting shared interests and responsibilities and enabling collective efforts to achieve common goals
- *Respect:* Embracing the dignity and diversity of individuals, groups, and countries
- *Integrity:* Assuring transparent, ethical, and accountable performance

PAHO's Leadership

Dr. Mirta Roses Periago of Argentina became the new Director of the Pan American Health Organization on January 31, 2003. The Ministers of Health of the Americas elected her to a 5-year term. She is the fourth Latin American and the first woman to lead the world's oldest international health agency. Dr. Joxel García, a native of Puerto Rico and former commissioner of the Connecticut Department of Public Health, was PAHO's deputy director until October 31, 2006. Dr. Carissa Etienne, a former official of the Ministry of Health of Dominica and head of its National AIDS Program, is PAHO's assistant director.

PAHO's Activities

PAHO works with the countries in a variety of areas such as disease prevention and control, family and community health, sustainable development and environmental health, technology and health services delivery, information and knowledge management, health analysis and information systems, emergency preparedness and disaster relief, strategic analysis and partnerships, and strategic health development and public information, among others. PAHO is committed to the United Nations' Millennium Development Goals and supports the drive to provide primary health care for all.

PAHO focuses on providing equal access to quality health care, safe drinking water and adequate sanitation to the most vulnerable groups, including mothers and children, laborers, the poor, the elderly, refugees, and displaced persons. PAHO also works to ensure that the countries' "have-nots" benefit from environmental protection against pollution, including toxic waste. Efforts are also directed at reducing pernicious gender inequity, reducing domestic abuse, and providing information on reproductive health.

A major priority for the Americas is to cut infant mortality, and PAHO is working to prevent infant deaths through the Integrated Management of Childhood Illness strategy, a simple and practical approach that teaches health workers a complete protocol for evaluating the health status of children brought to a health post or clinic. This helps reduce toll from diarrheal diseases, including cholera, and provides adequate diagnosis and treatment of acute respiratory infections, saving the lives of hundreds of thousands of children each year.

PAHO is committed to eliminating or controlling vaccine-preventable disease. One of the most notable successes was the eradication of smallpox from the Americas in 1973, a triumph that 5 years later led to global eradication of the dreaded disease. In 1994, PAHO led the eradication of polio from the Americas. Polio eradication is now a global goal. PAHO is on the way to eliminating measles from this hemisphere. Health officials have also agreed to seek to eradicate rubella and congenital rubella syndrome, responsible for many birth defects, and are pressing on with the introduction of new vaccines such as Haemophilus influenza B to reduce meningitis and respiratory infections.

PAHO also supports efforts to control malaria, Chagas' disease, dengue, urban rabies, leprosy, and other diseases that affect people in the Americas.

Action is ongoing to increase the supplies of safe blood by screening 100% of donated units of blood for disease and infection. Ensuring that volunteers free of disease donate all blood for transfusion is a critical goal for PAHO.

PAHO helps countries to identify and promote healthy lifestyles and to cope with issues of mental health, family health, reproductive health, and nutrition. It addresses major nutritional problems, including protein-energy malnutrition, and is working to

eliminate iodine and vitamin A deficiencies. It also assists countries with health problems typically found in developed and urbanized cultures, such as cardiovascular diseases, cancer, accidents, smoking, addiction to drugs and alcohol, and others. PAHO's governing bodies have mandated PAHO to move aggressively in the fight to reduce the use of tobacco, emphasizing the negative health consequences and high costs of tobacco use.

PAHO disseminates scientific and technical information through publications, its Internet site, and networks of journalists, libraries, and documentation centers. It is a leader in the use of advanced communications technologies for health promotion, education, and a variety of specialized public health fields.

PAHO coordinates emergency humanitarian relief and technical assistance to regions struck by natural disasters and helps them prepare adequately to mitigate the effects of disasters.

—Daniel Epstein

See also Child and Adolescent Health; Epidemiology in Developing Countries; Maternal and Child Health; Vaccination; World Health Organization

Web Sites

Pan American Health Organization: http://www.paho.org.

Panel Data

Panel data, also known as longitudinal data, are important in many areas of research, including epidemiology, psychology, sociology, economics, and public health. Data from longitudinal studies in clinical trials and cohort studies with long-term follow-ups are a primary example of panel data. Unlike data from traditional cross-sectional studies, panel data consist of multiple snapshots or panels of a study group or a cohort of subjects over time and, thus, provide a unique opportunity to study changes in outcomes of interest over time, causal effects, and disease progression, in addition to providing more power for assessing treatment differences and associations of different outcomes. Such data also present many methodological challenges in study designs and data analyses, the most prominent being correlated responses and missing data. As a result, classic models for cross-sectional data analysis such as multiple linear and logistic regressions do not apply to panel data.

Methodologic Issues for Panel Data Analysis

In cross-sectional studies, observations from study subjects are available only at a single time, whereas in longitudinal and cohort studies, individuals are assessed or observed repeatedly over time. By taking advantages of multiple assessments over time, panel data from longitudinal studies capture both between-individual differences and within-individual dynamics, offering the opportunity to study more complicated biological, psychological, and behavioral hypotheses than those that can be addressed using cross-sectional or time-series data. For example, if we want to test whether exposure to some chemical agent can cause a disease of interest such as cancer, the between-subject difference observed in cross-sectional data can provide evidence only for an association or correlation between the exposure and disease. The supplementary within-individual dynamics in panel data allows for inference of a causal nature for such a relationship.

Although panel data provide much richer information about the relationship among different outcomes, especially their causal nature, they raise challenging methodologic issues for study design, data analysis, and interpretation of analysis results. The two most important concerns are correlated responses and missing data. First, panel data create correlated responses because repeated assessments are collected from the same subject. For example, if we measure an individual's blood pressure twice, the two readings are correlated since they reflect the health condition of this particular individual; if he or she has high blood pressure, both readings tend to be higher than the normal range (positively correlated) despite the variations over repeated assessments. The existence of such within-subject correlations invalidates the independent sampling assumption required for most classic models, and as a result, statistical methods developed based on the independence of observations, such as the analysis of variance (ANOVA) and the multiple linear and logistic regression models, are not valid for panel data. In the blood pressure example, if we ignored the correlations between the two readings and

modeled the mean blood pressure using ANOVA, then for a sample of n subjects, we would claim to have $2n$ independent observations. However, if the two readings were collected within a very short time span, say 5 s apart, they would be almost identical and would certainly not represent independent data comparable to blood pressure readings taken from two different people. In other words, the variation between two within-subject readings would be much smaller than any two between-subject observations, invalidating the model assumption of independent observations and yielding underestimated error variance in this case. Although assessments in most real studies are not spaced as closely as in this extreme example, the within-subject correlation still exists. Ignoring such correlations by applying classic models may yield incorrect inferences.

Missing data present a more serious challenge in the analysis of panel data. Because panel data are collected over time, frequently in studies of lengthy duration, missing data are inevitable and happens for a variety of reasons. For example, in clinical trial studies, subjects may simply quit the study or not return for follow-up visits because of problems with transportation, weather, deteriorated or improved health condition, relocation, accidental death, treatment complications, and so on. Because missing data can seriously bias model estimates for panel data analysis, they are characterized in terms of their impact on inference through statistical assumptions or missing data mechanisms, which allow statisticians to ignore the multitude of specific reasons for missing data when performing data analysis. The *missing completely at random* (MCAR) assumption refers to a class of missing data that do not affect inference on model parameters when completely ignored. MCAR corresponds to a layperson's notion of data missing at random and includes all types of missing data that are unrelated to study or treatment, such as relocation, transportation problems, and bad weather conditions. Another important class of missing data is called *missing at random* (MAR), which generalizes MCAR to deal with treatment-related missing data. In many clinical trials, missing data are often associated with the treatment interventions under study. For example, a patient may quit the study if he or she feels that the study treatment has deteriorated his or her health condition, and any further treatment would only worsen the medical or psychological problems. Or, a patient may feel that he or she has completely responded to the treatment and does not see any additional benefit in continuing the treatment. In such cases, missing data do not follow the MCAR model since it is predicted by study or treatment-related responses. This class of reasons for missing data is modeled by the MAR assumption, which posits that the occurrence of a missing response depends on the response history or observed pattern. Thus, MAR postulates a plausible and applicable missing data condition that encompasses many treatment-related missing data and constitutes a sensible statistical approach to address bias in such situations. It is important to distinguish between these two common types of missing data, as some statistical models for panel data analysis only provide unbiased estimates under the stronger MCAR assumption, as we discuss next.

Modeling Approaches for Panel Data Analysis

As most classic models do not apply to panel data, specialized methods must be employed to address the within-subject correlation and missing data problems. The two most popular approaches for panel data analysis are the mixed-effects models (MM) and generalized estimating equations (GEE). MM is a general class of models that can be used for regression analysis with continuous, ordinal, categorical, and count responses. As panel data are common in an ever-widening variety of experimental and observational study designs in the biological, biomedical, behavioral, economic, epidemiological, and social sciences, various applications of MM have been found in these disciplines under different guises such as random coefficient models, random effects models, random regression, hierarchical linear models, latent variable models, mixed models, and multilevel linear models. Unlike the classic multivariate linear model, MM does not directly model the correlations among the repeated, within-subject responses but rather employs random effects (or latent variables) to account for such correlations. As a result, this approach enables one to model correlation structures without directly involving the within-subject responses, giving rise to a powerful and flexible way to deal with missing data as well as to address varying assessment times often found in cohort and longitudinal observational studies.

MM is a class of parametric models, requiring analytic distribution assumptions for both the response

(observed) and the random effects (latent). A fundamental problem with using random effects to account for correlated responses is the difficulty in empirically validating the distribution assumption for the random effects since they are latent variables. Further exacerbating this problem, MM relies on the distribution assumption for the response for inference. If either set of assumptions is violated, estimates will be biased, laying the basis for spurious findings.

A remarkable breakthrough underlying GEE-based inference is the elimination of both sets of assumptions, leading to inference of model parameters that are robust to data distributions and within-cluster correlation structures. Like MM, GEE is a general class of regression models capable of accommodating all types of responses. However, unlike MM, which relies on parametric distributions for the random effects and response for inference, GEE models the marginal mean of the response and uses estimating equations for inference, eliminating both layers of assumptions and thereby providing unbiased estimates regardless of the complexity of the correlation structure and the data distribution. Thus, under GEE, neither the specification of random effect nor the distribution of response is required.

A major weakness of GEE is that it requires MCAR to provide unbiased estimation of model parameters. In this regard, MM is more robust as it guarantees valid inference under both MCAR and MAR, provided, of course, the model assumptions (particularly the distribution ones) are met. This limitation of GEE is addressed by the latest development on GEE-based inference and in particular, the weighted GEE (WGEE), which extends the GEE to provide valid inference under MAR. Thus, WGEE is robust against not only data distributions but also the missing data types.

MM, GEE, and WGEE are implemented in most popular statistical software such as SAS, Splus, and SPSS. For example, MM is implemented in the SAS NLMIXED procedure and GEE and WGEE in the GENMOD procedure.

Other Related Topics

There are several panel-data-related topics that are not covered in this entry. For example, measurement error may arise in panel data that cannot be addressed by classic measurement error models. Also, in clinical trials that involve higher mortality rates, it may be necessary to jointly model the patient's dropout process and outcome of interest to provide more information about the progression of disease. In addition to MM, GEE, and WGEE, multiple imputation (MI) may also be used to address missing data. A great advantage of MI is that it uses software for complete data analysis to provide inference in the presence of missing data. However, as the specialized methods such as MM and GEE are implemented in most major statistical packages, MI is not widely used for panel data analysis. Finally, causal inference involving multiple outcomes collected over time cannot be addressed by MM, GEE, or WGEE and requires more specialized methods such as the structural equation models. The list of further readings provides detailed discussions of these and other panel-data-related research topics.

—Changyong Feng, Wan Tang, and Xin Tu

See also Descriptive and Analytical Epidemiology; Missing Data Methods; Robust Statistics; Study Design

Further Readings

Diggle, P. J., Heagerty, P. J., Liang, K. Y., & Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford, UK: Oxford University Press.

Dmitrienko, A., Molenberghs, G., Chuang-Stein, C., & Offen, W. (2005). *Analysis of clinical trials using SAS: A practical guide.* Cary, NC: SAS Institute.

Hanley, J. A., Negassa, A., Edwardes, M. D. deB., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology*, *157*, 364–375.

Hsiao, C. (2003). *Analysis of panel data* (2nd ed.). Cambridge, UK: Cambridge University Press.

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Rotnizky, A., & Robins, J. M. (1997). Analysis of semi-parametric regression models with non-ignorable non-response. *Statistics in Medicine*, *16*, 81–102.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall.

Tu, X. M., Kowalski, J., Zhang, J., Lynch, K., & Crits-Christoph, P. (2004). Power analyses for longitudinal trials and other clustered study design. *Statistics in Medicine*, *23*, 2799–2815.

Parasitic Diseases

Parasites can be defined as organisms that live in or on another organism called a host. In most situations, the

parasite benefits from this relationship, often at the expense of the host organism. Traditionally, parasites include protozoans and helminths. However, today, the term *parasite* is sometimes used to describe the multitude of viruses, bacteria, fungi, plants, and animals, including ticks, mites, and lice, that act in a parasitic fashion. Traditional parasites (protozoans and helminths) are responsible for many diseases in animals and humans and are transmitted to their host most often through the ingestion of contaminated food or water or arthropods, which act as intermediate hosts and vectors. Parasites pose health risks and economic costs in livestock and in humans and are often associated with epidemics when a disease occurs at a higher rate than would be expected within a defined area. The high prevalence of parasitic disease in humans provides opportunity for epidemiological studies that examine parasite pathogenicity, hosts, environment, and social conditions that may play a role in the spread of disease.

Traditional Parasites

Research has shown that parasites existed in ancient civilizations as evidenced by written records and the discovery of eggs of parasites in ancient Egyptian mummies. In 1875, Fedor A. Lösch demonstrated that the causative agent of dysentery was the protozoan *Entamoeba histolytica.* Protozoa are single-celled, heterotrophic eukaryotes, most of which are free-living. The discovery of *E. histolytica* as a pathogen led to the identification of other species of pathogenic protozoa. The flagellated protozoa *Trypanosoma rhodesiense* that is transmitted to humans through the bite of infected tsetse flies causes sleeping sickness (African trypanosomiasis). The organisms reproduce rapidly, avoiding recognition by antibodies in the blood and possibly outnumbering red blood cells. They travel through the bloodstream and eventually reach the spinal cord and brain, leading to coma and death. In a healthy human infected with *T. rhodesiense*, the disease may become a chronic condition, with the organism later becoming opportunistic if the immune system is weakened. The protozoa of the *Leishmania* species are transmitted to humans by infected sand flies and infect macrophages that attempt to engulf and digest the foreign pathogen. Eventually, the macrophages and immune defenses become overwhelmed causing leishmaniasis. This is a debilitating and fatal disease and epidemics have occurred in India, China, Africa, and Brazil.

Descriptions of malarial disease have dated back to ancient Chinese and Greek civilizations; however, the actual cause of malaria, protozoans of the genus *Plasmodium*, was not discovered until 1898, when it was found that humans could become infected through the bite of an infected mosquito. Once in the bloodstream, the *Plasmodium* travels to the liver where it infects and replicates within cells. The burden of organisms within a single cell will cause it to burst, releasing the *Plasmodium*, and allowing it to infect red blood cells, where it replicates rapidly, again causing the cell to rupture. The infection cycle into red blood cells may happen several times, resulting in a large quantity of *Plasmodium* and the symptoms of infection, such as intermittent fever.

Helminths include roundworms, also called nematodes, and flatworms, such as flukes and tapeworms. Filarial disease may be caused by one of several species of helminth nematodes. The filarial nematode *Wucheria bancrofti* is transmitted to humans by arthropods and is commonly found in Africa, the Middle East, Mexico, and Brazil. Humans and mosquitoes are the only suitable hosts in which *Wucheria* is able to complete its life cycle. Once injected into the bloodstream, the immature worms make their way into a lymph duct and mature into adults, which may take between 6 months and a year and result in a worm about 3 to 4 inches in length. Multiple adult females will exist in streams of clusters within a lymph duct where they reproduce, shedding thousands of microfilia every day, sometimes remaining in the lymph duct for 5 to 10 years. The accumulation of microfilia in the lymph ducts blocks the flow of lymphatic fluid causing swelling in affected body parts. Microfilia eventually migrate into the bloodstream and are drawn into the proboscis of a mosquito when it bites the host. In general, the immune system is able to defend against and kill the majority of microfilia, resulting only in minor illness associated with the lymphatic system and a low rate of morbidity.

Flatworms, such as the tapeworm, can cause infection and disease in humans. The tapeworm species *Diphyllobothrium latum* or one of several different species in the genus *Taenia*, in addition to several other genera, can be infectious. Tapeworms are highly specialized worms that attach themselves to the lining of the human intestinal wall using hooks that keep them firmly in place, or burrow into tissues such as muscle, the spinal cord, or the brain. Adult tapeworms tend to stay within their host as long as possible, with

just one adult tapeworm per host growing and reproducing continuously, shedding eggs that are excreted in the feces. Juvenile tapeworms pose the greatest health risks to humans because of their tendency to burrow through the intestinal wall, migrating to internal organs where they interfere with normal tissue function. Infection with adult tapeworms may be asymptomatic; sometimes abdominal pain or diarrhea occurs but is not immediately known to be the result of a tapeworm infection. Infection with juvenile tapeworms can cause cysticercosis, typified by the formation of cysts under the skin, inflammation, mental disorientation, and seizures.

Parasite Life Cycles

All parasites have a life cycle that involves a period of time spent in a host organism and the phases of which can be divided into growth, reproduction, and transmission. Life cycles of parasites can be divided into two categories: direct (monoxenous) and indirect (heteroxenous). Parasites with direct life cycles spend most of their adult lives in one host, known as the parasitic stage, with their progeny transmitted from one host to another, known as the free-living stage. Direct parasites often lack an intermediate stage and must leave their host. To do this, they must be able to survive in an environment outside their original host and then locate and establish in a new host. Parasites that depend on the host stage are called obligate parasites, whereas parasites that can skip the parasitic stage for several generations are called facultative parasites. Roundworms, trypanosomatids, and *Cryptosporidium* are examples of parasites with direct life cycles. Parasites with indirect life cycles are characterized by two host stages, which require a definitive host and an intermediate host. The definitive host stage is required for reproduction and the adult life phase. Within the intermediate host, parasite development occurs, after which it can be transmitted to a definitive host. Multiple developmental stages may take place in an intermediate host, which plays an important role in facilitating disease transmission in the form of vectors, such as mosquitoes, which pass immature parasites through their proboscis directly into the bloodstream of the definitive host. Filarial nemotodes, *Plasmodium*, and *Leishmania* are examples of parasites with indirect life cycles. Reservoir hosts typically tolerate parasites with no ill effects; however, the introduction of a new host into a population of reservoir hosts will often result in severe disease in the newly introduced host.

Epidemiology

Close to 3 billion people worldwide are infected with parasites. Parasites are often endemic and sometimes epidemic in certain regions of the world. For instance, the pathogenic parasite *Plasmodium* that causes malaria is a constant concern in Africa and often occurs in endemic and epidemic proportions, and sleeping sickness, caused by a species of *Trypanosoma*, has resulted in epidemic disease in Uganda. The emergence of diseases resulting from pathogenic parasites is always an issue of concern in many tropical regions around the world. Studies of infected populations and of pathogenic parasites have provided, and continue to provide, insight into ways to prevent, treat, and control disease. Following simple sanitation procedures, such as washing hands, cooking meat, and keeping human waste separate from humans, food supplies, pets, and livestock can prevent many diseases caused by parasites. However, these sanitary guidelines are not so easy to follow in Third World countries, which may lack monetary support to provide clean water sources or proper medications, or in cultures where humans maintain intimate dwellings with their livestock and their coexisting parasites. A more difficult problem to overcome is the presence of arthropod hosts in numbers sufficient to infect large number of people over a short period of time, especially in tropical climates. These diseases can be maintained indefinitely in human populations that have a lack of access to medical relief to break the parasite infection cycle.

Treatment

Unsanitary conditions are often the underlying cause of parasitic disease, especially in areas that are overpopulated or have poor water quality, or among populations that lack knowledge of parasites. In addition, parasites have evolved in ways that enable them to avoid antibody recognition and elimination by the immune system by entering into the cells of the body. This results in a cell-mediated immune response, including activation of helper and cytotoxic T cells, cytokines, and interleukins. Protozoa such as *Toxoplasma gondii*, *Leishmania*, and *Plasmodium* have each found ways to avoid or use the human immune

system to their advantage, facilitating their replication and increasing their pathogenicity. Antiparasitic drugs can be divided into antiprotozoan agents and antihelminthic agents. Antiprotozoans are typically designed to be effective in disrupting a specific stage in a parasite life cycle. Drugs against *Plasmodium* can be taken as prophylactics in the case of oral chloroquine or as treatment for acute attacks in the case of oral chloroquine in combination with sulfadoxine or oral quinine in combination with tetracycline. Other commonly used antiprotozoal drugs include metronidazole, amphotericin B, and suramin.

Antihelminthic drugs cause physical damage to parasitic worms, in most cases targeting adult worms, and inhibit their metabolism, inhibit their ability to lay eggs, or facilitate their excretion from the host. The class of benzimidazole antihelminthics, which includes mebendazole and albendazole, causes degeneration of microtubules that inhibits glucose uptake. These drugs are used as first line therapy for most roundworm and some tapeworm infections. Ivermectin, synthesized from a group of naturally occurring substances, is often used to treat leishmaniasis by causing paralysis of the infecting worms and is effective against *W. bancrofti.* Other antihelminthic agents include diethylcarbamazine and praziquantal.

—Kara E. Rogers

See also Epidemiology in Developing Countries; Malaria; Waterborne Diseases

Further Readings

Fernando, R. L., Fernando, S. S. E., & Leong, A. S.-Y. (2001). *Tropical infectious diseases*. Boston: Cambridge University Press.

Hardman, J. G., & Limbird, L. E. (Eds.). (2001). *Goodman and Gilman's the pharmacological basis of therapeutics* (10th ed.). New York: McGraw-Hill.

Mehlhorn, H. (Ed.). (2001). *Encyclopedic reference of parasitology: Biology, structure, function* (2nd ed.). New York: Springer.

Participation Rate

See Response Rate

Participatory Action Research

Research oriented toward action and/or change can take three forms: (1) It can apply the professional expert model, in which the researcher makes a study and recommends a course of action to decision makers in the organization studied; (2) it may involve action research controlled by the researcher, in which the researcher aims to be a principal change agent as well as controlling the research process; or (3) it may involve participatory action research (PAR), in which the researcher seeks to involve some members of the organization studied as active participants in all stages of the research or action process.

Investigations in the PAR context indicate a systematic effort to generate knowledge about specific conditions that can influence changes in a given situation. The term *action* in research indicates that the research is meant to contribute to change efforts or accompany action by the part of participants, such as workers and their representative trade unions, or change for employers, through the research learning process. PAR has its roots in social psychology and is a relatively new research technique in epidemiology but can be applied fruitfully in many contexts, for instance, to understand the causes of and reduce the number of worker injuries in an occupational setting or to reduce morbidity and mortality from diabetes in a community. The entry uses the example of PAR in a workplace setting to illustrate the principles of PAR. Workplace PAR is a process of systematic inquiry in which those who are experiencing a work-related problem participate with trained researchers in deciding the focus of knowledge generation, in collecting and analyzing information, and in taking action to improve the conditions or to resolve the problem entirely.

PAR as a Multidisciplinary Methodology

PAR methodology has been used in community development and health-related research, such as with community health workers and nurses, in industrial and other types of organizations, and in research in agriculture. In these settings, researchers using PAR have focused primarily on oppressed groups to empower and generate collective action, where new knowledge based on research led to local level and

industrial actions aimed at improvements for workers. PAR has also been used extensively in organizational development in industry and by management's application of human resource theories, particularly those with a systems perspective focusing on the fit, or lack of fit, between technical and social systems. When participatory systems work well, they produce results because they apply a wide range of information and ideas to problems in an organizational context.

Applying PAR as a methodology requires ensuring that those participating in the research feel that the researchers have genuine respect for them and their experiences, that their opinions are valued, and that they are perceived as partners in the process. Research tool choice should depend on what is being studied as well as why a subject is being studied. PAR can be used to extend the principles of education for empowerment, which espouse learning that is participatory, based on real-life experiences. The primary purpose of PAR has, thus, often been to encourage the poor and oppressed, and those who work with them, to generate and control their own knowledge. As a research methodology, PAR assumes that knowledge generates power and that people's knowledge is central to social change.

PAR for Workplace-Based Research

The application of PAR methodology in occupational health research starts from a belief that adults are self-motivated, rich with information that has immediate application to their lives and work. In research where workers' health is in question due to work processes, some degree of change in the organization is likely to be necessary to prevent further work-related injury and illness.

Few problems can be resolved in modern industrial organizations through the use of any single academic discipline; the complex nature of work calls for integrating ideas and methods from a variety of disciplines. The increasingly complex nature of the workplace gives rise to greater need to understand the causes and methods of prevention of work-related accidents, injuries, and illnesses, with the transmission of this information to workers, managers, and others being a critical dimension of preventing these negative outcomes. Important advantages of PAR methods include the qualitative information obtained about workers involved in the study, the potential to obtain a precise picture of work-related risks and the root causes, and the semiquantitative data on adverse health outcomes. Use of the research process as a means of catalyzing a process of consciousness raising and organization among workers is a further advantage, enabling positive actions for workplace improvements.

The need for a holistic view of occupational disease is a compelling reason for occupational physicians to view workers as a member of an occupational health team and gives rise to the fundamental need for a systems approach to health-related problems, involving workers as an indispensable source of information and valuable agent of change.

In conducting PAR-based occupational health research, both qualitative and quantitative techniques can be applied. Quantitative techniques may include survey-based data collection and systematic analysis by structured questionnaire with closed- and open-ended questions. Qualitative techniques may include structured individual interviews with open-ended questions, participant observation on the job, focus group interviews and discussions, videotaping, and examination of work process, workstation analysis, document and archival review, and broad or well-defined literature review. Engaging in discussion with workers and management at various stages of the research process is an important means of increasing confidence in the research findings by both groups, and it can strengthen the process and outcomes aimed at eventual change.

Comparisons by Laurell, Noriega, Martinez, and Villegas (1992) of workplace-based study results based on PAR methodology made with results from an individual questionnaire have affirmed that PAR-generated results are more revealing than individual questionnaire-generated results. Rosskam (2007) applied PAR methodology in a study of airport check-in workers; the process and findings led to direct improvements in various airports and contributed to policy changes in a number of countries. Hugentobler, Israel, and Schurman (1992) used PAR methods to implement a longitudinal multimethodological research and intervention project investigating occupational stress, psychosocial factors, and health outcomes; their findings were combined with intervention to improve worker health. Israel, Schurman, and Hugentobler (1992) used PAR methodology to better understand and try to reduce the negative effects of work-related stress. Ritchie (1996) described the workplace as a useful venue for

research, where social and environmental factors in a work environment can be relatively easily explored, a defined community where one can legitimately explore and help develop improvement-oriented and empowerment-based strategies for action. Schurman (1996) used a PAR approach to study stress in an automobile factory, involving the factory workers in the investigation. The process was designed to improve the system's performance through work organization redesign and to contribute to the body of scientific knowledge at the same time.

Participatory action researchers put into question domination and dominating research structures and relationships, including how actual organizational structures, processes, and practices shape and influence the ways in which those holding decision-making power relate to those not holding decision-making power. This questioning is particularly relevant and important to work-related research given the complex nature of worker-management relationships; indeed, it is the very process of human inquiry that provides the impetus toward action. At its core, PAR in work-related health issues promotes worker participation in decision making in the workplace, which implies an inherent redistribution of power between workers and management. In workplace PAR, this can also be conceived of as sharing and providing information and access to resource mobilization to help others as well as oneself.

PAR methodology incorporates the recognition that research can involve workers as an integral part of the process. The process can be a tool to encourage workers to question and challenge the very systems that may keep them passive and not able to participate in decision making, which in turn adversely affects their well-being. Dialogue and participation in the research process are means for workers to gain a critical understanding of the causes of workplace problems and their role in accepting or challenging these forces. Where workers are consulted and participate in projects likely to affect them, positive outcomes and changes are more likely to be sustainable, and participants can critically analyze the barriers to change in systems of work organization.

PAR in the workplace, like education, is most effective when it includes a holistic view of the context of behaviors, including an analysis of obstacles to safe or healthful work practices, without becoming narrowed to specific behaviors or competencies. Employing a holistic and systems view is concordant with the reality that improvement in working conditions takes place in a wider organizational context of worker and management relations.

PAR Compared With Other Research Methodologies

The professional expert model could be appropriate for examining contributing factors and moderating variables posited to cause a given health outcome. Case studies are not developed as a means of measuring a variety of causal factors. PAR, however, lends itself as a useful method by which multiple factors and outcomes can be examined. The contribution of detailed information from study participants is a key means of learning about factors thought to contribute to, or cause, outcomes, which is not inherent in the professional expert model. The descriptive aspects of a case study model may be useful for detailing a worker population and depicting working conditions. A detailed work analysis is indispensable for understanding how jobs are performed. While observation and questioning can be applied in case study development, the researcher maintains a more distant attitude than that used in a PAR methodology. The difference in researcher attitude between PAR and the professional expert model is significant in determining how a study process unfolds. Where participants feel themselves viewed as the object of scientific research rather than part of a process designed to benefit them, engendering a sense of trust and involvement becomes more difficult.

For a worker health study, where a change process is part of the study design, direct involvement with the participants is essential. Direct involvement with participants is of particular importance to identify problems associated with the job, as these are best known by the person performing the job. The professional expert model does not include the use of focus group discussions as part of the research process, as this could "contaminate," or influence the research process, compromising objectivity. In contrast, focus group discussions are accepted and even encouraged in a PAR methodology as a key means of obtaining rich, qualitative information, and for obtaining support for participation in the study. Focus group discussions can be also very useful for generating hypotheses.

In both the professional expert model and a case study methodology, there is no demarcation between theory and practice. In both these methods, it is the researcher who defines the problem to be studied. There is no

feedback mechanism built into these methodologies, no requirement that knowledge gained be shared directly with the study participants. The professional expert model contributes to the body of knowledge shared by only a limited group of "experts." These methods neither include an action component meant to contribute to change on the part of the participants nor are they designed to create a learning process for the participants. In contrast, PAR is designed to create a learning process among participants and to ensure the dissemination and application of knowledge and experience gained. For research on workers' health problems, where entire systems are questioned, one can argue that research needs to be directly relevant to those involved and that the findings are shared, and have the potential to contribute to organizational change, in addition to contributing to the existing body of knowledge.

Limitations of PAR

One of the difficulties in PAR is to ensure that all groups understand the process and feel validated and valued in their contributions to the research design, process, and any outcomes. PAR researchers must be careful not to speak only in academic terms with the various groups and not to presume the same understanding of research terminology and ability to interpret data—but without appearing condescending or technocratic. This entails a delicate balance on the part of PAR researchers, who should see themselves as learners and facilitators in addition to being scientists. Listening to and learning from workers and managers about workplace issues is valuable and enriching, albeit time-consuming. PAR researchers have the additional responsibility to help the various participants learn from the process, which can be time-consuming.

An additional difficulty in applying a PAR approach is arriving at consensus with workers and employers, establishing a relationship with each group, and bringing them into the process. These can be both difficult and lengthy processes for the researcher. The processes are more time-consuming than research based on, for example, the professional expert model, where input from various groups is not involved and where consensus building is not required at any stage. The consultative process in PAR methodology is likely, therefore, to be more costly in terms of researchers' time and in analysis of qualitative data.

As with any research methodology, there exists the potential for bias to be introduced. This can occur in defining the issues to be addressed, or if the participating groups want more emphasis on certain issues of greater importance or concern to them, which may also be areas of political concern. Stronger emphasis on particular issues by any one participating group could skew results. In working with management and unions, researchers must exercise caution to not fall into political traps, based on the agendas of any one group. While the research team must gain the confidence of all groups involved, if they are perceived to lean more toward one group's interest than another's, they may lose the confidence of the other group. Maintaining what can be a delicate balance is necessary, apart from being perceived as neutral with all groups, while ensuring that all concerns are addressed. There is also a risk of researcher bias through personal involvement with the participant community as well as in any change process that may develop. The researcher can influence how the change process unfolds and how the research findings are interpreted and applied within and beyond the participating groups. Care must be exercised to maintain a distance from the process, allowing the stakeholder groups to define how a change process is envisioned and formulated. In the end, it is the workers and managers who will have to live with the effects of any changes they implement, or any changes they do not implement, since nonimplementation of changes, after awareness has been raised, can also have consequences.

Validity and Reliability of Data Collected

A debate exists about the validity and reliability of data, in particular when they are qualitative and obtained through participatory appraisal. In the course of data collection, researchers' interpretation and conclusions can be confirmed or disputed by participants in the research process. Confirmation by the participant community of the accuracy of the study findings and analysis of the data greatly increases the credibility of the research findings. PAR is open to various interpretations that can include the researcher and participant community designing the study together or researchers designing the study and then collecting the data with the help of the participating group. Perhaps this is the best means of ensuring that research contributes to an organizational change process given the complex nature of today's organizations.

PAR is important to epidemiology and to epidemiologists. Some epidemiologists would challenge PAR as being "not objective." Many epidemiologists

do not involve themselves in PAR because intervention studies are often more difficult to carry out than strictly quantitative research and can take more time and, thus, may cost more. It can be easier to analyze an existing data set than to talk to groups of people, build consensus, and aim for real change to improve health. Yet it is precisely through intervention-based research and the adequate evaluation of this research that public health and occupational health get improved. For epidemiological research to not remain "ivory tower" in nature, epidemiologists need to be aware of the importance of PAR, perhaps today more than ever before. Epidemiologists stand only to gain from joint research together with working people.

—Ellen Rosskam

See also Environmental and Occupational Epidemiology; Health Disparities; Qualitative Methods in Epidemiology

Further Readings

Abrams, H. (1983). The worker as teacher. *American Journal of Industrial Medicine*, *4*, 759–768.

Dash, D. P. (Ed.). (1998). *Problems of action research: As I see it*. Lincoln, UK: Lincoln School of Management.

De Koning, K., & Martin, M. (Eds.). (1996). *Participatory research in health: Issues and experiences*. London: Zed Books.

Deshler, D., & Ewert, M. (1995). *Participatory action research: Traditions and major assumptions*. Ithaca, NY: Cornell University, Department of Education.

Elden, M. (1981). Political efficacy at work: The connection between more autonomous forms of workplace organization and more participatory politics. *American Political Science Review*, *75*(1), 43–58.

Elden, M. (1981). Sharing the research work: Participative research and its role demands. In P. Reason & J. Rowan (Eds.), *Human inquiry: A sourcebook of new paradigm research* (pp. 253–266). Chichester, UK: Wiley.

Hugentobler, M., Israel, B. A., & Schurman, S. J. (1992). An action research approach to workplace health: Integrating methods. *Health Education Quarterly*, *19*(1), 55–76.

Israel, B., Schurman, S. J., & Hugentobler, M. K. (1992). Conducting action research: Relationships between organization members and researchers. *Journal of Applied Behavioral Science*, *28*(1), 74–101.

Khanna, R. (1996). Participatory action research (PAR) in women's health: SARTHI, India. In K. De Koning & M. Martin (Eds.), *Participatory research in health: Issues and experiences* (pp. 62–81). London: Zed Books.

Laurell, A., Noriega, M., Martinez, S., & Villegas, J. (1992). Participatory research on workers' health. *Social Science and Medicine*, *34*(6), 603–613.

Mergler, D. (1987). Worker participation in occupational health research: Theory and practice. *International Journal of Health Services*, *17*(1), 151–167.

Ritchie, J. (1996). Participatory research in the workplace. In K. De Koning & M. Martin (Eds.), *Participatory research in health: Issues and experiences* (pp. 205–215). London: Zed Books.

Rosskam, E. (2007). *Excess baggage: Leveling the load and changing the workplace*. New York: Baywood.

Schurman, S. (1996). Making the New American workplace safe and healthy: A joint labor-management-researcher approach. *American Journal of Industrial Medicine*, *29*, 373–377.

Smith, S., Willms, D., & Johnson, D. (Eds.). (1997). *Nurtured by knowledge: Learning to do participatory action-research*. New York: Apex Press.

Whyte, W. F. (Ed.). (1991). *Participatory action research*. Newbury Park, CA: Sage.

Whyte, W. F. (1991). *Social theory for action: How individuals and organizations learn to change*. Newbury Park, CA: Sage.

Partner Notification

Partner notification (PN) is a fundamental component of sexually transmitted infection (STI) prevention and control programs in many state jurisdictions and may help prevent the spread of STIs and human immunodeficiency virus (HIV) among individuals who engage in risky sexual behaviors outside the context of a monogamous relationship. Traditional PN uses three different strategies for notifying the sexual partners of patients infected with an STI or HIV: provider referral (notification of sexual partners via a third party, such as an individuals' medical provider), partner referral (notification of sexual partners via the index patient), and contract referral (an agreement between the patient and provider, whereby the patient is given the opportunity to notify their sexual partners on their own, with the understanding that their partners will be notified by a third party if they have not been notified by a predetermined date).

PN was previously referred to as contact tracing, though more recently it includes the comprehensive category of partner services—the process of obtaining from individuals recently diagnosed with an STI information about their sexual partners and facilitating the triage procedure for the examination and treatment of those partners. Conducting PN assumes that the index patient has identifying and locating information

for their sexual partners, which is often not the case for individuals with multiple anonymous partners, such as individuals who meet their partners on the Internet sexual Web sites or public and private sex venues. It is also important to note that laws regarding PN for STIs and HIV infection differ from state to state, and that PN is traditionally a voluntary process, and that patient confidentiality is protected by law.

Partner Services for Individuals Diagnosed With STIs

In an attempt to avoid STI reinfection and/or continued transmission and acquisition of infection, it is imperative that examination and treatment of the sexual partner(s) be addressed when medical providers diagnose patients with an STI. The partner services process includes notification of partners in the case of STIs and HIV partner counseling and referral services (PCRS). PCRS programs provide services to HIV-infected persons and their sexual and needle-sharing partners in an attempt to curb infection or, if already infected, to prevent transmission to other individuals. In addition, these programs aid sexual partners of the index patient in gaining earlier access to individualized counseling, HIV testing, medical evaluation, treatment, and other prevention and health services.

Partner Counseling and Referral Services for HIV Infection

In addition to providing partner services for individuals diagnosed with an STI, some state-level PN programs also provide partner services for individuals infected with HIV. PCRS are typically available to individuals with HIV infection as long as contact with program staff is voluntary. Traditionally, the individual does not have to identify himself or herself and may choose whether or not to identify partners after discussion with the PN staff. Program staff do not have access to individuals' medical records related to the HIV infection unless they have the consent of the patient, typically requiring a medical release form.

Discussion about disclosing HIV infection to past and current sexual and injection-drug-using partners should routinely occur for all HIV-infected people and be integrated into the larger system of preventive and clinical care. Whereas the central goal of PCRS for STIs is the eradication of infection through treatment, the success of PCRS for HIV is evidenced by the prevention of new infections. HIV-infected patients should be informed of the benefits and advantages of PCRS, facilitating informed decision making about this service.

Nontraditional, Internet-Based PN

Traditional PN strategies may be appropriate for notification of partners under certain circumstances; however, much of the literature on PN strategies argues for cultural sensitivity and attention to special circumstances in assessing their appropriateness and potential for success as an intervention strategy. In particular, there are numerous considerations when evaluating the success of such strategies for men who have sex with men (MSM), such as the feasibility of notifying anonymous partners.

Internet-based interventions and PN efforts have evolved as a way to counteract the risks associated with seeking sexual partners, particularly anonymous encounters, on the Internet. Online health information on STI and HIV has become more pervasive, and there have been several initiatives to develop online PN systems that permit notification of sexual partners who may not otherwise be identified by the index patient. One such initiative, conducted by Klausner, Wolf, Fischer-Ponce, Zolt, and Katz (2000), found online PN via an Internet chat room for MSM to be moderately successful. These researchers designed a notification system in which the partners of index patients were notified of their possible exposure by an e-mail message sent via their online profile and found that an average of 5.9 partners per index patient sought testing.

In an attempt to assess the acceptability and perceived utility of a PN system for STI exposure among MSM, Mimiaga et al. (2006) recruited 1,848 men using one of the largest MSM sexual Web sites. Participants were recruited via a banner advertisement accessible by U.S. users of the Web site. The vast majority responded favorably to questions about receiving a PN e-mail containing information about being exposed to an STI (80.9%), education about the STI to which the person was exposed (77.8%), where to get tested (82.1%), a contact phone number (75.5%), and a number/e-mail address allowing the recipient to verify e-mail authenticity (78.8%). The majority were "somewhat to very likely" to use

the following components of a PN e-mail: a contact phone number (61.0%), a Web site link providing information about where to get tested (82.6%), an educational Web site about the STI (86.2%), and a number/e-mail address allowing the recipient to verify e-mail authenticity (70.5%). Overall, 70.0% of participants indicated they would use a public health specialist in some capacity to inform partners via an online notification e-mail of possible exposure to an STI had they been infected. These findings demonstrate that high-risk MSM reported high levels of willingness to use electronic media in conjunction with public health specialists for PN if they were infected with or exposed to an STI.

The Role of the Health Care Provider

For STIs that are priority cases, such as chlamydia, gonorrhea, and syphilis, it is helpful for the health care provider to inform patients that a state department of public health official will need to get in touch with them to discuss their sexual partners, specific behaviors that they engaged in with them, and when necessary, a mechanism of action for notification. With regard to STIs that are not priority cases, such as human papillomavirus (HPV) or herpes simplex virus (HSV), typically, the state department of public health relies on the health care provider to discuss the importance of examination and treatment of sexual partners. In most of these cases, the patients will inform their sexual partners themselves. It is recommended that patients infected with chlamydia or gonorrhea be informed to contact all sexual partners within 60 days preceding their diagnosis or contact the most recent partner if more than 60 days occurred since their last sexual contact. If a patient provides consent, their health care provider can contact his or her sexual partners to make them aware of their exposure to a given STI; typically, this process is confidential, in that the index patient is not identified. Health care providers are generally required to report priority cases of STIs promptly and directly to the appropriate state department of public health.

PN is an important public health strategy for preventing STIs and HIV among at-risk populations. Traditional and nontraditional PN efforts have yielded success with respect to curbing continued transmission of STIs and HIV, as well as getting asymptomatic patients in for evaluation and treatment. For continued success, evaluation of PN programs should be conducted and documented with various at-risk groups as nontraditional methods, such as the Internet, might need to be employed with populations who engage in sexual behavior with anonymous or otherwise nonnotifiable individuals.

—Matthew J. Mimiaga and Margie Skeer

See also Ethics in Health Care; Ethics in Public Health; HIV/AIDS; Sexually Transmitted Diseases

Further Readings

Centers for Disease Control and Prevention. (2002). Sexually transmitted diseases treatment guidelines. *Morbidity and Mortality Weekly Report*, *51*, 4–5.

Centers for Disease Control and Prevention. (2003). Using the Internet for partner notification of sexually transmitted diseases: Los Angeles County, California. *Morbidity and Mortality Weekly Report*, *53*, 129–131.

Klausner, J. D., Wolf, W., Fischer-Ponce, L., Zolt, I., & Katz, M. H. (2000). Tracing a syphilis outbreak through cyberspace. *Journal of the American Medical Association*, *284*, 447–449.

Mathews, C., Coetzee, N., Zwarenstein, M., Lombard, C., Guttmacher, S., Oxman, A., et al. (2002). A systematic review of strategies for partner notification for sexually transmitted diseases, including HIV/AIDS. *International Journal of STD and AIDS*, *13*, 285–300.

Mimiaga, M. J., Tetu, A., Novak, D., Adelson, S., VanDerwarker, R., & Mayer, K. H. (2006, August). *Acceptability and utility of a partner notification system for sexually transmitted infection exposure using an Internet-based, partner-seeking Web site for men who have sex with men*. Oral presentation at the XVI International AIDS Conference, Toronto, Ontario, Canada.

Pasteur, Louis
(1822–1895)

Louis Pasteur, a French chemist, is often called the father of modern microbiology. Through the development of vaccines for cholera, anthrax, rabies, staphylococcus, and streptococcus, he discovered much about the nature of infection and laid the groundwork for the microbial theory of disease. Pasteur also contributed greatly to the field of infectious epidemiology by demonstrating how pathogens spread through animal and human populations.

Pasteur examined the role of microorganisms in the transformation of organic matter, which at the time was greatly misunderstood. Instructed by Napoleon to investigate diseases infecting wines, he determined that fermentation results from the action of a specific microorganism. To enable fermentation, the right microorganism must be introduced, and microorganisms that could alter the process must be kept out. With Claude Bernard, Pasteur developed a process, eventually known as "pasteurization," in which wine, beer, vinegar, and milk were heated to kill bacteria and molds present within them.

Pasteur discredited the theory of spontaneous generation by demonstrating that microorganisms in a presterilized medium could be explained by outside germs. His research also showed that juice will not ferment if environmental yeasts are prevented from being deposited on grapes. By analogy, Pasteur believed that infectious diseases are probably caused by germs and that just as grapes can be protected against yeast, it might be possible to protect human beings against germs.

In 1879, Pasteur discovered that fowl cholera is caused by a type of bacteria now known as "Pasteurella." Chickens inoculated with a few drops of these bacteria would die. However, chickens inoculated with an old, weakened culture of Pasteruella did not die and were protected against a later inoculation with a more virulent culture. Through this chance observation, Pasteur discovered the principle of vaccination with attenuated pathogens. Because of Edward Jenner's work on vaccination, scientists knew that a weakened form of a disease could provide immunity to a more virulent version. However, whereas Jenner's vaccines used cowpox, a naturally occurring infection similar to but much less severe than smallpox, Pasteur's cholera and anthrax vaccines used artificially generated, weakened forms of disease organisms.

Pasteur grew the rabies virus in rabbits and then weakened it by drying the affected nerve tissue. He gave these artificially weakened diseases the generic name of "vaccines" to honor Jenner. In 1885, Pasteur conducted the first experimental rabies inoculations on a human. Joseph Meister, a 9-year-old boy who had been bitten multiple times by a rabid dog, was brought to Pasteur by his mother. Since the death of the child appeared inevitable, Pasteur attempted a method of inoculation that had proved consistently successful on dogs. Pasteur was not a licensed physician and could have faced prosecution for this. Ultimately, Meister's health improved after 12 inoculations, and Pasteur was hailed as a hero.

—Emily E. Anderson

See also Public Health, History of; Vaccination; Zoonotic Disease

Further Readings

Debré, P. (1998). *Louis Pasteur* (E. Forster, Trans.). Baltimore: Johns Hopkins University Press.

Geison, G. L. (1995). *The private science of Louis Pasteur.* Princeton, NJ: Princeton University Press.

Schwartz, M. (2001). The life and works of Louis Pasteur. *Journal of Applied Microbiology*, *91*, 597–601.

Pearson Correlation Coefficient

The sample Pearson product-moment correlation coefficient (r) is a measure of the linear association between two independent continuous variables, namely X and Y, measured on the same individuals or units. The values of the Pearson correlation coefficient measures the strength of the linear relationship between X and Y, while the sign of the correlation coefficient indicates the direction of the relationship between X and Y.

Given two continuous variables, X and Y, the Pearson correlation coefficient, r_{XY}, is obtained as the ratio of the covariance between the two variables over the product of the respective standard deviations.

$$r_{XY} = \frac{\text{Degree to which } X \text{ and } Y \text{ vary together}}{\text{Degree to which } X \text{ and } Y \text{ vary separately}}$$

$$= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2},$$

where $\bar{x}$ and $\bar{y}$ are the sample means for the variables X and Y, respectively.

The correlation coefficient is defined only if both the standard deviations are finite and both of them are nonzero.

Assumptions

To be able to correctly interpret and make valid inferences about the Pearson correlation coefficient, the following assumptions must hold:

- The observation $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ of X and Y are independent and identically distributed.
- The variables X and Y are jointly normally distributed with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2 and correlation ρ_{XY}.

Under these assumptions, the sample Pearson correlation coefficient r_{XY} represents a valid estimate of the correlation ρ_{XY}.

Properties

The Pearson correlation coefficient assumes values within the $(-1; 1)$ range. A correlation coefficient equal to -1 indicates a perfect negative linear relationship between two variables (see Figure 1a), while a correlation coefficient of 1 indicates a perfect positive linear relationship between two variables (Figure 1b). The correlation coefficient is equal to zero when either the two variables are independent (Figure 1c) or they are associated through a nonlinear relationship (Figure 1d).

Values in the middle of the $(-1; 1)$ range indicate the degree of linear dependence of the X and Y variables. A correlation coefficient >0 is called a positive correlation and indicates that the variables X and Y tend to increase or decrease together. A correlation coefficient <0 is called a negative correlation and indicates that increases in one variable correspond to decreases in the other. There are no rules on what defines a high or a low correlation, and the interpretation of the correlation coefficient depends on the context and the data on which it is calculated.

The Pearson correlation coefficient is not affected by changes in location or scale in either variable.

Although r_{XY} can be used to determine the degree of association between two variables, it is not a measure of the causal relationship between X and Y.

The value of r_{XY} can be affected greatly by the range of the data values, and extreme observations (outliers) can have dramatic effect on r_{XY}. Thus, the full range of scores should always be used when calculating the correlation coefficient. Extreme observations should be treated with caution in the calculation of the correlation coefficient.

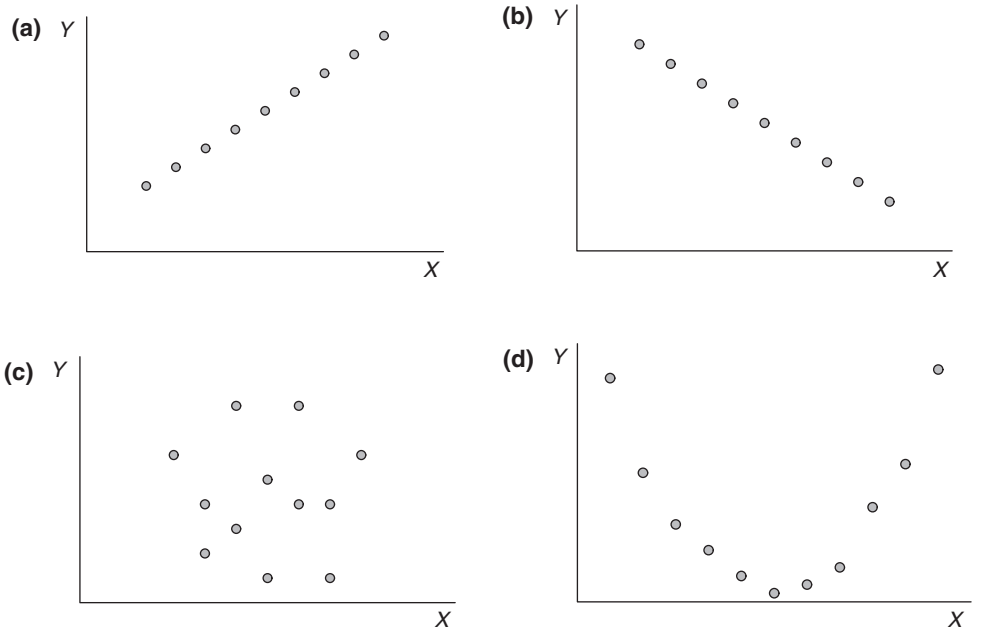


Figure 1 Correlation Coefficient Under Different Scenarios

Applications

The Pearson correlation coefficient can be used in a number of different applications.

- *Prediction.* Knowing that a strong relationship exists between two variables allows one to make an accurate prediction about one of them using the other.
- *Validity.* The Pearson correlation coefficient is often used to validate a new measurement scale. A high correlation between the new scale and an established one would assure that the new instrument is measuring what it is supposed to.
- *Reliability.* The Pearson correlation coefficient may also be used to establish reliability of an instrument. A high correlation between successive measures on the same individuals, for example, would indicate that the instrument is reliable.

Hypothesis Testing About ρ_{XY}

When interest is in testing the null hypothesis that there is no linear association between two continuous variables (H_0: $\rho_{XY} = 0$) against an alternative hypothesis that such association exists (H_A: $\rho_{XY} \neq 0$), then a Student's t approximation can be used to test this hypothesis. Under the assumption that the distribution of X and Y is bivariate normal, the test statistic

$$t^* = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}$$

follows a Student t distribution with $(n-2)$ *df*, under the null hypothesis. Thus, values of this test that exceed the critical value $t_{(n-2),(1-\alpha/2)}$ for a prespecified Type I error α would lead to reject the null hypothesis of no linear association or independence of X and Y.

When interest is in testing a more general null hypothesis that specifies a particular value for ρ_{XY}, H_0: $\rho_{XY} = \rho_0$ against the alternative H_A: $\rho_{XY} \neq \rho_0$, then an inference is carried out using the Fisher's transformation:

$$z = \frac{1}{2}\log_e\left(\frac{1+r_{XY}}{1-r_{XY}}\right).$$

When the sample size n is large, z is approximately normally distributed with mean

$$\varsigma = \frac{1}{2}\log_e\left(\frac{1+\rho_0}{1-\rho_0}\right)$$

and variance $\sigma^2(z) = 1/(n-3)$, under the null hypothesis.

The test statistic $\lambda = (z - z_0)\sqrt{n-3}$ will reject the null hypothesis if its value exceeds the critical value $z_{1-\alpha/2}$ on the standard normal distribution, for a prespecified Type I error α.

Confidence Intervals for ρ_{XY}

The upper and lower bound of a $100(1-\alpha/2)\%$ confidence interval for the Fisher's transformation z are given by $z_L = z - z_{1-\alpha/2}\sqrt{n-3}$ and $z_U = z + z_{1-\alpha/2}\sqrt{n-3}$.

A $100\%(1-\alpha)$ confidence interval for ρ_{XY} is given by

$$\rho_L = \frac{e^{2z_1}-1}{e^{2z_1}+1}; \rho_U = \frac{e^{2z_2}-1}{e^{2z_2}+1}.$$

Relationship Between r_{XY} and Regression Parameters

In a straight line model $r_{Y|X} = \hat{\beta}_1(S_X/S_Y)$, where $\hat{\beta}_1$ is the estimate of the slope, r_{XY} and $\hat{\beta}_1$ have the same sign.

When $r_{Y|X} = \pm 1$, then $\hat{\beta}_1 = \pm(S_Y/S_X)$. Thus, r_{XY} does not indicate the magnitude of the slope of the regression line. But in a straight line model, testing for $\rho_{XY} = 0$ is the same as testing for $\beta_1 = 0$.

When $\rho_{XY} = 0$, then $\beta_1 = 0$. There is no linear relationship between X and Y although a nonlinear relationship may exist.

In a linear model $r_{YX} = \text{sign}(\hat{\beta}_1)\sqrt{R^2}$, where R^2 is the model coefficient of determination. Note that although r_{XY} is related to the coefficient of determination, it should never be interpreted as a proportion (e.g., the proportion of Y predicted by X).

—Emilia Bagiella

See also Coefficient of Determination; Hypothesis Testing; Normal Distribution; Regression

Further Readings

Gravetter, F. J, & Wallnau, L. B. (1992). *Statistics for the behavioral sciences* (3rd ed.). St. Paul, MN: West.

Neter, J., Kutner, H. M., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models* (3rd ed.). Chicago: Irwin.

Peer Review Process

Peer review is a process whereby experts help judge the value of a work that they were not part of creating. Editorial peer review involves scientific or academic manuscripts submitted for publication or meeting presentation, while grant peer review involves review of funding applications. This entry focuses on editorial peer review of work submitted to scientific journals, although some of the issues discussed apply to other types of peer review as well. The primary function of editorial peer review is gate-keeping—selecting the best from a pool of submissions. In addition, peer review often involves constructive criticisms intended to improve a submitted work prior to publication. A common misunderstanding is that peer review validates the scientific integrity of a published article. Expecting such validation is unrealistic, as reviewers typically have access only to what the author or authors present in the manuscript. Important logistical and methodological decisions made along the way will be unknown to the reviewers, as will key details that the authors might omit. In essence, we must trust the authors.

Early forms of editorial peer review go back as far as the beginning of the 18th century, most notably within the Royal Societies of London and Edinburgh. The first modern peer review system was developed in the late 19th century by Ernest Hart, editor of the *British Medical Journal.* Yet it was only after World War II, as medical research methods became more sophisticated and journals became more selective, that peer review systems became institutionalized in the scientific and academic journals of the United Kingdom, the United States, and elsewhere.

In this context, the term *peer* is loosely interpreted to include subject-area experts, statisticians and methodologists, journal editors, editorial boards, and sometimes others, such as graduate students or nonexperts. Outside experts are usually sought because of their specialized knowledge in the submitted manuscript's content area or their advanced statistical or methodological skills. Ideally, these reviewers will have more expertise than the authors of the submitted work. But because of the proliferation of scientific journals and the many competing demands on experts' time, this specialized ideal is not always reached. It is therefore not uncommon for less qualified reviewers to assume this role.

A variety of peer review systems are employed across the world's 10,000 or so journals. Systems vary in their relative reliance on external reviewers, such as outside experts, versus in-house reviewers, such as editors and editorial boards. Acceptance rates can differ widely across journals, ranging from around 2% for the most highly selective journals to 90% and above for some electronic journals where publication space is less of an issue, and pay-per-page journals such as those that serve primarily as outlets for routine pharmaceutical studies.

Postpublication peer review is an important, if often neglected, type of review. This can include letters to the editor or full articles critiquing a published work, sometimes going so far as to involve reanalysis of the original data. While the need for postpublication review is now receiving more attention, challenges remain. Authors sometimes choose to ignore a published critique or respond minimally to peripheral issues in place of the specific criticisms made. Even when serious errors are detailed in a critique, retractions or corrections are the exception. Medline and other databases rarely link postpublication critiques to the original article, and literature reviews that cite a criticized work frequently ignore the critique.

Criticisms of Peer Review

Many criticisms of peer review have been raised by authors and reviewers as well as journal editors. A common criticism is that peer review is prone to bias. Bias can take various forms, including ad hominem bias, affiliation bias, ideological bias, and publication bias, among others. Ad hominem bias and affiliation bias are found when a review is influenced, either consciously or unconsciously, by knowledge of the author's identity or affiliation. Mixed evidence has been found on the presence and extent of these two biases. Some have argued that such influences are not necessarily biases but can be valid considerations in reviewing a manuscript. Yet the prevailing opinion remains that these potential influences are inappropriate for editorial peer review. It is for this reason that the norm is blinded review, where the author's name and affiliation are not known to the reviewer. The opposite is true, however, in grant peer review, where

authors' names and affiliations, as well as other detailed information about the authors' past experience and accomplishments, are typically an integral part of the reviewed application.

Ideological bias, where a reviewer's antecedent value-based views for or against an author's position unduly influence a review, has been demonstrated in several studies on the peer review process. Closely related is confirmation bias, the more general and well-documented tendency to less critically evaluate evidence that is consistent with one's existing beliefs. Numerous experimental demonstrations of these types of biases have led some to call for use of only the introduction and methods section of a paper in publication decisions, but this strategy has not yet been widely embraced or studied. Also related is publication bias, the selective publication of manuscripts based on the direction and magnitude of their results. It is widely accepted that research reporting statistically significant positive results is more likely to be published than research with null or nonsignificant results, and that this can, among other problems, lead to serious negative consequences for meta-analyses and other types of systematic reviews. Interestingly, studies have failed to support the belief that editorial decisions are biased in this matter. Instead, publication bias appears to result primarily from authors' reduced likelihood of submitting papers with null or negative results.

Other commonly voiced criticisms are that peer review is conservative and stifles innovation; is secretive and without accountability of reviewers to authors; suffers from low interrater reliability (typically found to be .30 or less); produces reviews of low quality; allows too many papers to slip thorough the system; frequently lacks adequate statistical and methodological review; is slow and expensive and delays publication; and is unscientific, with little or no evidence of its effectiveness. It is sometimes said that, like democracy, the peer review system is deeply flawed, yet better than all the alternatives. While no serious candidates for replacement of peer review have emerged, the many challenges to the current system have been increasingly publicized, and many suggestions for improvements in current practice have attracted attention.

Improving the Peer Review Process

One proposal to increase reviewer accountability and to generate more constructive reviews is to employ a signed rather than anonymous review process, in which reviewers' names are provided to the reviewed author. Several major journals, including the *British Medical Journal* (*BMJ*), now use signed review systems, while other journals encourage but do not require signed reviews. Yet most journals—and reviewers—do not yet support such a process, primarily due to reviewers' concerns about retribution, especially in the case of younger researchers who are fearful of criticizing senior colleagues. With reviewer recruitment already a difficult and time-consuming task, additional disincentives to accept review invitations would not be ideal.

Other suggestions for improving overall review and editorial decision quality include providing more training and support to reviewers and employing statistical and methodological review of all manuscripts. Reviewer training can take the form of workshops, tutorials, guides, apprentice models, and other strategies and should be targeted not only to existing reviewers but also to graduate students as future reviewers. Statistical and methodological review can be done concurrently with review by subject-area experts or subsequent to initial reviews by these experts. It can involve either in-house or external methodologists. Some journals, such as the *BMJ* and *Lancet*, already institutionalize separate methodological reviews, yet the majority of medical journals do not. In spite of the challenges in recruiting sufficient numbers of methodologists, this strategy has tremendous potential power, given that it has been found across several different fields that most published articles do contain nontrivial methodological flaws.

It is said that if peer review were not already widely in use and someone were to propose it today, the evidence of its effectiveness would be so lacking that it would not be given serious consideration. But evidence is accumulating as research on the peer review process becomes more widespread and sophisticated. An International Congress on Peer Review in Biomedical Publication is held every 4 years since 1989. These meetings are an attempt to encourage systematic research on the peer review process. And a large-scale collaborative mixed-methods study is currently underway to investigate the peer review processes at three prestigious medical journals: *Lancet*, *Annals of Internal Medicine*, and *BMJ*.

—Norman A. Constantine

See also Meta-Analysis; Publication Bias

Further Readings

Altman, D. G. (2002). Poor quality medical research: What can journals do? *Journal of the American Medical Association*, *287*, 2765–2767.

Cicchetti, D. V. (1991). The reliability of peer-review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, *14*, 119–134.

Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer-review. *Science*, *214*, 881–886.

Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist*, *45*, 591–598.

Giles, J. (2006). Journals submit to scrutiny of their peer-review process. *Nature*, *439*, 252.

Godlee, F., & Jefferson, T. (Eds.). (2003). *Peer review in health sciences* (2nd ed.). London: BMJ Books.

Rennie, D. R. (1998). Freedom and responsibility in medical publication: Setting the balance right. *Journal of the American Medical Association*, *280*, 300–302.

Rennie, D. R. (2002). Fourth international congress on peer review in biomedical publication. *Journal of the American Medical Association*, *287*, 2759–2758.

Schmelkin, L. P. (2003, August). *Peer review: Standard or delusion?* Division 5 Presidential Address delivered at the annual meeting of the American Psychological Association, Toronto, Ontario, Canada.

Shatz, D. (2001). *Peer review: A critical inquiry.* New York: Rowman & Littlefield.

Wagner, E., Godlee, F., & Jefferson, T. (2002). *How to survive peer review.* London: BMJ Books.

Web Sites

Fifth International Congress on Peer Review and Biomedical Publication home page: http://www.ama-assn.org/public/peer/peerhome.htm.

Pellagra

See Goldberger, Joseph

Percentiles

The percentile is a concept often used to summarize data and place the score or measurement taken on an individual into the context of a larger population. For any particular number p between 0 and 100, the pth percentile of a set of n measurements arranged in order of magnitude is the value that has at most $p\%$ of the observations below it and at most $(100-p)\%$ above it. Roughly speaking, the first percentile is the number that divides the bottom 1% of the data from the top 99%; the second percentile is the number that divides the bottom 2% of the data from the top 98%; and so on. Therefore, if a man has a body mass index score at the 98th percentile for his age, it means roughly 98% of men his age have a body mass index score lower than him, and only 2% have a higher score.

A percentile may be viewed as the division of a data set into 100 equal parts. Smaller groupings are often used; for instance, the median of a data set is also the 50th percentile, which specifies that at least half the observations are equal or smaller than it. Other commonly used percentile groupings include deciles, which divide a data set into tenths (10 equal parts), quintiles, which divide a data set into fifths (5 equal parts), and quartiles, which divide a data set into quarters (4 equal parts). Of these, quartiles are the most commonly used.

Percentiles are often used to describe large data sets; for instance, in the body mass index example above, the percentiles may have been calculated using a sample of thousands of American men. However, researchers sometimes want to calculate percentiles, quartiles, and so on, for a smaller data set, in which case the following procedure may be used to establish cut points.

1. Arrange the observations into increasing order from smallest to largest.
2. Calculate the product of the sample size n and proportion p you wish to include in each division (for quartiles, $p = 0.25$; for deciles, $p = 0.10$; etc.)
3. If np is an integer, say k, calculate the average of the kth and $(k+1)$th ordered values; if np is not an integer, round it up to the next integer and find the corresponding ordered value.

For example, a study of serum total cholesterol (mg/L) levels recorded the following ordered levels for 20 adult patients (the data were adapted by the author from data presented in Ott and Longnecker (2001, p. 83).

To determine the first quartile, we take $p = 0.25$, and calculate $np = (20)(0.25) = 5$, then the first quartile is the average of the fifth and sixth observations,

Table 1 Serum Total Cholesterol (mg/L) Levels Recorded the Following Ordered Levels for 20 Adult Patients

Ordered Observation	*Cholesterol (mg/L)*
1	133
2	137
3	148
4	149
5	152
6	167
7	174
8	179
9	189
10	192
11	201
12	209
13	210
14	211
15	218
16	238
17	245
18	248
19	253
20	257

Source: Adapted from data presented in Ott and Longnecker (2001, p. 83).

$$Q_1 = \frac{152 + 167}{2} = 159.5.$$

Therefore, data points falling at or below this cut point are in the first quartile of the data set. To calculate the cut point for the median, we take $p = 0.5$, and $np = (20)(0.5) = 10$, so the median is the average of the 10th and 11th observations, the

$$\text{median} = \frac{192 + 201}{2} = 196.5.$$

Values falling between 159.5 and 196.5 are in the second quartile. For the third quartile, we take $p = 0.75$, and $np = (20)(0.75) = 15$, so the cut point for the third quartile is the average of the 15th and 16th observations:

$$Q_3 = \frac{218 + 238}{2} = 228.$$

Values falling between 196.5 and 228 are in the third quartile, and values above 228 are in the fourth quartile.

A related concept, the interquartile range (IQR) of a data set is defined to be the difference between the upper and lower quartiles—that is,

$$IQR = Q_3 - Q_1.$$

The IQR measures the distance needed to cover the middle 50% of the data only, so it totally ignores the variability in the lower and upper 25% of data. Thus, the IQR does not provide a lot of useful information about the variability of a single set of measurements but can be quite useful when comparing the variability of two or more data sets. This is especially true when the data sets are skewed or contain outliers. For the above data set, IQR = 228 − 159.5 = 68.5.

If we need to calculate the 87th percentile of this data set, we can take $p = 0.87$ and calculate $np = (20)(0.87) = 17.4$. Because this is not an integer, we take the next largest integer, 18, so that the 18th ordered observation, 248 is at the 87th percentile.

—Renjin Tu

See also Box-and-Whisker Plot; Histogram; Measures of Central Tendency; Measures of Variability

Further Readings

Johnson, R. A., & Bhattacharyya, G. K. (2006). *Statistics principles and methods* (5th ed.). Hoboken, NJ: Wiley.

Ott, R. L., & Longnecker, M. (2001). *Statistical methods and data analysis* (5th ed.). Pacific Grove, CA: Duxbury.

Person-Time Units

It is common in medicine and epidemiology to express the frequency of occurrence of some event in terms of the number of events per person-time unit; for instance, the number of complications per 100 patient-days or the

number of deaths per 100,000 person-years. A good example is the incidence rate, also known as the incidence density or force of morbidity or mortality. The incidence rate is calculated as

$$\frac{\text{No. of new cases of a disease}}{\text{Total person-time of observation}}.$$

The numerator is always the number of new cases of the disease in the time period studied, and the denominator is the sum of the time of observation for all the subjects in the study. This fraction is usually converted to a standard unit such as cases per 100 to facilitate comparisons.

Person-time units are used when the subjects in a study have been observed for different lengths of time and have, therefore, been at risk for the event in question for longer or shorter times. Using a person-time denominator allows each subject to contribute to the denominator in proportion to the length of time they were observed and allows comparison across units (e.g., complication rates in different hospitals or mortality rates in different countries).

Consider the following example. We want to compare the quality of care for a particular condition in a particular year at two hospitals using the mortality rate for that condition. Table 1 presents hypothetical data on eight patients treated for this condition at two hospitals and includes the days observed (i.e., the number of days they were in that hospital and thus eligible for the event of death to occur at that hospital). Assuming this is the total patient population treated at those hospitals for that condition in the year under study, we can see that in Hospital A, two patients died (because they have a "Y" for "yes" in the "Event" column), whereas in Hospital B, only one patient died. We might interpret this as meaning that Hospital A was somehow less safe or had a lower quality of care for this condition because they had two deaths per year versus one death per year for Hospital B, but we would be ignoring the fact that Hospital A had more patients at risk of death from this condition during this time period.

A more sensible comparison would be made using the mortality rate per 100 patient-days. In this case, Hospital A had 2 deaths per 100 patient-days, while Hospital B had 1 event in 20 patient-days or a rate of 5 deaths per 100 patient-days. By the criterion of mortality rate, Hospital A seems to be doing a better job in treating this condition. Of course, this example is greatly simplified, and hospital-to-hospital comparisons are generally done after correcting for expected mortality and morbidity, considering factors such as patient mix, but it illustrates why person-time units are commonly used in epidemiology.

Table 1 Data to Calculate Complication Rate per 100 Patient-Years for Two Hospitals

Hospital	*Patient*	*Days Followed*	*Event?*
A	1	10	Y
A	2	20	Y
A	3	25	N
A	4	30	N
A	5	15	N
Total Person-Days		100	
B	6	10	Y
B	7	5	N
B	8	5	N
Total Person-Days		20	

—*Sarah Boslaugh*

See also Incidence; Mortality Rates; Public Health Surveillance; Rate

Further Readings

Hennekens, C. H., & Buring, J. E. (1987). Measures of disease frequency. In S. L. Mayrent (Ed.), *Epidemiology in medicine* (ch. 4, pp. 54–98). Boston: Little, Brown.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott-Raven.

Pharmacoepidemiology

Pharmacoepidemiology is the study of the use and effects of medical products (drugs, biological products, and medical devices) in human populations. One of the newer branches of epidemiology, pharmacoepidemiology has emerged as a unique field of study in parallel with the development of large,

comprehensive, health care databases. However, it is not the reliance on large databases, but the nature of medical products as exposures that truly differentiates pharmacoepidemiology as a subspecialty of epidemiology. First, medical products are regulated by government entities. They are approved for a particular use or indication with dosing, labeling, and monitoring requirements. Second, exposures to medical products are made consciously for the treatment of a known medical condition or to prevent or delay the occurrence of a disease. The regulatory nature of pharmaceutical products drives the type and timing of studies and often the source of funding and perspective as well. The nonrandom nature of treatment decisions can introduce bias and, as such, drives many research design decisions and ultimately affects interpretation of study results. The entry is organized into the following sections: (1) the drug development and approval process, (2) adverse drug events, (3) postmarketing safety, (4) risk management, and (5) training and careers.

Drug Development and Approval

Approval of a pharmaceutical product for marketing in the United States requires extensive testing and a determination by the Food and Drug Administration (FDA) that its benefits outweigh its risks for the intended use or indication. Before a medicine is tested in humans, it is studied and evaluated extensively in the laboratory and in animal models. The type and extent of testing depends on the nature of the chemical being studied and the indication for use that the sponsor is pursuing. Before clinical testing can begin, the product sponsor submits an investigational new drug application (IND) to the FDA. An IND summarizes the preclinical study results and contains a detailed plan for clinical testing.

Clinical testing is divided into three unique phases. Phase I is the first use of a medication in humans and, as such, is usually limited to a small group of healthy individuals. The primary purpose of this first phase is to determine the safety of the drug in humans. Within Phase I, scientists seek to determine a safe dosing range, how the drug is handled by the body (pharmacokinetics), and its action or effect at various dosages (pharmacodynamics). Phase II studies are generally small clinical trials in which a drug is tested in patients to further characterize its safety profile and determine which dosages and dosing schedules will be tested for approval. Phase III clinical trials are the randomized studies in the intended patient population. In Phase III trials, the new drug is compared with a placebo, or in cases where it is unethical to deny or delay treatment, an alternative treatment—typically the standard of care. The use of such "active-control" trials relies on historical data for assurance that the alternative treatment is more effective than placebo, while testing for equivalence between the new and control drugs.

From 1,000 to 5,000 patients are typically exposed to a new drug on submission of a New Drug Application (NDA) to the FDA. If the FDA determines that the NDA is complete, a team of reviewers evaluates all the study results and determines whether or not the medication can be marketed in the United States. It takes approximately 15 years and $800 million to take a drug from discovery through approval, according to the Pharmaceutical Research and Manufacturers Association (PhRMA).

Adverse Drug Events

The basis for approval of a new pharmaceutical agent is that it is safe and effective for its intended or labeled use. This does not mean that it is absolutely safe but that relative to its benefits as established in the clinical trials, the risks are acceptable. At the time of approval, the FDA may require additional studies to follow-up on outstanding questions. These postmarketing studies are often called Phase IV studies or commitments. Additionally, all sponsors are required to monitor and report any serious adverse events that may be related to the use of the drug. Pharmacovigilance refers to the process of collecting, monitoring, and evaluating adverse events reports.

Often, adverse effects of a drug are related to its pharmacokinetics (how it is processed by the body) or pharmacodynamics (how it affects the body). A drug may cause adverse events by the same mechanism that provides the intended therapeutic benefit. For example, an agent that effectively prevents blood clotting can be a contributing cause of excessive bleeding and subsequent hemorrhage.

Genetic variation, concomitant drugs, and other medical conditions can contribute individually or in unison to increase the risk of an adverse event. For example, the amount of an active metabolite can be increased to toxic levels by genetic polymorphisms that impede its breakdown, renal or liver impairment

(depending on the route of metabolism), and/or the use of a concomitant drug that successfully competes for binding sites. These "Type A" adverse drug events are considered predictable because they are based on known properties of the drug. Less common are Type B adverse events, those that are idiosyncratic or unpredictable such as allergic or immunologic reactions. Type B events include anaphylaxis and Stevens Johnson syndrome.

Extensive preclinical and clinical testing ensures that on marketing approval, the more common adverse events, occurring at rates of 1% or greater, are well characterized. As approval is conditioned on the benefits of therapy outweighing the risks, commonly occurring adverse events are typically nonserious. Preapproval studies are limited not only in the total number of persons exposed but also on the duration of exposure and the diversity of the patient populations. Therefore, rare adverse events, occurring at rates below 1/10,000 persons, and events associated with extended duration use are often not identified until after a drug is marketed. New safety problems, for example drug-drug or drug-gene interactions, may also emerge on use within a larger and more diverse patient population.

Postmarketing Safety

Once a product is marketed, manufacturers are required to monitor, evaluate, and submit reports of serious adverse events potentially caused by their products to the FDA. Health care professionals and the general public may voluntarily report problems to the manufacturer or directly to the FDA through their MedWatch® program. The FDA maintains databases of these reports: the Adverse Event Reporting System (AERS) for drugs and biologics, the Vaccine Adverse Event Reporting System (VAERS) for vaccines, and the Manufacturer and User Facility Device Experience Database (MAUDE) for devices.

The initial evaluation of a potential product safety problem has many similarities to outbreak investigations, beginning with the development of a case series of adverse event reports. Once a case definition is created, ineligible reports are excluded and the remaining reports evaluated in terms of person, place, and time. A crude reporting rate may be estimated and a causal assessment conducted. Case series investigations guide the decision to conduct further research, either observational or experimental, and provide information about exposures, latency period, and potential risk factors.

There is no magic number of case reports that identify a real problem. The volume of adverse event reports varies over time and is typically highest during the first several years of marketing. Labeling changes, marketing programs, and publicity have all been shown to affect reporting levels. Also, the uniqueness and severity of the adverse event may also affect reporting. For example, an abnormal laboratory test for liver enzymes is less likely to be reported as an adverse event than a case of acute liver failure. Similarly a unique syndrome of birth defects is more likely to be associated with an exposure and reported because of its uniqueness, than a more common pregnancy outcome such as a spontaneous abortion. A clinical expert evaluates each new report of an adverse drug event within the context of previous reports. The application of data mining techniques, which use statistical algorithms to aid in the identification of new and otherwise unexpected adverse events, shows promise as an additional screening tool. Data mining algorithms identify drug-event or drug-drug-events that occur in excess of expected rates.

Once a hypothesis is formulated, the feasibility for conducting an epidemiological study must be determined. In rare situations, one or more cases can be so definitive that no further information is needed to prove a causal link between exposure and adverse outcome. Pharmacoepidemiological investigations use the methods and study designs of epidemiology. The field relies heavily on the use of databases such as health care service utilization data, automated medical records, and health care and pharmaceutical claims. Databases have a distinct advantage over original data collection in the speed it takes to complete a study as well as the size of the underlying population. The terms *historic cohort*, *nonconcurrent prospective cohort*, and *retrospective cohort studies* are used synonymously to describe research studies that rely on data from health care or claims databases to identify their study cohort and create variables to characterize subjects by exposure, outcome, demographics, and so on.

Each database has distinct characteristics that can affect internal and external validity. For example, claims databases are established to provide reimbursement for a covered service. As such, services that are not covered or are charged to another insurer (e.g., patients who are covered by both Medicaid and Medicare) may not be captured. Medical practice and changes in coding practices may also influence the validity of diagnostic codes. The likelihood that

a patient was exposed to a medication also varies across and within databases. An assumption of an exposure is based on a record in a database though there is no proof that the patient ever took the medication. A claim for reimbursement by the pharmacy for a filled prescription is a step closer to a potential exposure than a record of a prescription having been written by the physician. Two filled prescriptions, one following the other at an interval equivalent to the allotted days supply, further increases the likelihood of an exposure compared with a single filled prescription. Knowledge of database characteristics, local medical practice, and the covered population, and how these unique database characteristics influence the potential to be prescribed a particular drug and the likelihood of diagnosing the study outcome, are critical to designing a valid epidemiological study.

Physicians make a variety of treatment choices, first whether or not to prescribe a medication, and if so, which medication and dosage schedule to use. These decisions are not random. Thus, therapeutic decision making can introduce a number of potential biases that, if not accounted for in the study design, can obscure a true association between exposure and outcome. "Confounding by indication" refers to a bias introduced into a study when the choice of treatments is in some way (noncausally) related to the outcome being studied. Consider an observational study comparing rates of oral cancer among men with and without heavy use of mouthwash. Because heavy mouthwash users (exposed) were predominantly smokers and the unexposed predominantly nonsmokers, the statistical association with mouthwash use and oral cancer is confounded by the "indication" for the study exposure. Confounding by indication is best accounted for in the design of a study. In the mouthwash example, the study population could be limited to heavy smokers. Protopathic bias occurs when a particular treatment is used to treat a symptom or other factor that is directly associated with the risk of the outcome under study. For example, an antidiabetic agent may be associated with an increased risk of birth defects because gestational diabetes itself increases the risk of adverse fetal outcome.

Risk Management

In March 2005, the U.S. FDA released an industry guidance document on the Development and Use of Risk Management Action Plans (RiskMAP) simultaneously with guidances on premarketing risk assessment and postmarketing pharmacovigilance practices. These guidance documents articulated the need to articulate and assess known and potential risks of a product at the earliest stages of preclinical development. They outline a framework to identify, evaluate, and monitor safety throughout a product's life cycle. This life cycle approach, particularly the anticipation and evaluation of the risk to cause serious adverse events, such as blood dyscrasias, liver toxicity, and Q-T prolongation, is designed to save lives by identifying and characterizing safety issues earlier in the process. Potentially, this may allow important therapies, which would have previously been withdrawn for safety reasons, to be available to patients for whom the benefits do outweigh the risks.

At the time of approval for marketing, the FDA may require the sponsor to conduct further studies, often called Phase IV studies or Phase IV commitments, as they come after the definitive Phase III clinical trials conducted for approval. Phase IV studies are done to answer questions that arise during clinical development but are outside the scope of the Phase III trials. This might include questions on the impact of an extended duration of use or the safety of use in special populations, such as patients with common chronic conditions or pregnant women.

The ultimate risk management decision is the FDA's approval (or withdrawing of approval) to market a medication in the United States. Product labeling is the next line of risk management describing approved indications as well as contraindications, dosing and prescribing considerations, and known safety issues. There are additional levels of regulation that can be used to improve the benefit-risk ratio for products with unique risks that may be avoided or managed with appropriate knowledge and/or action. In escalating levels of intervention, this includes education and outreach, informed consent or prescribing checklists, and performance-based access. The most stringent level of regulation, performance-based access, requires certain conditions for medication access such as laboratory assessment to rule out pregnancy before a prescription for a teratogen such as isotretinoin or thalidomide can be filled. There is a great need to evaluate ongoing risk management programs as well as develop and assess new methods to communicate product risk and risk management to patients and providers.

Training and Careers

The practice of pharmacoepidemiology requires sound knowledge of epidemiology, including both its methods and limitations. Clinical expertise in medicine and pharmacy is important as is an understanding of pharmacology. Practitioners have come to pharmacoepidemiology from many disciplines, clinical and nonclinical. This same diversity is also seen in training programs, which are based within schools of public heath, medicine, pharmacy, or in multidisciplinary programs. Training is almost exclusively at the graduate and postgraduate levels.

Pharmacoepidemiologists apply epidemiological methods to study the risks, benefits, and utilization of drugs, vaccines, biologics, and/or devices. They are employed in three principal areas: industry, regulatory agencies, and academia. An industry pharmacoepidemiologist may work for a pharmaceutical company or a company that provides consulting, pharmacovigilance support, research, or some combination of these services. Within a regulatory agency, a pharmacoepidemiologist might evaluate adverse drug events, conduct independent research, evaluate industry studies, oversee external research, and develop standards. Academic researchers typically teach, train graduate and postgraduate students, conduct research, and consult. There are a number of professional societies that provide forums for the exchange of new knowledge, professional development, and training in pharmacoepidemiology. The International Society for Pharmacoepidemiology (ISPE) is the most specific professional organization within the field. Through the society, ISPE members develop and promulgate standards for pharmacoepidemiology research, provide training programs, provide forums for scientific exchange, advertise jobs and training programs, and recognize expertise through their Fellow program. There are a number of more broad-based professional societies that include Pharmacoepidemiology—for example, the Society for Epidemiologic Research (SER), Drug Information Association (DIA), International Society for Pharmacovigilance (IsoP), and the International Society of Pharmacoeconomics and Outcomes Research (ISPOR). ISPE maintains a list of hyperlinks to these and other professional associations, as well as Pharmacoepidemiology-related resources, including training programs, government agencies, research centers, professional journals, and tools.

—Sheila Weiss Smith

See also Case Reports and Case Series; Clinical Trials; Confounding; Food and Drug Administration; Secondary Data

Further Readings

Begaud, B. (2000). *Dictionary of pharmacoepidemiology*. New York: McGraw-Hill.

Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. (2005). *Guidance for industry development and use of risk minimization action plans*. Rockville, MD: U.S. Department of Health and Human Services, Food and Drug Administration. Retrieved July 24, 2007, from http://www.fda.gov/cber/gdlns/riskminim.pdf.

Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. (2005). *Guidance for industry good pharmacovigilance practices and pharmacoepidemiologic assessment*. Rockville, MD: U.S. Department of Health and Human Services, Food and Drug Administration. Retrieved July 24, 2007, from http://www.fda.gov/cder/guidance/6359OCC.htm.

Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. (2005). *Guidance for industry premarketing risk assessment*. Rockville, MD: U.S. Department of Health and Human Services, Food and Drug Administration. Retrieved July 24, 2007, from http://www.fda.gov/cder/guidance/6357fnl.htm.

Goldman, S. A. (2004). Communication of medical product risk: How effective is effective enough? *Drug Safety*, *27*, 519–534.

Goldman, S. A., Kennedy, D. L., & Lieberman, R. (2005). *Clinical therapeutics and the recognition of drug-induced disease*. MedWatch continuing education article. Rockville, MD: Staff College, Center for Drug Evaluation and Research, Food and Drug Administration. Retrieved July 24, 2007, from http://www.fda.gov/medwatch/articles/dig/ceart.pdf.

Hartzema, A. G., Porta, M. S., & Tilson, H. H. (Eds.). (1998). *Pharmacoepidemiology: An introduction* (3rd ed.). Cincinnati, OH: Harvey Whitney.

International Society for Pharmacoepidemiology. (2003). *Hyperlinks of interest*. Retrieved September 8, 2006, from http://www.pharmacoepi.org/resources/links.cfm.

Nelson, R. C., Palsulich, B., & Goglak, V. (2002). Good pharmacovigilance practices technology enabled. *Drug Safety*, *25*, 407–414.

Pharmaceutical Research and Manufacturers Association. (n.d.) *Drug development in the United States*. Retrieved September 28, 2006, from http://www.phrma.org/innovation.

Strom, B. (Ed.). (2005). *Pharmacoepidemiology* (4th ed.). West Sussex, UK: Wiley.

Waning, B., & Montagne, M. (2000). *Pharmacoepidemiology: Principles & practice*. New York: McGraw-Hill.

Wysowski, D. K., & Swartz, L. (2005). Adverse drug event surveillance and drug withdrawals in the United States, 1969–2002: The importance of reporting suspected reactions. *Archives of Internal Medicine*, *165*, 1363–1369.

Phenotype

The *phenotype*, a term used extensively in the field of genetics since its development in the early 1900s, comprises the characteristics, traits, values, or abnormalities that we observe, measure, test, or evaluate in an individual. As such, the phenotype may include behavioral, biochemical, clinical, molecular, morphological, physical, and physiological characteristics, as well as the presence or absence of disease. In genetics, we think of the phenotype as the outcomes and results that are determined by the interplay between the genotype and environmental factors. It is important to consider that the environment can also include the so-called genetic environment—that is, the genes at other genetic loci whose products might interact with a specified gene or its product during development or during processes later in life.

With respect to genetic disease, the phenotype includes the clinical signs and symptoms of the disease, clinical features that are observed or measured on an individual as the disease progresses, and various disease outcomes such as impairments, disabilities, and quality of life measures. Clinical diagnosis of a genetic disease usually occurs following the initial recognition of the phenotypic manifestations of the disease, which then spurs the physician to request genetic testing for confirmation and specification of the genetic mutation. For example, the muscular dystrophies comprise a group of hereditary diseases characterized by the progressive wasting of skeletal and sometimes cardiac and smooth muscles. The Muscular Dystrophy Association recognizes 9 different muscular dystrophies among 34 other various types of diseases affecting neuromuscular function. The presence of progressive muscle wasting and weakness often suggests the diagnosis of a form of muscular dystrophy. The presence of these symptoms in early childhood further narrows the diagnostic consideration to (1) Duchenne muscular dystrophy (DMD), (2) Becker muscular dystrophy (BMD), (3) one of the various limb-girdle muscular dystrophy (LGMD), or (4) several other less common dystrophies. The presence of progressive weakness with a predominantly limb and trunk distribution with weakness occurring more in the lower versus upper limbs suggests the possibility of either DMD, BMD, or one of the LGMD, at which point genetic testing for mutations in the appropriate genes underlying these disorders may be ordered based on the phenotypic manifestations.

In DMD, genetic mutations in the dystrophin gene lead to incorrect coding for the protein dystrophin with varying degrees of severity. The decreased expression of the fully functioning form of dystrophin leads to a loss of internal muscle structure, leading to slowly progressive muscle weakness and wasting, the clinical phenotype of DMD. The phenotypic expression of a disease may also correlate with the severity of the genetic mutation. This information may enhance or even redefine what we know about disease. DMD and BMD were originally considered to be two separate genetic diseases because of the variation in clinical disease symptom severity and prognosis of death between them. It is now known that DMD and BMD represent a spectrum of the phenotypic expression of differing mutations in the dystrophin gene. DMD (clinically severe, leading to death in the third decade of life) and BMD (milder clinical disease than DMD with significantly improved survival) are now classic examples of how minor variations in genotype may cause major variations in phenotype.

Much current research focuses on the relationship of the genotypes in individuals affected with a genetic disorder to the resultant phenotypic outcomes. Different genotypes of a genetic disorder could be predictive of the severity of the disease among affected individuals or even the phenotypic manifestations of the disease. For example, if we could predict the resultant phenotype of individuals with mutations in the dystrophin gene, specifically whether and when common secondary complications, such as cardiomyopathy, scoliosis, and pulmonary disease, will manifest, then it is possible that physicians may provide more effective prevention and treatment. It has been well documented that improved clinical management of these complications does prolong the life and improve the quality of that life in patients with DMD. Understanding the relationship between the genotype and resultant phenotype will likely prove to enhance our understanding of these disorders and our abilities to care for those affected.

New methods for detecting mutations in specific disease-causing genes have enabled the identification of mutant genotypes heretofore not possible in many genetic disorders. The hope has been that this would

allow us to predict phenotypic variance that in turn would improve prognosis and treatment of genetic diseases. However, as with most diseases, the phenotypic outcomes that are observed in the clinical evaluation are the result of a complex and lengthy series of biological events, both genetically and environmentally influenced. These events occur during development and well into adult life, adding to the complexity of the situation. By and large, what has followed is the realization of the complexity of the phenotypic expression in genetic diseases. It is this realization that has caused many geneticists to question the common ways that we have classified some of the genetic disorders, and there is currently a trend to view even single gene (monogenic) traits as complex and multifactorial. Contributing to this changing perspective is the realization that if we examine the multitude of phenotypic outcomes in any genetic disorder, we find a variety of factors that can contribute to the widely different phenotypic expressions among affected individuals. What accounts for differences in phenotypic outcomes include, but are not limited to, variation among the alleles for a single gene, the effects of interaction between the disease gene and other genes (modifier genes), and the effects of environmental exposures on the products of disease genes during development. Even for simple monogenic disorders, the biology remains complicated and multifaceted.

—F. John Meaney, Jennifer Andrews, and Timothy Miller

See also Association, Genetic; Gene; Gene-Environment Interactions; Genotype; Mutation

Further Readings

Nussbaum, R. J., McInnes, R. R., & Willard, H. F. (2001). *Thompson and Thompson genetics in medicine* (6th ed.). Philadelphia: W. B. Saunders.

Scriver, C. R., & Waters, P. J. (1999). Monogenic traits are not so simple: Lessons from phenylketonuria. *Trends in Genetics*, *15*, 267–272.

Strachan, T., & Read, A. P. (2004). *Human molecular genetics 3* (3rd ed.). New York: Garland Science.

Physical Activity and Health

Prior to 1900, virtually every aspect of life, including transportation, work, food preparation, and caring for one's property required physical exertion or movement. However, beginning with the Industrial Revolution, an immense number of inventions have provided convenience and relief from physical effort. This has created an environment in which people can be almost completely sedentary on any given day. This changed environment has unintended consequences as we are now beginning to fully understand the negative impact a sedentary lifestyle can have on health.

The benefits of physical activity (PA) have been extolled throughout Western history, but it was not until the latter half of the 20th century that scientific evidence supporting these beliefs began to accumulate. A significant amount of this evidence has come from prospective epidemiology studies involving large numbers of people followed for several years in which the relationship between PA and various health outcomes have been documented.

This entry summarizes the evidence from such studies to provide an understanding of the association between PA and different health benefits and risks. Where sufficient evidence exists, answers to the question, "How much physical activity is enough?" are provided. However, since the vast majority of the epidemiology studies have involved participants above 18 years of age, only evidence on adults is considered. Research with adults shows that regular PA independently confers significant health benefits as indicated by marked reductions in the risk of developing several chronic diseases that are today's leading causes of death and disability. In some cases, the optimal dose of PA is unclear. However, current data suggest that at least 30 min/day of moderate intensity PA is very beneficial and, in some cases, more is better. The resultant health benefits are available to all persons across the adult life span and, therefore, a lifetime of PA should be a priority.

Terminology

Physical activity is defined as bodily movement produced by skeletal muscle that increases energy expenditure above the resting level. As such, PA involves all movement associated with occupational, household, leisure time, recreational, sport, or transportation activities. *Exercise* is a subcategory of PA and is planned, structured, repetitive, and for the purpose of improving or maintaining one or more components of physical fitness. Both PA and exercise can be categorized by type, duration, frequency, and intensity.

Physical fitness is the ability to carry out daily tasks with vigor and alertness, without undue fatigue. Health-related fitness includes cardiorespiratory (aerobic) endurance, muscle endurance, muscle strength, flexibility, and body composition. *Health* is a human condition with physical, social, and psychological dimensions. Positive health is associated with a capacity to enjoy life and withstand challenges, not just the absence of disease. Negative health is associated with morbidity and, sometimes, premature mortality.

The intensity of PA is known to influence the health benefits derived. Thus, a correct knowledge of this component is helpful. *Moderate intensity* PA requires 3 to 6 times as much energy as rest. This is equivalent to brisk walking. *Vigorous intensity* PA requires 7 times as much energy as rest, or greater. This is equivalent to jogging. *Energy expenditure* is a product of the frequency, intensity, and duration of PA and is commonly reported as kilocalories per week (kcal/week).

Physical Activity and Mortality

Does PA add years to life? The resounding answer is yes. A number of studies indicate that physically active men and women live longer than sedentary people, meaning that the health benefits of PA outweigh the risks.

The cumulative evidence indicates a linear reduction in mortality risk with an increased level of PA. On average, a threshold of about 1,000 kcal/week (4,200 kJ/week) of energy expenditure from PA is associated with a 20% to 30% reduction in mortality from all causes. This amount of energy expenditure is attainable by walking briskly for 30 min/day. Further risk reduction may be observed with energy expenditure greater than 1,000 kcal/week.

The protective effect of cardiorespiratory fitness has also been reported with the most fit men and women having 70% to 80% lower death rates than the least fit men and women. Interestingly, adults in the next to lowest fitness group have exhibited 48% to 60% lower rates of all-cause death than the least fit group, indicating that even modest increases in aerobic fitness promote longevity. The amount of PA required for achieving even minor improvements in aerobic fitness and significant risk reductions in premature death equates to 130 to 140 min/week of walking, 100 to 130 min/week of aerobics, or 90 min/week of jogging.

Of special note is that the research indicates that higher-activity and higher-fitness groups had lower risk of death whether or not they smoked, had high cholesterol, had high blood pressure, had high blood glucose, had a family history of heart disease, had a healthy baseline examination, or were overweight. Thus, PA and fitness can improve health for men and women, regardless of their health and risk factor status.

One limitation to most of the epidemiology studies is that PA is only measured once. However, it can change over time. Stronger evidence that PA and fitness *cause* health improvements comes from studies illustrating that changes in PA predict changes in risk of mortality. In the Harvard Alumni Study, more than 10,000 men aged 45 to 84 years reported their PA at two time periods with subsequent deaths monitored. Compared with those who remained inactive at both times, those who became active decreased their risk of dying by 15%. Those who were active the first time but became inactive by the second assessment increased their risk of death by 10%. This cause and effect relationship has been confirmed with research linking changes in aerobic fitness to changes in mortality. Another important finding is that gains in longevity are noted for persons across a large age range demonstrating that changes in PA or fitness at any age are beneficial.

The evidence is irrefutable for regular PA significantly reducing the risk of premature death. To put this in perspective, one can determine the number of deaths for which sedentary lifestyle is responsible. Considering the most common causes of death, more than 250,000 deaths in the United States could be prevented *each year* if sedentary lifestyle was eliminated. This figure accounts for approximately 23% of all deaths as compared with those caused by smoking (33%), obesity (24%), and high cholesterol (23%). As such, sedentary lifestyle is one of the most important public health challenges facing us today.

Physical Activity and Cardiovascular Disease (CVD)

Physical activity reduces the risk of premature death by primarily decreasing the risk of cardiovascular diseases, which are the leading causes of death in the United States and other industrialized countries. Many associate CVD with men, but heart disease is also the leading cause of death in women, resulting in approximately 500,000 deaths in U.S. women each year.

Coronary Heart Disease

Coronary heart disease (CHD) resulting from plaque accumulating in the coronary arteries is the most deadly form of CVD, and there is overwhelming evidence that PA and aerobic fitness are protective factors. Results show that the least active or fit persons had an 80% higher risk of dying from CHD than the most active or fit groups. In addition, the largest risk reduction typically occurs between persons in the least active or fit group and persons in the next highest activity or fitness group, once again indicating that even slight increases in activity or fitness confers significant benefits. For example, women have experienced 20% to 50% reductions in risk of CHD death with as little as 1 hr/week of walking. As with all-cause mortality, the association between PA or aerobic fitness and CHD risk is significant for both men and women across different races, levels of body fatness, preexisting medical conditions, and age groups. These results signify the capacity to use PA as an effective intervention for both primary and secondary prevention of CHD.

Similar to findings for premature death from all causes, *changes* in PA and fitness have been shown to affect CHD risk. Men who were unfit or inactive at baseline but increased in fitness or activity level over time reduced their risk of CHD by 52% and 45%, respectively, compared with men who remained unfit or inactive at both assessments.

There appears to be a window of protection from CHD death by expending 750 to 2,000 kcal/week through *moderate* intensity, dynamic, endurance PA (such as walking or jogging 7.5–20 miles/week). In the absence of other activity, at least 1 hr/week of intermittent hard physical labor also significantly reduces the risk of CHD. Most evidence shows that the largest reductions occur with moderate levels of activity or fitness as compared with those who are least active or fit. Therefore, something is better than nothing, but exactly how much is enough remains unclear. Regardless, in very inactive or unfit persons, brisk walking daily for at least 30 min should stimulate fitness gains and energy expenditure associated with CHD benefits.

Stroke

Recent reviews on PA and stroke risk found that moderately or highly active individuals had lower risk of stroke occurrence or mortality than did low active persons. Being moderately or highly active during leisure time was associated with a 15% to 20% and 20% to 27% lower risk of total stroke occurrence and mortality, respectively, compared with being inactive. Although there were relatively few studies available, results showed that being moderately and highly active at work was associated with a 36% and 43% lower risk of stroke, respectively, compared with being sedentary. It has also been demonstrated that daily commuting PA on foot or by bicycle was modestly associated with a decreased risk of stroke in both men and women. The risk of having a stroke was decreased by 8% and 11% in persons accumulating 1 to 29 min/day and >30 min/day of active commuting, respectively, compared with those who did not actively commute to work. Overall, the results indicate that moderate PA achieved through a variety of daily activities is protective against stroke and additional benefits may be realized with greater than moderate amounts of PA. However, more information is needed to develop specific recommendations with regard to the intensity, duration, and frequency of PA associated with a meaningful reduction in stroke risk.

Hypertension

Hypertension (HTN) is a significant risk factor for all-cause and CVD death, stroke, CHD, heart failure, kidney malfunction, and peripheral vascular disease. Fortunately, PA can serve as a low cost intervention in the prevention and management of HTN.

Epidemiology studies reveal that regular PA has potential for reducing or preventing mild hypertension. Participation in vigorous sports was associated with a 19% to 30% reduction in risk of developing HTN in U.S. men. Investigations with Japanese and Finnish men have also demonstrated significant inverse associations between baseline levels of commuting and leisure-time PA and future HTN. High levels of aerobic fitness have also been reported to reduce risk of HTN by 50% to 90% as compared with the lowest levels of fitness. None of the studies in women have observed significant relationships between PA and future HTN, although one did report a 30% lower risk for developing HTN in active versus sedentary women. In the only study to date including black men, PA was not associated with HTN risk; however, more studies are needed before definitive conclusions can be made.

Clinical studies indicate that regular exercise is effective for reducing systolic blood pressure by about 6 to 11 mmHg and diastolic blood pressure by about 6 to 8 mmHg in men and women with mild HTN. The recommendation to achieve such benefits is to accumulate at least 30 min of moderate intensity endurance type PA on most (at least 5) days of the week. For now, it doesn't appear that PA of a higher intensity confers any further benefits above those attained with moderate intensity activity.

Physical Activity and Cancer

Cancer is the second leading cause of death in many industrialized nations with the most common cancers being lung, colon, breast, and prostate. Because each cancer is likely to have somewhat different causal factors, the protective effects for PA have been examined for specific types of cancer.

Colon Cancer

The cumulative evidence from more than 50 studies clearly shows that physically active men and women have about a 30% to 40% reduction in risk of developing colon cancer, compared with inactive persons. It appears that 30 to 60 min/day of moderate to vigorous intensity PA is required to decrease risk. The optimal amount, intensity, duration, and frequency of PA associated with a reduced risk of colon cancer remains uncertain. However, the findings do indicate that total energy expenditure, whether from a job or leisure-time activities, is associated with colon cancer risk reduction.

Breast Cancer

Substantial evidence documents that physically active women have a 20% to 30% reduction in risk of breast cancer, compared with sedentary women. As with colon cancer, 30 to 60 min/day of moderate to vigorous intensity PA is needed to decrease breast cancer risk with risk declining further at the higher levels of PA. Although the biological mechanisms explaining how PA affects breast cancer risk remain unknown, lifetime moderate intensity PA appears to be a protective measure against breast cancer for all women.

Other Cancers

The available data reveal that PA is not associated with the risk of future rectal cancer. Data do suggest that physically active persons have a lower risk of lung cancer, but separating the effects of smoking is difficult. There is scant information on the role of PA in preventing other cancers.

Physical Activity and Diabetes

A substantial reduction in the occurrence of type 2 diabetes is consistently found among physically active persons compared with their sedentary peers. The magnitude of the reduced risk is 30% to 50% for active individuals with the benefit related to favorable effects of PA on body weight, insulin sensitivity, blood glucose control, blood pressure, and blood lipids. Most of these studies have observed significant benefits with daily walking for 30 min or more with additional benefits exhibited through participation in regular vigorous intensity PA. Some studies indicate that persons with an elevated risk of diabetes at baseline (such as higher body weight and fasting blood glucose) also demonstrate marked reductions in type 2 diabetes risk via regular PA or attainment of high aerobic fitness.

Physical activity combined with modest weight loss may exert optimal reduction in type 2 diabetes risk. In the U.S. Diabetes Prevention Program, combining 150 min/week of PA and a low-fat diet resulted in a 5% weight loss and 60% reduced risk of future type 2 diabetes. This benefit was superior to the nearly 30% risk reduction after treatment with an oral drug. This lifestyle intervention was found to be effective in men and women of all ethnic groups, including persons aged 60 years and older.

Regardless of the underlying biological mechanisms involved, regular PA is strongly related to a reduced risk of type 2 diabetes. The general recommendation is 30 min/day of moderate intensity endurance PA. Even further benefits may be attained with higher intensity aerobic activity, and performing strength-training exercises 2 to 3 days/week has also resulted in positive changes in biological markers of type 2 diabetes. However, further research is needed to uncover the ideal methods and intensities of PA, and more studies with women and minority groups need to be conducted.

Physical Activity and Obesity

Obesity is a problem of epidemic proportions with nearly two thirds of American adults suffering from overweight or obesity. The prevalence of overweight and obesity continues to increase among all age

groups and ethnicities, but PA plays a vital role in reversing this trend. The research shows an inverse relationship between levels of PA and body fatness, and men and women who are minimally active are 3 to 4 times more likely than their more active counterparts to experience weight gain.

Thirty minutes of moderate intensity PA, preferably all days of the week, is adequate to minimize health risks for chronic diseases, yet it may be insufficient for prevention of weight gain, stimulation of weight loss, or prevention of weight regain. The International Association for the Study of Obesity recommends 45 to 60 min/day of moderate intensity PA to prevent unhealthy weight gain and 60 to 90 min/day of moderate intensity PA or lesser amounts of vigorous PA to prevent weight regain in formerly overweight and obese individuals. There is no conclusive evidence regarding the amount of PA needed to incur significant weight loss; however, most studies indicate that exercise combined with modest caloric restriction is most effective in promoting weight loss.

Regular PA, independent of substantial weight loss, can provide improvements in health. Those who are overweight, yet engage in regular PA, have been termed the *fit fat*. Research shows that unfit lean adults have twice the risk of all-cause mortality as fit lean and fit obese adults. Although there is a direct relationship between body fatness and all-cause and CVD mortality, being active or fit decreases high mortality risk in obese persons.

—Steven P. Hooker and Anna E. Price

See also Cancer; Cardiovascular Disease; Obesity

Further Readings

Bassuk, S. S., & Manson, J. E. (2005). Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. *Journal of Applied Physiology*, *99*, 1193–1204.

Dishman, R. K., Washburn, R. A., & Heath, G. W. (2004). *Physical activity epidemiology.* Champaign, IL: Human Kinetics.

Lee, I. (2003). Physical activity and cancer prevention: Data from epidemiologic studies. *Medicine and Science in Sports and Exercise*, *35*, 1823–1827.

Pescatello, L. S., Franklin, B. A., Fagard, R., Farquhar, W. B., Kelley, G. A., & Ray, C. A. (2004). American College of Sports Medicine position stand. Exercise and hypertension. *Medicine and Science in Sports and Exercise*, *36*, 533–553.

Sallis, J. F., & Owen, N. (1999). *Physical activity and behavioral medicine.* Thousand Oaks, CA: Sage.

Saris, W. H. M., Blair, S. N., van Baak, M. A., Eaton, S. B., Davies, P. S. W., Di Pietro, L., et al. (2003). How much physical activity is enough to prevent unhealthy weight gain? Outcome of the IASO 1st stock conference and consensus statement. *Obesity Reviews*, *4*, 101–114.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. (1996). *Physical activity and health: A report of the surgeon general.* Atlanta, GA: Author.

Physicians' Health Study

The Physicians' Health Study (PHS) is a randomized, double-blind, placebo-controlled trial that was initially designed as a cohort study to test the effect of two medications: (1) the effect of aspirin on mortality due to cardiovascular disease and (2) the effect of beta-carotene on reducing the incidence of cancer. The initial planning for the PHS began in 1978, with Phase One (PHS-I) beginning in 1982 and ending in 1995. Phase Two of the cohort study (PHS-II) began in 1997 and is expected to conclude in 2007. This entry provides a general overview of the two phases of PHS and briefly discusses the major findings from PHS-I.

Physicians' Health Study Phase I

Study Population

The first phase of the Physicians' Health Study began in 1982, with funding from the National Cancer Institute and the National Heart, Lung, and Blood Institute. The study had two arms. One group of study participants were used to test whether aspirin prevented cardiovascular events such as a heart attack (myocardial infarction). A second group of participants were used to determine whether beta-carotene was useful in preventing cancer. Physicians aged 40 to 84 years were recruited between 1980 and 1982 as the study participants, and the study was conducted by mail-in survey between 1984 and 1995. A total of 22,071 physicians were eventually randomized into the trial. Principal investigators used physicians as the study population to obtain more accurate medical history and other pertinent health information.

Study Design

PHS-I was constructed to assign study participants to one of four possible treatment scenarios. Study participants received one of the following: two active medications (aspirin and beta-carotene), one active drug and one placebo (active aspirin and a beta-carotene placebo, or an aspirin placebo and active beta-carotene), or two placebos (neither pill was an active medication). Using blood samples and follow-up questionnaires, information was obtained regarding each participant's ability to adhere to the medication regimen, their use of other medications, significant health outcomes, and whether a participant had any illness or disease during the course of the study.

Major Findings

The aspirin arm of the study was stopped in 1988 after a finding that aspirin reduced the risk of myocardial infarction by 44%. This was a highly significant result when compared with the experience of those participants taking the aspirin placebo. Partly as a result of this finding, low-dose (325 mg) aspirin is now recommended as standard care for patients with cardiovascular disease.

The beta-carotene arm of PHS-I concluded in 1995. Although the medical literature had suggested that individuals consuming fruits and vegetables high in beta-carotene had lower rates of cancer, the results of PHS-I failed to demonstrate any positive or negative effect of beta-carotene supplementation on the incidence of cancer.

Physicians' Health Study Phase II

The second phase of the Physician's Health Study was initiated in 1997 and is expected to conclude in 2007. Funding is provided by the National Institutes of Health and additional private sponsors. This particular study is designed to determine whether certain dietary supplements have any effect on reducing the incidence of certain chronic diseases. Specifically, PHS-II investigates whether vitamin C, vitamin E, multivitamins, and beta-carotene serve to prevent colon cancer, prostate cancer, diseases of the eye, memory loss, and cardiovascular disease.

Participants in this study are physicians aged 50 years or older and who did not participate in the first Physician's Health Study. Similar to the design of PHS-I, physicians are assigned to one of 16 treatment scenarios. These include supplementation with one of the following: active forms of vitamins C and E, beta-carotene and a multivitamin, a set of four placebos, or a combination of both active supplements and placebos. A total of 14,642 physicians are randomized into the trial. The collection of information from each participant is also similar to the design of PHS-I: Blood samples and annual follow-up questionnaires are being used to obtain information about individual adherence to the supplement regimen, use of other medications, significant health outcomes, and incidence of major illness or disease during the course of the study.

Results from PHS-II are yet to be published, although information about the design and rationale for the trial is available in the medical literature. Findings regarding the long-term use of vitamin supplementation and the impact of supplements on the prevention of chronic disease are expected to follow the conclusion of the study, which is expected in December of 2007.

—Ashby Wolfe

See also Cancer; Cardiovascular Disease; Chronic Disease Epidemiology; Nutritional Epidemiology

Further Readings

Christen, W. G., Gaziano, J. M., & Hennekens, C. H. (2000). Design of Physicians' Health Study II: A randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. *Annals of Epidemiology*, *10*(2), 125–134.

Hennekens, C. H., Buring, J. E., Manson, J. E., Stampfer, M., Rosner, B., Cook, N. R., et al. (1996). Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *New England Journal of Medicine*, *334*(18), 1145–1149.

Steering Committee of the Physicians' Health Study Research Group. (1989). Final report on the aspirin component of the ongoing Physicians' Health Study. *New England Journal of Medicine*, *321*(3), 129–135.

Web Sites

Physicians' Health Study: http://phs.bwh.harvard.edu.

Pie Chart

A *pie chart* is a graphical representation of data as a disk divided into wedge-shaped "pieces" or "slices" whose sizes are proportional to the relative frequencies of the categories they represent. Because pie charts are most useful when they include only a small number of categories, they are most often used to display the relative frequencies of a categorical data set. Pie charts are a logical choice for graphical presentation when the focus of interest is on the relative frequencies of the data—that is, how much of the whole each category represents, rather than the absolute frequency of each category (in the latter case, a bar chart would be more appropriate).

Florence Nightingale, a founder of modern nursing, was also well versed in statistics. She invented a type of pie chart to show that in the Crimean War, far more soldiers died of illness and infection than those who died of battle wounds. Her campaign succeeded in improving hospital conditions and nursing so that many lives were saved.

Pie charts are most often created using statistical software but may also be created by hand: to obtain the angle for any category—that is, the size of the "slice," we multiply the relative frequency by 360°, because there are 360° in a complete circle.

For example, there were 863 kidney transplant patients who had their transplant performed at the Ohio State University Transplant Center during the period 1982 to 1992. There were 432 white males, 92 black males, 280 white females, and 59 black females.

Table 1 Race and Sex of Kidney Transplant Patients: Demographics of Kidney Transplant Recipients at the Ohio State University Transplant Center (1982–1992)

Race and Sex	*Frequency*	*Relative Frequency*
White male	432	0.501
Black male	92	0.107
White female	280	0.324
Black female	59	0.068
Total	863	1.000

Source: Adapted from data presented in Klein and Moeschberger (2003, p. 262).

Table 1 gives the race and sex information and calculates the relative frequencies for the four response categories.

To make a pie chart for this data set, we need to divide a disk into four wedge-shaped pieces that comprise 50.1%, 10.7%, 32.4%, and 6.8% of the disk. We do so by using a protractor and the fact that there are 360° in a circle. Thus, four pieces of the disk are obtained by marking off 180.36°, 38.52°, 116.64°, and 24.48°, which are 50.1% of 360°, 10.7% of 360°, 32.4% of 360°, and 6.8% of 360°, respectively. The pie chart for the relative frequency distribution for the above table is shown in Figure 1.

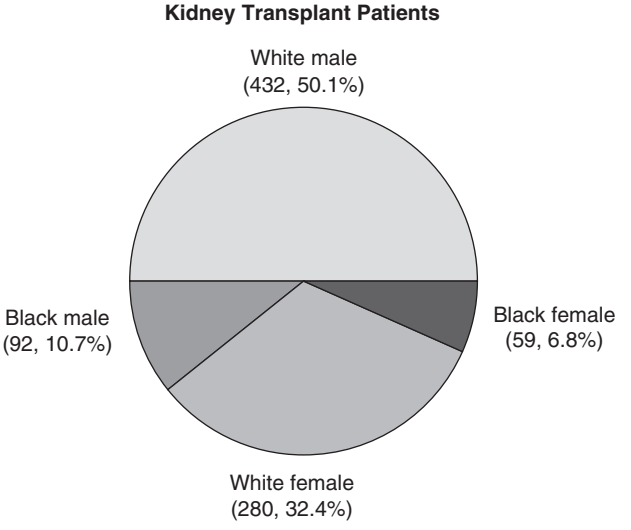


Figure 1 Race and Sex of Kidney Transplant Patients: Demographics of Kidney Transplant Recipients at the Ohio State University Transplant Center (1982–1992)

—Renjin Tu

See also Bar Chart; Graphical Presentation of Data; Nightingale, Florence; Proportion

Further Readings

De Veaux, R. D., Velleman, P. F., & Bock, D. E. (2005). *Stats data and models.* Boston: Addison-Wesley.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). New York: Springer.

Nightingale, F. (1958). *Polar-area diagram: Biographies of women mathematicians.* Retrieved November 10, 2006, from http://www.scottlan.edu/lriddle/women/nightpiechart.htm.

Placebo Effect

The placebo effect is an improvement in an individual's medical condition or an alleviation of adverse symptoms that occurs when the person receives an inert treatment. It may result from the person's expectation of improvement or from the increased motivation to make improvements in general health that may result. The placebo effect was first described in 1955 by Henry K. Beecher, an American physician, who described it in his frequently cited article, "The Powerful Placebo." The placebo effect has been explained as a result of the Pavlovian conditioning theory, the expectancy-value theory, and increased motivation on the part of the participant.

A placebo may be contact with a physician, cognitive or behavioral intervention, lifestyle changes in diet or level of physical activity, or a sugar pill. Regardless of the form, the aim of a placebo is to have no biologic effect at all. Placebos are typically used in placebo-controlled clinical trials where a treatment group receives the medical intervention being tested and the control group receives a placebo. The aim of this experimental design is to ensure that the study participants do not know whether they are receiving the treatment or the placebo. When the experimenter knows who is receiving the treatment and who is receiving the placebo, the study is single blind; if neither the participants nor the researcher knows, the study is double blind. Such studies minimize potential bias that may distort the true relationship between the exposure to the treatment and the outcome.

Although the aim of a placebo is primarily related to improving the methodology of a trial by blinding the participants to the status of the received treatments, one of the results of offering a placebo is that some individuals actually feel better and experience a beneficial effect despite the fact that the placebo has no known mechanism of action that may induce this effect. It seems that for illnesses such as depression, headache, stomach ailments, and pain, about a third of patients taking a placebo actually start to feel better because they believe they are receiving medical treatment, when in fact they are receiving an inert treatment.

The biologic mechanism by which a placebo can create this effect is unclear. However, it has been suggested that it is primarily a psychological effect that results from the individual's expectation that the treatment will work. Another theory of the mechanism of the placebo effect is that it is a conditioned response reflecting people's experience of treatment followed by symptom relief.

There are several ethical issues in the use of a placebo. Some bioethicists suggest that patients participating in trials cannot truly give informed consent if they do not know which treatment they will be receiving. Another criticism is that single-blind studies introduce an element of deception into health care and practice, because patients are told or allowed to believe that they are receiving a drug when the researchers know otherwise. Furthermore, in some cases, the placebo effect may actually result in adverse side effects, rather than only beneficial ones; this phenomenon is often called the nocebo effect. This may occur when individuals expect to experience negative side effects from the treatment.

—Kate Bassil

See also Ethics in Human Subjects Research; Hawthorne Effect; Randomization

Further Readings

Beecher, H. K. (1955). The powerful placebo. *Journal of the American Medical Association*, *159*, 1602–1606.

Turner, J. A., Deyo, R. A., Loeser, J. D., Von Korff, M., & Fordyce, W. E. (1994). The importance of placebo effects in pain treatment and research. *Journal of the American Medical Association*, *271*(20), 1609–1614.

Walsh, B. T., Seidman, S. N., Sysko, R., & Gould, M. (2002). Placebo response in studies of major depression: Variable, substantial, and growing. *Journal of the American Medical Association*, *287*, 1840–1847.

Plague

Yersinia pestis is the causative organism of plague, an enzootic vector-borne disease usually infecting rodents (e.g., rats) and fleas. Over the past 2,000 years, three devastating pandemics have occurred. Plague pandemics have caused social and economic

devastations on a scale unmatched by any other infectious disease except for smallpox. Although at the present time the organism is not considered a major health concern, approximately 2,500 cases annually are reported worldwide, and recently the World Health Organization categorized plague as a reemerging infectious disease. Despite major advances in diagnosis and treatment that were made since the discovery of the causative organism, the disease persists in several parts of the world, causing significant recurrent outbreaks in rodents and humans.

History

Reports of plague date back to ancient times, but the first undoubted account of bubonic plague is the *Great Plague of Justinian.* This first plague pandemic originated around AD 532 in Egypt and quickly spread to the Middle East and around the Mediterranean basin. In the following years, the disease spread as far north as into the territories of France and Germany. The estimated population losses in North Africa, Europe, and central/southern Asia were between 50% and 60% of the population. In contrast, the second pandemic—also known as the great medieval plague, Black Death, or Great Pestilence—is well described by many authors and many documents. It originated around the year 1334 in China and spread westward along the trade routes in Tauris on the Black Sea and eventually reached Constantinople (today's Istanbul) and the Crimea in 1347. From the Crimea, the disease was imported into Venice, Genoa, and Sicily by Italian merchant ships. The disease spread slowly but inevitably from village to village and eventually extended all over Europe, killing more than one third of its population. Despite the high mortality rate of the Black Death pandemic, the most devastating effects resulted from smaller, recurrent outbreaks that continued well into the 18th century. The third pandemic originated in China around 1855, rapidly spreading to its southern coast. The disease reached the city of Hong Kong in the 1890s. At this time, larger epidemics occurred all over China, marking the beginning of the next pandemic. Plague rapidly spread throughout the world to all inhabited continents, except for Australia.

Since then, smaller outbreaks have occurred around the world, with most recent outbreaks in Africa and Madagascar. In 1900, plague was introduced into North America (San Francisco), and between 1900 and 1924 most plague cases in the United States occurred in port cities along the Pacific and Gulf coasts. The disease spread slowly eastward with sporadic cases now being reported mainly in Arizona, New Mexico, Colorado, Utah, and Texas.

The causative organism of plague was discovered in 1894 during the early years of the third pandemic. Independent from each other, the Japanese microbiologist Shibasaburo Kitasato and the French microbiologist Alexandre Yersin conducted the experiments that led to the identification of the causative organism. Yersin's descriptions and explanations were published only a few days after Kitasato's; however, they seemed to be somewhat more accurate. Over the past decades, the literature has been quite inconsistent in crediting Yersin or Kitasato with the discovery of the plague bacillus. Finally, in 1970, the organism was officially named *Yersinia pestis.* In 1898, Paul-Louis Simond discovered that plague is transmitted by fleas. In 1927, Ricardo Jorge found an explanation for the occurrence of sporadic cases of plague.

Epidemiology and Clinical Manifestations

Plague occurs worldwide, with most cases reported in rural underdeveloped areas of Third World countries. In developed countries, advances in living conditions, public health, and antibiotic treatment made outbreaks of urban rat-borne plague less likely to occur in the decades after the third pandemic. However, the disease continues to be a problem in rural areas in the Americas, Africa, and Asia. This form of plague, termed the *sylvatic plague*, is maintained in wild rodents. Most recently, plague has resurged in sub-Saharan Africa and particularly in East Africa and Madagascar. In addition to its occurrence in nature, plague has been extensively researched for its role in biowarfare during the times of World War II and the Cold War. The possibility of plague being used as a biological weapon in the hands of military or terrorists remains as an important national security threat requiring special measures for medical and public health preparedness.

The organism is a gram negative, nonmotile bacillus and belongs to the family of Enterobacteriaceae. Based on historical data and bacteriological characteristics of strains isolated from remnant foci, Devignat described the three biovars Antiqua, Medievalis, and

Orientalis, which caused the first, second, and third pandemic, respectively. In nature, plague is primarily an infection of wild rodents and is transmitted by fleas. Worldwide, the domestic rats *Rattus rattus* and *Rattus norvegicus* are the most important reservoirs. The most common vector for transmission is the oriental rat flea.

Infection of *Y. pestis* in humans occurs in one of three clinical forms: *bubonic plague* is characterized by regional lymphadenopathy resulting from cutaneous or mucous membrane exposure to the organism; *primary septicemic plague* is an overwhelming plague bacteremia usually following cutaneous exposure; *primary pulmonary plague* follows the inhalation of aerosolized droplets containing *Y. pestis* organisms. Most cases of naturally occurring human plague represent the classic form of bubonic plague when victims are bitten by infected fleas. Plague can effectively be treated using antibiotics such as gentamicin, doxycycline, or tetracycline.

—*Stefan Riedel*

See also Bioterrorism; Epidemic; Insect-Borne Disease; Zoonotic Disease

Further Readings

Mandell, G. L., Bennett, J. E., & Dolin, R. (Eds.). (2005). *Principles of infectious diseases* (6th ed.). Philadelphia: Churchill Livingstone.

Murray, P. R., Baron, E. J., Jorgensen, J. H., Pfaller, M. A., & Yolken, R. H. (Eds.). (2003). *ASM manual of clinical microbiology* (8th ed.). Washington, DC: ASM Press.

Point Estimate

Most statistical analysis is done with the desire to reach a conclusion or decision about one or more parameters associated with a population of interest (statistical inference). Two types of estimators are used to assist in reaching a conclusion: point estimators and interval estimators. The goal of estimation is to provide a "best guess" at the true value of an unknown population parameter. To this end, a point estimator is a rule or function for computing a single quantity from a sample that will be used to approximate most closely a population parameter. Statistically, the point estimate is the value itself that is obtained when the rule is applied to sample data. Often the term point estimate is used to refer to any value computed from the data that is used to estimate a population parameter, even if it is not the "best" estimator.

The point estimate is the most common way that an estimate is expressed. Table 1 contains a list of the names or symbols for commonly used estimators along with the population parameters they estimate. Note that often point estimators are descriptive statistics.

Point estimates are quick and easy to calculate. They allow for a first look at the population based on sample data. Researchers hope that the single value obtained for a point estimator will be close to the parameter it is estimating. Since a point estimate is a random variable, it is in some way distributed about the true value of the population parameter. However, since the point estimate consists of a single number or a single point on the real number scale, there is room for questions. For example, a point estimate does not tell us how large the sample was on which it is based. Nor does it tell anything about the possible size of the error.

Since a researcher will infer that the population parameter is equal to the value of the point estimator, a little more knowledge of statistical inference is necessary. Statistical inference differs from ordinary inference in that not only is an inference made, but typically a measure is provided of how good the inference is. The error of estimation for a particular point estimate is defined to be the absolute value of the difference between the point estimate and the true population value. For example, the error of estimation for the mean is $|\bar{x} - \mu|$. However, the magnitude of the error of estimation is unknown since the true population parameter value is unknown. If a probability sample was taken, then statistical reliability may be calculated for the estimate by either using a confidence interval or by using the bound on the error of estimation.

Table 1 Common Parameters and Their Point Estimators

Parameter	*Population*	*Point Estimator*
Mean	μ	$\overline{X}$
Variance	σ^2	s^2
Proportion	p or π	$\hat{p}$
Relative risk	RR	$\widehat{RR}$

Sometimes a point estimate is referred to as the "realized value" since it is the actual numerical value of a random variable. Also, the phrase "point estimate" is sometimes used as an infinitive verb, as in the following: To point estimate is to compute a value from a sample and accept that value as an estimate of the unknown parameter.

Estimation of the Mean

Most often, researchers are interested in the value of the population mean for some variable. There are two points about estimating the mean. First, each of the measures of central tendency is a valid estimator for the population mean. However, the "best" or most robust estimator of the population mean is the sample mean. Thus, it is the estimator used in this article. Second, if the sampling distribution is approximately normal (i.e., the data have a bell-shaped curve or a mound-shaped histogram), then the Empirical Rule applies. By the Empirical Rule, the error of estimation will be less than $2\left(\sigma_{\overline{X}}\right) = 2(\sigma/\sqrt{n})$ approximately 95% of the time. The quantity $2(\sigma/\sqrt{n})$ is called the bound on the error of estimation. This quantity is a measure of how good our inference is. The smaller the bound on the error of estimation, the better the inference is. Since σ is unknown, we estimate it with the sample standard deviation (s) and obtain an approximate bound on the error. Point estimation with a bound on the error of estimation is used when no more than a crude statement of precision is required. If more precision is needed, confidence intervals are used.

—Stacie Ezelle Taylor

See also Confidence Interval; Histogram; Inferential and Descriptive Statistics; Measures of Central Tendency; Probability Sample; Random Variable; Robust Statistics; Sampling Distribution

Further Readings

Freund, J. E., & Smith, R. M. (1986). *Statistics: A first course* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Hurlburt, R. T. (2006). *Comprehending behavioral statistics* (4th ed.). Belmont, CA: Thomson/Wadsworth.

Ott, R. L., & Longnecker, M. (2001). *An introduction to statistical methods and data analysis* (5th ed.). Pacific Grove, CA: Duxbury.

Poisson Regression

See Regression

Polio

Polio is a viral disease that has caused considerable suffering for much of human history. The oldest clearly identifiable reference to paralytic poliomyelitis is an Egyptian stone engraving from 14th century BCE. Prior to the introduction of effective vaccines in the 1950s, polio was a common infection of childhood, with a small proportion of infections resulting in death or lifelong paralysis. Control began in 1955 after the first inactivated poliovirus vaccine (IPV) and subsequently several years later an oral polio vaccine (OPV) was also introduced. In most developed countries, a good level of control was achieved by the mid 1960s. In 1985, the Pan American Health Organization (PAHO) launched an initiative to eradicate polio in the Americas by 1990. Based on the success of the PAHO program, in May 1988, the 41st World Health Assembly committed the Member States of the World Health Organization (WHO) to the global eradication of poliomyelitis by the year 2000 (resolution WHA41.28). Despite great progress, by the end of 2006, there were still four countries in which polio was endemic (Nigeria, India, Pakistan, and Afghanistan), with another eight countries experiencing importations (Angola, Cameroon, Ethiopia, Indonesia, Nepal, Niger, Somalia, and Yemen).

Infectious Agent and Transmission

The poliovirus is an enterovirus with man as the only reservoir. There are three antigenic types: 1, 2, and 3. Type 1 most commonly causes paralysis, Type 3 less frequently, and Type 2 uncommonly. Most epidemics historically are due to Type 1. The risk of vaccine-associated poliomyelitis per million persons vaccinated ranged from .05 to .99 (Type 1), 0 to .65 (Type 2), and 1.18 to 8.91 (Type 3).

Infection is spread from person-to-person with fecal-oral transmission most common in developing countries where sanitation is poor, while oral-pharyngeal transmission is more common in industrialized countries and

during outbreaks. The mouth is the usual site of entry, and the virus first multiplies at the site of implantation in the lymph nodes in the pharynx and gastrointestinal tract. Incubation is usually from 7 to 10 days and may range from 4 to 40 days. The virus is usually present in the pharynx and in the stool before the onset of paralytic illness. One week after onset, there is low virus concentration in the throat, but the virus continues to be excreted in the stool for several weeks. Cases are most infectious during the first few days before and after onset of symptoms. For poliomyelitis, the ratio of inapparent (either subclinical or mild) infections to paralytic cases is very high, somewhere between 100 and 1,000 to 1. Long-term carriers are not known to occur.

Immunity

Susceptibility to poliomyelitis is universal. Epidemiologic evidence indicates that infants born to mothers with antibodies are naturally protected against paralytic disease for a few weeks. Immunity is obtained from infection with the wild virus or from immunization. Immunity following natural (including inapparent and mild) infections, or a completed series of immunizations with live OPV, results in both humoral (related to antibody production) and local intestinal cellular responses (a more localized response). Such immunity is thought to be lifelong and can serve as a block to infection with subsequent wild viruses and, therefore, helps in breaking chains of transmission. Vaccination with the IPV confers humoral immunity, but relatively less intestinal immunity; thus, vaccination with IPV does not provide resistance to carriage and spread of wild virus in the community. There is thought to be little, if any, cross-immunity between poliovirus types.

Clinical Features

Many infected with the wild poliovirus exhibit minor illnesses, but these cannot be distinguished clinically from illnesses caused by a number of other etiologies. Symptoms associated with minor illnesses include mild fever, muscle pains, headache, nausea, vomiting, stiffness of neck and back, and less frequently, signs of aseptic (nonbacterial) meningitis. Other conditions that may present similar to paralytic poliomyelitis include traumatic neuritis and tumors, followed less frequently by meningitis/encephalitis and illnesses produced by a variety of toxins. The most prominent difference between poliomyelitis and other causes of acute flaccid paralysis (AFP) is that for polio, the paralytic sequelae is generally severe and permanent, while for many other causes of AFP, paralysis tends to resolve or improve by 60 days after onset.

Susceptible older children and adults, if infected, are at greatest risk of paralytic illness. For persons with paralytic disease, the case-fatality rate varies between 2% and 20%; however, with either bulbar or respiratory involvement, case-fatality rates may reach as high as 40%. The majority of the deaths occur within the first week following onset of paralysis.

Epidemiology

Polio epidemics in both developed and developing countries were common during the first half of the 20th century. For example, in the United States more than 20,000 cases of paralytic disease were reported in 1952. Dramatic reductions in polio incidence was achieved in countries incorporating either IPV or OPV into their routine schedule. For example, in the United States cases dropped to < 100 in 1965 and < 10 in 1973. The last cases of indigenously transmitted wild-type poliovirus in the United States were in 1979.

The molecular epidemiology of wild poliovirus has recently proved to be useful in helping identify whether different virus isolates originate from a common ancestral source of infection. With this information, geographic foci or reservoirs of transmission can be defined and help trace sources of outbreaks throughout large geographical areas, such as have been seen in Africa during 2006.

Vaccines

There are currently two effective polio vaccines available: IPV, which first became available in 1955, and live attenuated OPV, first used in mass campaigns in 1959. In 1987, an enhanced inactivated poliovirus vaccine (eIPV) was introduced. In developing countries, OPV has been the vaccine of choice due to ease of administration, since it simulates natural infection and induces both circulating antibody and intestinal resistance, and by secondary spread protects susceptible contacts. In the Americas, using OPV mass campaigns interrupted transmission in areas where routine delivery had failed. Under ideal conditions in temperate countries, a primary series of three doses of OPV produces seroconversion to all three virus types in

more than 95% of vaccine recipients and is thought to have a clinical efficacy of nearly 100%. Three properly spaced doses of OPV should confer lifelong immunity. In developing tropical countries, the serologic response to OPV may be only 85%. This may be due to breaks in the cold chain, interference with intestinal infection by other enteroviruses, presence of diarrhea that causes excretion of the virus before it can attach to the mucosal cell, and other factors. Schedules may vary; WHO recommends that children receive four doses of OPV before 1 year of age. In endemic countries, a dose should be given at birth or as close to birth as possible. This is called the "birth dose," or "zero dose." The other three doses should be given at least 4 weeks apart and usually at the same time as DPT. For IPV, the current schedule is for four doses of the vaccine (2, 4, 6 to 18 months, and at 4 to 6 years) although the duration of immunity is not known with certainty.

OPV Versus IPV

In countries where polio is no longer endemic or there is little or no threat for reimportation, and cost is not a major consideration, increased use of IPV has been recommended since 1996. This has successfully reduced the risk of vaccine-associated paralytic polio. The overall risk in the United States for OPV vaccine-associated paralytic polio in vaccine recipients was one case per 5.2 million doses distributed. The risk of vaccine-associated paralytic polio in vaccine recipients for first dose was one case per 1.3 million doses. On the other hand, there are a number of advantages that favor OPV over IPV for use in an eradication programs. The rationale to use OPV includes the following: the development of intestinal immunity and ability to reduce intestinal spread of wild virus, duration of immunity, ease of administration in both routine and mass campaigns, and cost. Probably, the most critical issue relates to the effect of the vaccine on wild poliovirus transmission. It has been well documented that the use of OPV can successfully interrupt wild poliovirus transmission in both developed and developing countries. IPV protects against clinical disease and suppresses pharyngeal excretion of the virus but has little effect on intestinal excretion. Vaccinating children with IPV would reduce the number of paralytic cases due to the vaccine but, comparatively, would have little effect on the transmission of the wild poliovirus, which in developing countries is primarily by the fecal-oral route.

Vaccination Strategies

High immunization coverage is a key factor in the success of maintaining a polio-free environment or for eradication. Vaccination coverage of 90% or higher at 1 year of age with three or more doses of OPV or IPV must be maintained, not only at the national level but also at local levels. Immunization activities at the local level should be evaluated as to (1) the availability of routine immunizations, including the reduction of missed opportunities (e.g., false contraindications are one of the major causes of missed opportunities), (2) the extent of infant and preschool immunization programs, and (3) the availability of vaccination coverage data. If vaccination coverage is low, it is necessary to improve routine and outreach immunization activities and to determine whether mass immunization campaigns are needed to substantially raise low levels of coverage.

Mass Vaccination Campaigns

Conducting vaccination days (a selected time period in which a large number of people are vaccinated en masse) is an integral part of the polio eradication strategy, and without such campaigns polio is unlikely to be eradicated. Widespread vaccination produces extensive dissemination of the vaccine virus that competes with circulation of the wild virus and can abruptly interrupt virus transmission. Such activities are intended to supplement the routine immunization programs and can be held at the local or national levels. During the organization of these vaccination days, special attention needs to be paid to those locations in which coverage is below the national average. This is particularly true in areas with deficient health services. Not all endemic countries can successfully reach all high-risk populations with routine delivery and national vaccination days; therefore, it is necessary to mount special efforts to reach pockets of children in areas of potential wild poliovirus foci or in those areas not being served by existing health resources, a process known as mop-up.

Control and Eradication

Surveillance

Surveillance is the key to controlling and ultimately eliminating any disease threat. For polio, the

reporting system must cover key hospitals and clinics with at least one reporting source for each geopolitical unit. A concept of weekly reporting of all AFP cases rather than only poliomyelitis cases is critical to this effort. A concept of negative reporting of AFP (i.e., reporting even when no cases occur) must be integrated in the reporting system. Surveillance systems need to be continually monitored and feedback issued. Immediate response to reports in the surveillance system by trained epidemiologists must occur with every suspected case within 48 hr. Cooperation from the private medical community is essential for all surveillance efforts, and of course the public needs to be informed about reporting AFP.

Environmental Monitoring

In the latter stages of eradication when no cases are being reported, community monitoring may be considered. However, countries planning to start environmental surveillance should consult the WHO regional office at an early stage.

Methods of collection, concentration, and identification of viruses in the environment differ depending on the type of system that is being sampled, as well as the type of virus. Different methods for each of these steps have advantages and disadvantages, although the best chance of finding poliovirus is in stools of the cases and their contacts.

Global Eradication

In 1962, just 1 year after the introduction of Sabin's OPV in most industrialized countries, Cuba began using the oral vaccine in a series of nationwide polio campaigns. The success of these efforts demonstrated that polioviruses could be successfully eliminated from a developing country. In 1985, PAHO, under the leadership of Dr. Ciro de Quadros, launched an initiative to eradicate polio in the Americas by 1990.

On September 7, 1993, PAHO announced that 2 years had elapsed since the occurrence of the last case of poliomyelitis associated with wild poliovirus isolation in the Americas (Peru, August 1991). Although in 2000 and 2001, outbreaks occurred in the Dominican Republic and Haiti, respectively, these outbreaks were caused by a virus derived from the Sabin vaccine (a reversion of the vaccine virus to neurovirulence), and the same strategies that had eliminated polio previously were successful in bringing these outbreaks under control. The overall achievement in the Americas presented a new milestone in efforts to eradicate a disease.

An initiative to eradicate polio globally was launched in 1988. This has become one of the largest public health initiatives in history. In 1988, polio existed in more than 125 countries on five continents, and there were more than 35,251 reported cases of children paralyzed that year. By the end of 2006, there were 1,902 cases reported (provisional), with two countries, India and Nigeria, reporting more than 90% of the cases. The dominant type reported is Type 1 although some cases of Type 3 are still being reported.

Poliomyelitis transmission has been interrupted in the American, European, and Western Pacific Regions, and by end 2002, more than 180 countries and territories were polio free. With the eradication of polio and the eventual cessation of polio immunization, it is estimated that the world will save U.S.$ 1.5 billion (2006) dollars per year. The only other infectious disease eradicated previously was smallpox, the last case of which was reported during August 1977 in Somalia, and now it is hoped that polio will soon join this short list.

—Marc Strassburg

Acknowledgment: The author worked as a consultant for the PAHO polio eradication program, and in that capacity he assisted in writing a number of earlier versions of the polio eradication field guide. Sections from both previous and current field guides were adapted for this entry.

See also Disease Eradication; Pan American Health Organization; Public Health Surveillance; Vaccination

Further Readings

Andrus, J. A., de Quadros, C. A., Olivé J.-M., & Hull, H. F. (1992). Screening of cases of acute flaccid paralysis for poliomyelitis eradication: Ways to improve specificity. *Bulletin of the World Health Organization*, *70*(5), 591–596.

Cáceres, V. M., & Sutter, R. W. (2001). Sabin monovalent oral polio vaccines: Review of past experiences and their potential use after polio eradication. *Clinical Infectious Diseases*, *33*, 531–541.

Centers for Disease Control and Prevention. (2005). Poliomyelitis. In W. Atkinson, J. Hamborsky, L. McIntyre, & S. Wolfe (Eds.), *Epidemiology and prevention of vaccine-preventable diseases* (8th ed., pp. 89–100). Washington, DC: Public Health Foundation.

Cherry, J. D. (2004). Enteroviruses and parechoviruses. In R. D. Feigin, J. Cherry, G. J. Demmler, & S. Kaplan (Eds.), *Textbook of pediatric infectious diseases* (5th ed., pp. 1984–2041). Philadelphia: W. B. Saunders.

Halsey, N. A., & de Quadros, C. A. (Eds.). (1983). *Recent advances in immunization: A bibliographic review.* Scientific and Technical Publication No. 451. Washington, DC: Pan American Health Organization.

Heymann, D. (Ed.). (2004). *Control of communicable diseases in man* (18th ed.). Washington, DC: American Public Health Association.

Kew, O. M., Sutter, R. W., de Gourville, E. M., Dowdle, W. R., & Pallansch, M. A. (2005). Vaccine-derived polioviruses and the endgame strategy for global polio eradication. *Annual Review of Microbiology*, *59*, 587–635.

Pan American Health Organization. (2006). *Poliomyelitis eradication field guide* (3rd ed.). Washington, DC: Author.

de Quadros, C. A., Andrus, J. A., Olivé, J.-M., Da Silveira, C. M., Eikhof, R. M., Carrasco, P., et al. (1991). Eradication of poliomyelitis: Progress in the Americas. *Pediatric Infectious Disease Journal*, *10*, 222–229.

de Quadros, C. A., Andrus, J. A., Olivé, J.-M., de Macedo, C. G., Henderson, D. A. (1992). Polio eradication from the Western hemisphere. *Annual Review of Public Health*, *12*, 239–252.

World Health Organization, Department of Vaccines and Biologicals. (2003). *Guidelines for environmental surveillance of poliovirus circulation*. Geneva, Switzerland: Author. Retrieved July 24, 2007, from http://www.who.int/vaccines-document.

Pollution

Pollution can be defined as the presence of a substance or agent in the environment that is potentially harmful to health, safety, or comfort. In addition to affecting the health of humans or the ecosystem, pollution may have adverse effects on agricultural products or infrastructure such as buildings or monuments. Pollutants include naturally occurring and industrial chemicals, biological pathogens, and forms of energy such as noise.

The primary significance of pollution to epidemiology lies in its relation to human health. In some cases, this has been well studied, but for thousands of chemicals, it has not. The potential for a pollutant to cause adverse health outcomes is related not only to its toxicity but also on the extent of exposure. Briggs (2003) has estimated that 8% to 9% of the total global burden of disease is attributable to environmental or occupational pollution. Children and people in developing countries are disproportionately affected, and the most important routes of exposure are water and indoor air. Although this is only one estimate, it serves to underscore the impact of pollution on human health. This entry describes characteristics, sources and health effects of key pollutants, focusing on those affecting air and water.

Air Pollution

Human use of fire was perhaps the first anthropogenic source of air pollution, but it was the beginning of industrialization that really initiated a rapid escalation of the phenomenon. Several early-20th-century events brought with them recognition that circumstances of extreme air pollution could be threatening to health and even deadly. In 1930, in the Meuse River valley of Belgium, an atmospheric inversion during a period of cold, damp weather trapped pollutants of industrial origin close to the ground, resulting in 60 deaths, mostly among older persons with preexisting heart or lung disease. A similar event occurred in Donora, Pennsylvania, in 1948, and the infamous London smog of 1952 also involved similar meteorological conditions with pollutants created by burning of coal. In this case, thousands of deaths resulted. Following these incidents, efforts were made to reduce air pollution levels in the United States and Western Europe. However, air pollution remains a serious problem for humankind, as health effects of pollutants are discovered at even lower concentrations, formerly less developed nations undergo rapid industrialization, and greenhouse gas concentrations increase on an unprecedented scale.

Ambient Air Pollutants

Pollutants are released to the ambient air from an array of stationary or point sources (e.g., power plants, industrial sites), area sources (e.g., forest fires), and mobile sources (e.g., motor vehicles, boats, lawn mowers). Indoor sources are discussed separately below. Human exposure generally involves complex mixtures rather than individual pollutants.

The U.S. Environmental Protection Agency (EPA) is required by the Clean Air Act to set national ambient air quality standards for the protection of public health and welfare. The six major pollutants for which these standards are set, known as ‘‘criteria’’ pollutants, are carbon monoxide (CO), sulfur dioxide

(SO_2), nitrogen dioxide (NO_2), ozone, airborne particulates, and lead.

- *CO* is formed by incomplete combustion of carbon-based fuels such as gasoline or natural gas. The primary outdoor source is motor vehicles, while indoor sources include gas appliances and environmental tobacco smoke. CO has a high affinity for hemoglobin and interferes with oxygen transport. It causes acute poisoning at high levels; for typical environmental levels, associations with cardiovascular endpoints have been observed.

- *SO_2* is formed by the combustion of fossil fuels containing sulfur (primarily coal). Levels were very high in urban areas of the United States and Europe in the early to mid-20th century and have declined since the 1970s. However, high levels are now observed in other regions such as China, where coal use is currently high. SO_2 is associated with decreased lung function and respiratory symptoms, especially in asthmatics, and contributes to the formation of acid rain.

- Primary sources of *NO_2* include motor vehicle and power plant emissions and burning of fossil fuels. Local levels vary with the density of traffic. NO_2 is associated with lung irritation and lowered resistance to respiratory infection. In the presence of sunlight, NO_2 contributes to the formation of ozone.

- *Ozone* is a secondary pollutant formed by solar radiation and other pollutants such as nitrogen oxides and volatile organic chemicals (VOCs) from sources such as gasoline vapors, solvents, and consumer products. Due to the role of solar radiation in its formation, ozone levels are higher in summertime and in sunnier areas. Concentrations are higher downwind of urban centers than in cities themselves because of the time needed for these photochemical reactions to occur. Ozone in our ground-level air is considered a pollutant that contributes to formation of urban smog and is associated with reduced lung function and sensitization to other irritants; however, depletion of stratospheric ozone is associated with global warming and decreased protection from UV exposure.

- *Particulate matter* (PM) is a heterogeneous mixture of small particles and liquid droplets. Components include acids, organic chemicals, metals, soil, and dust particles from sources, including fuel combustion, high temperature industrial processes, atmospheric reactions of other pollutants, and mechanical processes such as demolition or road wear. PM is classified according to the diameter of the particles, which affects how far they can penetrate into the respiratory system: coarse PM (PM10) is <10 μm, fine (PM2.5) is <2.5 μm, and ultrafine particles are <0.1 μm. Respiratory and cardiovascular outcomes and visibility impairment have been associated with PM.

- *Lead* is a naturally occurring metal. For much of the last century, motor vehicles were the principal source of lead in air, and ambient concentrations have decreased markedly in the United States since the phase-out of lead from gasoline. Current sources for lead pollution in air include metal processing and waste incinerators. Lead exposure can have adverse effects on many of the body's organs and systems. It is considered of particular concern for infants and young children as one of the primary targets is the nervous system and exposure can lead to impaired neurodevelopment and reduced IQ.

Numerous air pollutants in addition to these six are known or suspected to pose health threats. The Clean Air Act amendments of 1990 designated 188 of these as "hazardous air pollutants." Chemicals in this category have been associated with cancer or other adverse health effects, including neurological, reproductive, developmental, immune, and respiratory outcomes. Examples include benzene, formaldehyde, perchloroethylene (used in dry cleaning), polycyclic organic matter (produced as combustion byproducts), and compounds of metals such as mercury and cadmium. The World Health Organization (WHO) and European Union (EU) also set air quality guidelines.

Indoor Air Pollutants

Because many people spend a large portion of time indoors, indoor air pollution may be as important as that of outdoor air in determining potential exposures. The quality of indoor air is influenced by that of the ambient outdoor air, but there are additional concerns arising from the built environment or indoor activities.

According to the WHO, more than half of the world's population uses solid biomass fuels (i.e., dung, wood, crop waste) or coal for cooking and heating. Burning these materials in open fires or simple, nonvented stoves produces indoor smoke, with specific components, including CO, particulate matter, and VOCs. Depending on the composition of coal, its

combustion can also produce SO_2 and toxins such as fluorine and arsenic. Exposure to smoke from these solid fuels is a risk factor for pneumonia and other lower respiratory infections, especially in young children, chronic obstructive pulmonary disease (COPD), and lung cancer. It may also be associated with other adverse outcomes, including cataracts, tuberculosis, asthma, and low birth weight.

Modern buildings raise air quality concerns related to building, furnishing, and consumer products contained within. They are often composed of synthetic materials and contained within airtight structures. Asbestos insulation and lead-based paints are instances in which products that were once widely used were subsequently recognized as posing serious threats to health. Other chemicals have been associated with sensory irritation, nervous system symptoms, and cancer. Common concerns include formaldehyde occurring in pressed-wood building materials, organic chemicals in paints or cleaning products, and pesticides. Microorganisms and allergens arising from sources, including humidifiers, air-cooling equipment, household pets, insects, and mold have been associated with allergies, asthma, and infections such as Legionnaires' disease. The phenomenon of *sick building syndrome*, a term used to describe various types of medically unexplained symptoms reported by people living or working in the same building, has drawn attention to possible health effects of indoor air exposures.

Environmental tobacco smoke (ETS), like smoking itself, is a concern worldwide. ETS consists of exhaled smoke plus secondary smoke produced by burning tobacco, and it is associated with lower respiratory infections, asthma, lung cancer, and adverse perinatal outcomes.

Radon is a naturally occurring, odorless, and colorless gas formed during the decay of uranium in the earth's crust. Higher levels tend to occur in homes on sandy or gravelly soil. Radon and its decay products result in radiation exposure in persons who breathe the affected air. The primary concern related to radon exposure is lung cancer; this association has been demonstrated in uranium miners. Because of a synergistic effect, risks are greater for smokers than for nonsmokers.

Water Pollution

The importance of clean drinking water for human life has been accepted for thousands of years. However, direct consumption accounts for only a minute fraction of human water use. Other uses include cooking, bathing, washing laundry, personal and industrial waste disposal, recreation, and irrigation. While some of these activities provide additional occasions for human exposure to pollutants in water, they may also be mechanisms of water contamination in and of themselves. Recognition of the idea that water can transmit disease is typically attributed to John Snow, who mapped cholera cases and traced them to a common water source during an 1854 cholera outbreak in London.

Sources of Water Pollution

There are numerous ways in which pollutants can enter water supplies. Point sources are those such as industrial or sewage treatment facilities, where discharges occur at an explicit location and are more easily identified and controlled than nonpoint or diffuse sources. These include agricultural and urban runoff. Deliberate discharges of pollutants may occur legally or illegally, and other sources include leakage or spills, seepage from landfills, and atmospheric deposition. Naturally occurring contaminants, such as arsenic, can leach into groundwater from geological formations. Water distribution systems are another source where water may pick up contaminants—for example, iron, lead, or copper due to leaching or corrosion from pipes. Finally, disinfectants added to water to treat microbial contamination may react with organic matter to form halogenated chemicals known as disinfection by-products.

Health Effects of Water Pollutants

In developing countries, many cities lack infrastructure for waste treatment and discharge the majority of sewage directly into water sources such as rivers and streams. Hence, in many areas of the world, human feces is the most important contaminant, and waterborne diseases (such as diarrheal illness, cholera, typhoid, and amebic dysentery) are the greatest human health risks associated with water contamination. Waterborne disease outbreaks still occur in developed countries as well. For example, a 1993 cryptosporidium outbreak in Milwaukee, Wisconsin, caused more than 400,000 cases of illness.

Contamination from naturally occurring substances such as arsenic and fluoride can be a concern in both developed and developing countries, although they

are associated with greater morbidity for the latter, where water testing and treatment facilities are often not in place. Arsenic exposure is associated with skin, lung, and bladder cancer; keratosis; and peripheral vascular damage and is particularly prevalent in Bangladesh. Areas with high levels of fluoride in drinking water include India, Africa, China; this can cause fluorosis, which involves dental discoloration, decay, and skeletal deformity.

Chemical contaminants of water supplies include agricultural products such as nitrates, pesticides, and fertilizers; chemicals from urban or industrial sources, especially heavy metals and solvents; pharmaceuticals or their breakdown products; and disinfection by-products such as trihalomethanes. Some of these pollutants are volatile so that exposure may occur via absorption through the skin or inhalation, for example, during showering. Epidemiological studies on possible health effects of these contaminants have tended to focus on cancer and reproductive/developmental effects such as miscarriage or low birth weight; however, the overall evidence for association is generally inconclusive.

Other Types of Pollution

Pollution affecting air and water are highlighted here as two categories of major import for the health of populations worldwide. However, there are other media affected by pollution and other possible ways of classifying pollutants. Some of these are mentioned below.

Food in its various manifestations is at once life sustaining and a potential source of exposure to a variety of pollutants. Plant products are vulnerable to sources, including pesticides that are applied intentionally, deposition from traffic or industry, sludge application, and waste disposal or spills. Pollutants may directly deposit on plants, as in a sprayed pesticide, or be taken up from soil in which they are contained. For example, cadmium concentrates in leafy vegetables. Contaminated soil itself may be a source of exposure by ingestion, especially for young children playing near the ground with abundant hand-to-mouth behavior. Animals can also ingest contaminated soil, water, plants, or feed, and then become a source of exposure for humans who consume their milk, eggs, or meat. Bioaccumulation is a process whereby concentration of a chemical within organs or tissues of an organism exposed to it increases over the concentration of the chemical in the surrounding environment. This can lead to increasing concentrations up the food chain, the potentially detrimental consequences of which were illustrated in Minamata Bay, Japan, in the 1950s. Wastes containing mercury were discharged by a chemical company and concentrated in fish and shellfish living in the contaminated waters. People who then ate large quantities of fish caught from the bay were affected by methyl mercury poisoning, resulting in neurological impairment and in some cases, death. A congenital form affecting infants exposed in utero showed cerebral-palsy-like symptoms.

Persistent organic pollutants (POPs) are chemicals that persist in the environment, accumulate in fat tissue of animals, circulate globally achieving a wide geographic distribution, and have health effects on humans and animals, most notably disruption of endocrine systems and reproduction. POPs include DDT and other pesticides, polychlorinated biphenyls (PCBs), dioxins, and furans. Of relatively recent concern are the polybrominated diphenyl ethers (PBDE), which are used as flame retardants. POPs may be present in the food supply, air, and water; infants can also be exposed through breast milk.

Greenhouse gases are gases, especially carbon dioxide, methane, and nitrous oxide that trap and retain heat from the sun. The primary anthropogenic source is carbon dioxide from the burning of fossil fuels. Evidence indicates that rising atmospheric levels of these gases are contributing to an observed increase in global temperatures over the last century. Global warming and associated climate change may eventually have a variety of effects with direct impact on human health, including illness or injury from extreme weather events, changes in geographic distribution of disease vector organisms, impaired crop or livestock production, and displacement of populations.

Physical agents may also be considered as pollutants. For example, noise pollution can be described simply as unwanted noise, coming from sources such as airports and traffic. Radiation and heat are other examples that fall into this category.

—Keely Cheslack-Postava

See also Environmental and Occupational Epidemiology; Harvard Six Cities Study; Lead; Love Canal; Mercury; Sick Building Syndrome; Urban Health Issues; Waterborne Diseases

Further Readings

Briggs, D. (2003). Environmental pollution and the global burden of disease. *British Medical Bulletin*, *68*, 1–24.

Gordon, B., Mackay, R., & Rehfuess, E. (2004). *Inheriting the world: The atlas of children's health and the environment.* Geneva, Switzerland: World Health Organization. Also available online at http://www.who.int/ceh/publications/atlas/en.

Moeller, D. W. (2005). *Environmental health* (3rd. ed.). Cambridge, MA: Harvard University Press.

Patz, J. A., & Balbus, J. M. (2001). Global climate change and air pollution. In J. L. Aron & J. A. Patz (Eds.), *Ecosystem change and public health: A global perspective* (pp. 379–408). Baltimore: Johns Hopkins University Press.

Tibbetts, J. (2000). Water world 2000. *Environmental Health Perspectives*, *108*, A69–A73.

World Health Organization. (2005). *Indoor air pollution and health* (Fact sheet no. 292). Retrieved December 20, 2006, from http://www.who.int/mediacentre/factsheets/fs292/en/print.html.

Web Sites

U.S. Environmental Protection Agency. Office of Air and Radiation: http://www.epa.gov/air.

Population Pyramid

The population pyramid is a graphical representation of the age and gender composition of a specific population. The shape of the graph depends on the age and gender structure of the population. The representation may take the form of a pyramid, but it may have a columnar shape, with vertical sides rather than sloped sides, or it may have an irregular profile.

Population pyramids provide a summary view of the overall age-gender structure of a specific population. The size of the population is depicted on the horizontal axis, and age is aligned on the vertical axis. The depiction actually contains two graphs, in mirror image format, on either side of a central vertical axis; the female population is represented on the right side of the axis, and the male population is shown on the left side.

The population pyramid is made up of bars stacked on top of one another, each representing an age category, typically in 5-year age groups, with the youngest age group represented by the bottom bar, and the oldest age group by the uppermost bar. The length of each bar, on either side of the central vertical axis,

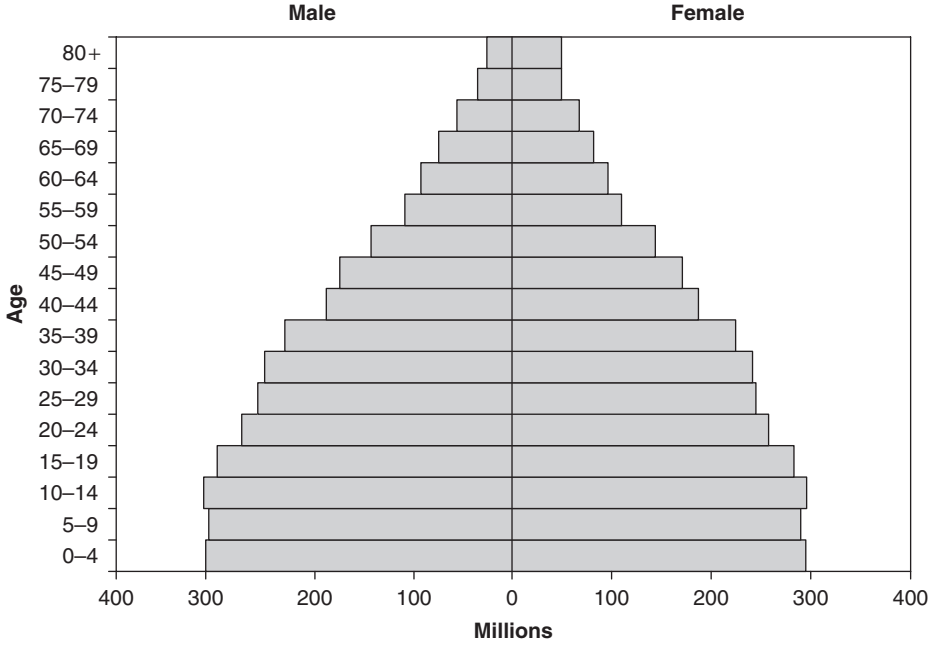


Figure 1 Age-Sex Structure of Global Population: 2002

Source: U.S. Census Bureau (2004a).

represents the number of males (left side) and females (right side) in the specific age group, in the population depicted. The age groups are displayed along the central axis or along one side, and often the years of birth for each age category are also displayed on the graph. To maintain proportionality, the age groups are all of the same size (typically in 1-year, 5-year, or 10-year age groups), and the bars are all of equal height. However, the age axis is often truncated at the age group 80 to 84, depending on the data available for the population depicted. For some populations, the data for the older age groups are incomplete or inaccurate, or there are few people in the older age categories.

The population pyramid can depict the proportion of the total population in each age-gender group rather than the actual count. In this case, for example, the length of the bar for females in the 5- to 9-year-old group would represent the proportion that group consists of within the total population. When calculating the proportions, the denominator used is always the number in the entire population, and the numerator is the number in the specific age-gender group. The sum of all the groups represented by the bars should add up to 100% of the population depicted.

When comparing population pyramids, it is important to note whether proportions or counts are represented and whether the scale of the bars and the age categories are the same. Population pyramids intended for comparison should be drawn to the same scale, and should depict the same age categories. The population pyramid can be used to represent additional characteristics of the population, such as marital status, race, or geographic location. In this case, the bar for each age-gender group is further subdivided and formatted to represent the additional categories. The formatting system used to depict the additional categories should be applied consistently throughout the graph. The same sequence should be used on either side of the vertical axis, in mirror image form. For example, if race is depicted, and the categories are white, black, and other, the categories would be arranged in the same sequence for males and for females, working outward from each side of the central axis.

The shape of the population pyramid efficiently communicates considerable information about the age-gender structure of a specific population. A broad-based pyramid indicates people in the younger age categories make up a relatively large proportion of the population, and a narrow or pointed top indicates older people make up a relatively small proportion of the population. In the older age groups in many populations, the number of females is much greater than the number of males, and this is reflected in the shape of the pyramid; the bars on the right side of the central axis (the female side) are longer than those on the left (male) side. The median age of the population would be the age group (bar) represented by the point on the vertical axis that equally divides the area within the pyramid, that is, about equal areas within the pyramid fall above and below the age represented by this bar. The fertility and mortality of the population are also reflected in the shape of the population pyramid. A broad base and sharply tapering sides (a true pyramid shape) reflects a high fertility rate and high mortality rates in younger age groups. Irregularities in the profile of the population pyramid convey information about changes in the population or aberrations. A bulge or an indentation in the profile of the population pyramid may indicate unusually high fertility or mortality, or changes in the population due to in-migration or out-migration. For example, the bulge in the age groups of 35 to 54 years in the pyramid representing the 2000 U.S. population reflects a period of high fertility, the post–World War II baby boom (Figure 2).

Demographers who have studied the historical changes in the age and gender composition, fertility, and mortality of the world's populations have articulated

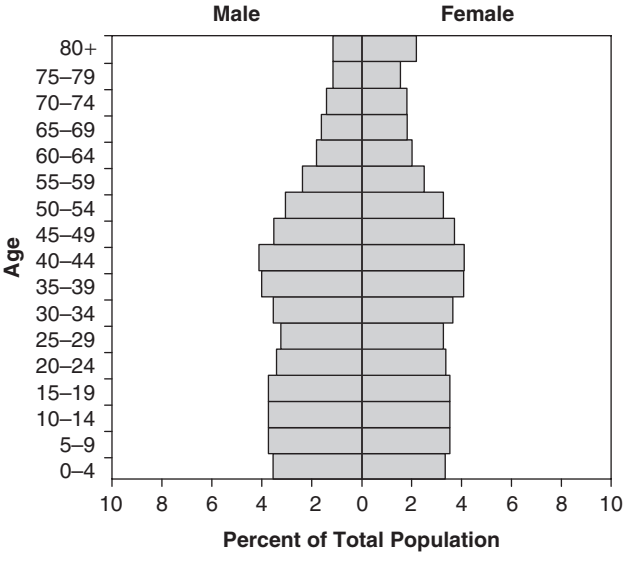


Figure 2 Bulge in a Population Pyramid Due to a Baby Boom (United States, 2000)

Source: U.S. Census Bureau (2004a).

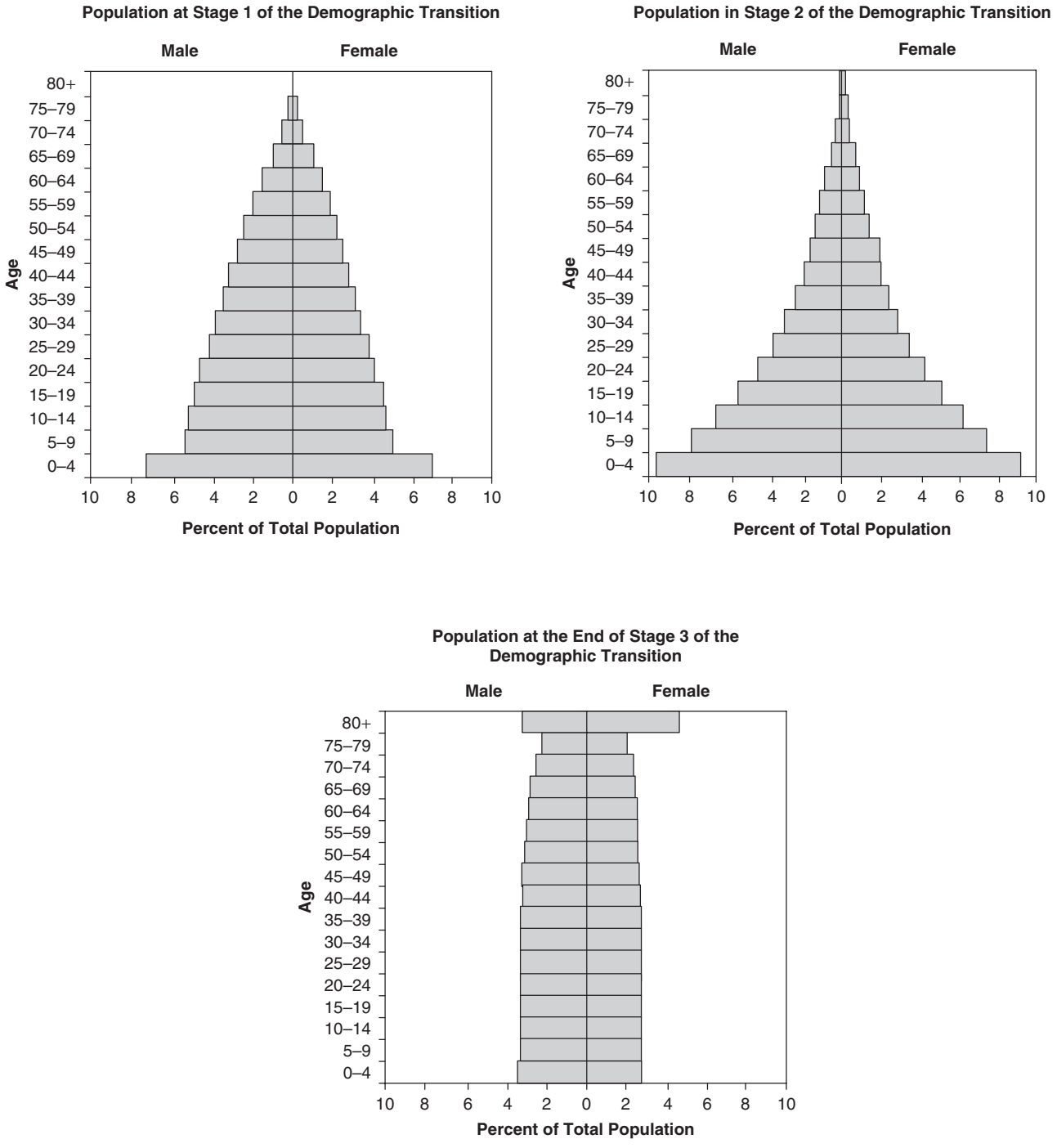


Figure 3 Stages in the Demographic Transition

Source: U.S. Census Bureau (2004a).

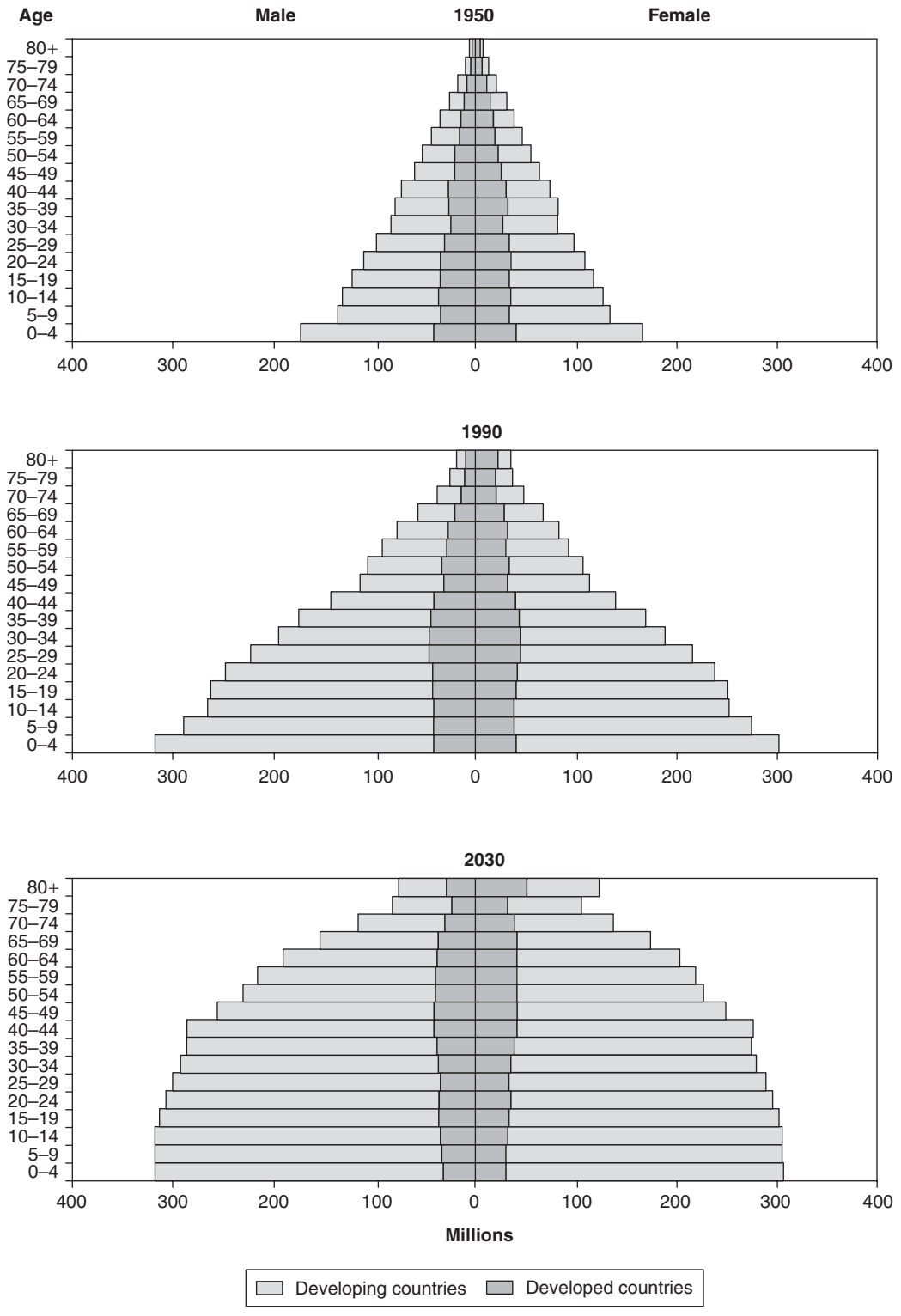


Figure 4 Population by Age and Sex: 1950, 1990, and 2030

Source: Kinsella and Velkoff (2001).

Figure 5 Age-Sex Structure of World Population: 2002 and 2050

Source: U.S. Census Bureau (2004b)

a theory of "demographic transition." This theory seems to provide a useful approximation of the historical changes that have taken place in the populations in many different regions of the world. The stages of this transition are represented by dramatically different population pyramids (Figure 3).

Stage 1 is represented by a tapering pyramid sitting on a broad base, reflecting high fertility, but also a high mortality rate among the younger age groups, so the population increases slowly, and remains relatively small. The shape of the population pyramid for Stage 2 of the demographic transition reflects lower mortality, especially among the youngest age groups, coupled with high fertility; the population increases rapidly but remains relatively young. The population pyramid that represents Stage 3 in the demographic transition is roughly rectangular, reflecting lower fertility, lower childhood mortality, and longer survival; the older age categories make up a larger proportion of the population than in earlier stages, and the size of the population stabilizes.

The world's population does appear to be progressing through the stages of the demographic transition. Figure 4 shows changes in the age structure of the global population over time and shows the contrast between developed countries and developing countries. It is apparent from this figure that the developed countries are further ahead in the demographic transition than are the developing countries, but the structure of the world's population as a whole is approaching the third stage of the demographic transition.

The overall aging of the world's population can be seen in Figure 5, which shows the dramatic changes that are expected to occur by 2050.

The variety in the age and gender structure of the population in different regions of the world is graphically depicted in Figure 6, which superimposes regional population pyramids on the global population pyramid.

—Judith Marie Bezy

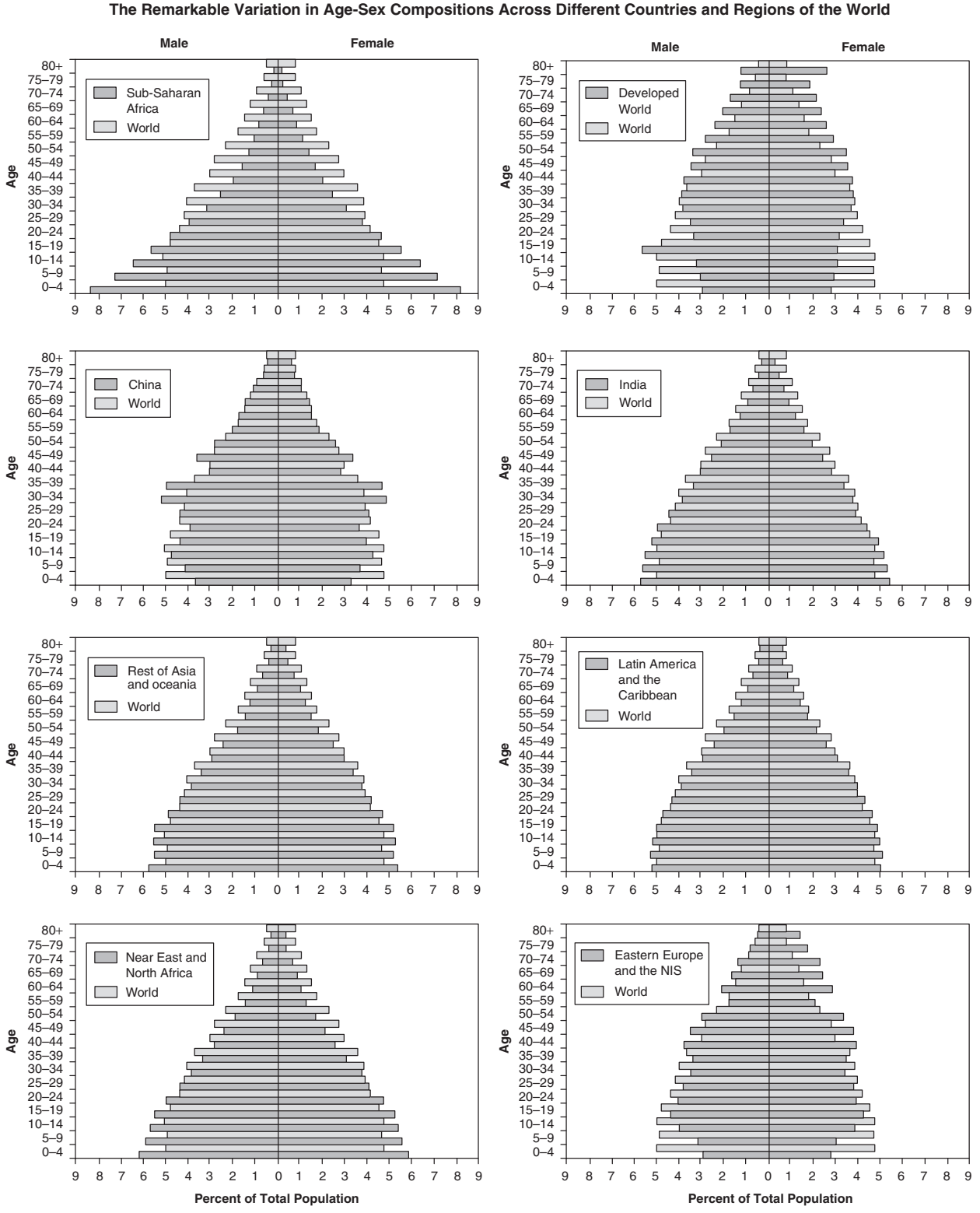


Figure 6 Population Pyramids by Region and Selected Countries: 2002

Source: U.S. Census Bureau (2004a).

See also Demography; Graphical Presentation of Data; Proportion

Further Readings

Kinsella, K., & Velkoff, V. A. (2001). *An aging world: 2001* (U.S. Census Bureau, Series P95/01-1). Washington, DC: Government Printing Office.

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The methods and materials of demography* (2nd ed.). San Diego, CA: Elsevier Academic Press.

U.S. Census Bureau. (2004a). *Global population profile: 2002* (International Population Reports, WP/02). Washington, DC: Government Printing Office.

U.S. Census Bureau. (2004b). *Global population at a glance: 2002 and beyond* (International Brief, WP/02-1).Washington, DC: Government Printing Office.

Positive Predictive Value

See Clinical Epidemiology

Post-Traumatic Stress Disorder

Post-traumatic stress disorder (PTSD) was adopted by the American Psychiatric Association (APA) as part of the official classification of psychiatric disorders in the third edition of the *Diagnostic and Statistical Manual of Mental Disorder* (*DSM-III*), published in 1980. The adoption of PTSD in *DSM-III* was motivated by pressures from advocates on behalf of Vietnam War veterans. The definition of PTSD in the *DSM-III* and subsequent *DSM* editions, *DSM-III-R* and *DSM-IV*, is based on the concept that traumatic events, in contrast with other stressful events, are linked etiologically to a specific syndrome. The PTSD syndrome is defined by three symptom groups: (1) reexperiencing the traumatic event, (2) avoidance of stimuli that resemble the event and numbing of emotional responsiveness, and (3) increased arousal. These features are defined in terms of their connection with the traumatic event that caused them. Temporal ordering is also required: These disturbances must not have been present before the trauma occurred.

Since 1980, research on PTSD has focused chiefly on Vietnam War veterans and to a lesser degree on victims of specific types of traumas, such as natural disasters or rape. With the growth of the field of psychiatric epidemiology, PTSD has been studied in samples of the general population in the United States and other countries. From the time of its introduction into the official psychiatric nosology, PTSD has been a controversial diagnosis. Some critics question its validity as a distinct disorder. Others contest the trend toward a broader, more inclusive definition of what constitutes a traumatic event. Concerns have been expressed about the proliferation of PTSD-related disturbances other than *DSM-IV* PTSD (e.g., subthreshold PTSD) and about potential distortion in recalling traumatic experiences, especially when the diagnosis of PTSD entitles victims to compensation. Despite extensive efforts, neurobiological research has not yielded laboratory tests that can be used diagnostically.

Exposure to *DSM-IV* Traumatic Events and PTSD

In the latest edition of the *DSM*, the *DSM-IV* (APA, 1994), the definition of traumatic events that can potentially cause PTSD has been enlarged to include a wider range of stressors than the typical stressors of the initial definition (combat, concentration camp confinement, natural disaster, rape, or assault). The stressor definition in *DSM-IV* requires that "the person experienced, witnessed or was confronted with an event(s) that involved actual or threatened death or serious injury or a threat to the physical integrity of self and others," and which evoked "intense fear, helplessness, or horror." Thus, learning that someone else was threatened with harm qualifies as a traumatic event. The defining features of the PTSD syndrome have remained unchanged, although the specific configuration of symptoms was revised somewhat. *DSM-IV* introduced a new condition—that the disturbance causes clinically significant distress or impairment—in recognition that distress in itself or in commonly experienced symptoms, such as sleep problems, is not equivalent to a mental disorder. Survey data from the United States, where results based on earlier definitions are available for comparison, show that the broader definition of stressors has resulted in a considerably higher proportion of the population having experienced traumatic events that qualify for PTSD. However, prevalence estimates of *DSM-IV* PTSD have not increased. In the United States, the vast

majority of the population (approximately 80%) has experienced one or more traumatic events. A similarly high figure has been reported in a Canadian study. Much lower figures have been reported in surveys in Germany and Switzerland (from 20% to 28%).

Although most of the U.S. population has been exposed to one or more traumatic events, only a minority of victims has succumbed to PTSD ($< 10\%$). The lifetime cumulative incidence of *DSM-IV* PTSD in a national sample of the U.S. population, circa 2000, was 6.8%. Table 1 presents published estimates of lifetime cumulative incidence and 12-month prevalence of PTSD from surveys in the United States, Canada, Germany, the Netherlands, Switzerland, Lebanon, and Australia. Both lifetime and 12-month estimates are higher in the United States than in other countries.

A consistent finding across epidemiologic studies is the higher PTSD prevalence in women compared with men. Although men are more likely to experience trauma, the likelihood of developing PTSD following exposure to traumatic events is higher in women. Rape and sexual assault occur more frequently in women, and victims of either sex are more likely to succumb to PTSD than victims of other traumatic events. However, the sex difference in PTSD is not accounted for by women's greater risk of experiencing rape and sexual assault.

Table 1 Cumulative Incidence in Lifetime and 12-Months Prevalence of *DSM-IV* Post-Traumatic Stress Disorder (PTSD)

			Total	
Study	*Sample*	*Interview*	*Lifetime*	*12-Months*
Breslau et al. (1998)	Detroit PMSA, United States Age 18–45 ($n = 2{,}181$)	WHO-CIDI (telephone interview)	12.2% (8.3%)	—
Breslau, Wilcox, Storr, Lucia, and Anthony (2004)	Mid-Atlantic City, United States Age 19–22 ($n = 1{,}698$)	WHO-CIDI (personal interview)	7.1%	—
Kessler et al. (2005a, 2005b)	U.S. national Age $\geq$ 18 ($n = 9{,}282$)	WMH-CIDI (personal interview)	6.8%	3.5%
Stein, Walker, Hazen, and Forde (1997)	Winnipeg, Canada Age >18 ($n = 1{,}002$)	PTSD Symptom Scale (telephone interview)	—	2.0%
Hapke et al. (2005)	Luebeck, Germany Age 18–64 ($n = 4{,}075$)	M-CIDI	1.4%	—
Perkonigg, Kessler, Storz, and Wittchen (2000)	Munich, Germany Age 14–24 ($n = 3021$)	WHO-CIDI (personal interview)	1.3%	0.7%
Van Zelst, de Beurs, Beekman, Deeq, and van Dyck (2003)	The Netherlands Age 55–85 ($n = 422$)	WHO-CIDI (personal interview)	—	0.9%
Hepp et al. (2006)	Zurich, Switzerland Age 34–35 and 40–41	Zurich Cohort Study (personal interview)	—	0.0%
Karam et al. (2006)	Lebanon national Age $\geq$ 18 ($n = 308$)	WHO-CIDI (personal interviews)	—	2.0%
Creamer, Burgess, and McFarlane (2001)	Australian national Age 18 ($n = 10{,}641$)	WHO-CIDI modified (personal interview)	—	1.3%

Sources: Breslau et al. (1998), Breslau, Wilcox, Storr, Lucia, and Anthony (2004), Creamer, Burgess, and McFarlane (2001), Hapke et al. (2005), Hepp et al. (2006), Karam et al. (2006), Kessler et al. (2005a, 2005b), Perkonigg, Kessler, Storz, and Wittchen (2000), Stein, Walker, Hazen, and Forde (1997), and Van Zelst, de Beurs, Beekman, Deeq, and van Dyck (2003).

Note: Dash represents not published. CIDI, Composite International Diagnostic Interview; WMH, World Mental Health; *DSM-IV*, *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (1994).

Suspected Risk Factors, Course, and Co-Occurring Disorders

Because traumatic events cause PTSD in only a small fraction of victims, researchers have sought to identify risk factors that predict who succumbs to the disorder among those who have experienced trauma. Suspected risk factors include personality traits (neuroticisms), preexisting psychiatric disorders, family history of psychiatric disorders, and prior exposure to traumatic events. High intelligence (more than 1 *SD* above the population mean) has been found to protect individuals against PTSD effects after exposure to traumatic events. PTSD is more likely to occur following events involving assaultive violence than other types of traumatic events, such as natural disasters or severe accidents. Recently, there has been a growing interest in the relationship between PTSD and terrorist attacks on civilian populations. Surveys following the September 11, 2001, terrorist attacks on the World Trade Center indicate that, with the exception of persons directly involved, any increase in the prevalence of PTSD among New York City residents was transient.

The onset of PTSD symptoms among victims with the disorder occurs within days of the traumatic experience. Cross-sectional surveys of the general population shows recovery over time, although the disorder is generally chronic, lasting longer than 6 months in the majority of cases. Victims who develop PTSD are at an increased risk for the first occurrence of other psychiatric disorders, chiefly major depression, anxiety disorders, drug use disorders, and nicotine dependence. Trauma victims who do not succumb to PTSD (i.e., most victims) are not at a markedly increased risk for subsequent onset of other psychiatric disorders, compared with community residents who have not experienced traumatic events.

Treatment

Both psychological and pharmacological approaches have been developed for patients with PTSD. Of the psychological treatments, specialized versions of cognitive behavior therapies (CBT) that help patients confront fear and avoidance in a structured format have been found in randomized clinical trials to be efficacious in improving PTSD symptoms. With respect to medication, to date, two serotonin reuptake inhibitors, sertraline and paroxetine, have been approved by the Food and Drug Administration (FDA) for the treatment of PTSD. Debriefing, a popular technique involving immediate post-trauma intervention that encourages victims to recount their experiences, has been found to be ineffective and at times damaging.

—Naomi Breslau

See also Psychiatric Epidemiology; Stress; Violence as a Public Health Issue; War

Further Readings

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: The 1996 Detroit area survey of trauma. *Archives of General Psychiatry*, *55*, 626–632.

Breslau, N., Wilcox, H. C., Storr, C. L., Lucia, V. C., & Anthony, J. C. (2004). Trauma exposure and posttraumatic stress disorder: A study of youths in urban America. *Journal of Urban Health*, *81*, 530–544.

Brewin, C. R., Andrews, B., & Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology*, *68*, 748–766.

Creamer, M., Burgess, P., & McFarlane, A. C. (2001). Post-traumatic stress disorder: Findings from the Australian National Survey of Mental Health and Well-being. *Psychological Medicine*, *31*, 1237–1247.

Foa, E. B., Keane, T. M., Friedman, M. J., & International Society for Traumatic Stress Studies. (2000). *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies.* New York: Guilford Press.

Hapke, U., Schumann, A., Rumpf, H. J., John, U., Konerding, U., & Meyer, C. (2005). Association of smoking and nicotine dependence with trauma and posttraumatic stress disorder in a general population sample. *Journal of Nervous and Mental Disease*, *193*, 843–846.

Hepp, U., Gamma, A., Milos, G., Eich, D., Ajdacic-Gross, V., Rossler, W., et al. (2006). Prevalence of exposure to potentially traumatic events and PTSD: The Zurich Cohort Study. *European Archives of Psychiatry and Clinical Neuroscience*, *256*, 151–158.

Karam, E. G., Mneimneh, Z. N., Karam, A. N., Fayyad, J. A., Nasser, S. C., Chatterji, S., et al. (2006). Prevalence and treatment of mental disorders in Lebanon: A national epidemiological survey. *Lancet*, *367*, 1000–1006.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593–602.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and

comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 617–627.
Perkonigg, A., Kessler, R. C., Storz, S., & Wittchen, H. U. (2000). Traumatic events and post-traumatic stress disorder in the community: Prevalence, risk factors and comorbidity. *Acta Psychiatrica Scandinavica*, *101*, 46–59.
Rose, S., Bisson, J., Churchill, R., & Wessely, S. (2002). Psychological debriefing for preventing post traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews*, Article No. CD000560.
Stein, M. B., Walker, J. R., Hazen, A. L., & Forde, D. R. (1997). Full and partial posttraumatic stress disorder: Findings from a community survey. *American Journal of Psychiatry*, *154*, 1114–1119.
Van Zelst W. H., de Beurs, E., Beekman, A. T., Deeq, D. J., & van Dyck, R. (2003). Prevalence and risk factors of posttraumatic stress disorder in older adults. *Psychotherapy and Psychosomatics*, *72*(6), 333–342.

Poverty and Health

The adverse effects of poverty on health are well documented and continue to be a major public health concern all over the world. Much of the early pioneering work in public health has its roots in the study of the health consequences of poverty. And, although poverty has long been known to cause numerous public health problems, billions of people globally continue to be affected by poverty. This entry examines the extent of poverty, its impact on health, methods for measuring poverty, and strategies for alleviating poverty and addressing the health needs of the poor.

Demographics

The World Bank estimated that in 2001, 2.7 billion people worldwide lived on less than $2 per day and more than a billion subsisted on less than $1 per day. Some regions, such as eastern and southern Asia, have seen reductions in extreme poverty (from a rate of 33% in 1990 to 14% in 2002 in eastern Asia and from 39% in 1990 to 31% in 2002 in southern Asia). Yet the percentage of persons living in extreme poverty has increased in some of the transition economies of southeastern Europe and many of the countries of the former Soviet Union (from 0.4% in both regions in 1990 to 1.8% and 2.5% in 2002, respectively). While Latin America and the Caribbean have seen marginal reductions in poverty rates, more than 47 million people continue to live in poverty in those regions. With more than 300 million people living in extreme poverty, sub-Saharan Africa continues to have the largest regional proportion of extreme poverty in the world.

While poverty is often a great concern for developing nations, poverty continues to affect developed nations as well. In the United States, the Census Bureau reported that in 2005, more than 12% of Americans were living in poverty (37.0 million people). Furthermore, poverty rates in the United States are higher for some racial/ethnic groups than for non-Hispanic whites. Nearly 25% of blacks, 25% of American Indian and Alaska Natives, and nearly 22% of Hispanics lived in poverty in 2005 in the United States, compared with 8.3% of non-Hispanic whites and 10.9% of Asians. Poverty rates vary widely by geographic areas within the country as well. For example, the proportion of persons in poverty for some areas of the United States is in excess of 40%, whereas other areas have poverty rates of less than 5%.

Poverty and Health Outcomes

Poverty has been shown to influence many health outcomes, including all-cause mortality, infectious diseases, chronic diseases, and health behaviors. Millions of people die from poverty-related diseases each year. Many of the diseases associated with poverty are infectious, such as tuberculosis, diarrheal illnesses, and malaria. Lack of appropriate health care, malnourishment, and disease are likely responsible for more than half a million childbirth-related deaths in women each year. However, with increased industrialization and globalization in developing nations, social and behavioral changes that lead to chronic diseases such as diabetes, hypertension, circulatory diseases, respiratory diseases, and cancer are being observed at higher rates. While these diseases affect persons of all socioeconomic strata, socioeconomic gradients in disease incidence, prevalence, treatment, and survival have been shown, most often revealing that poorer individuals have poorer health outcomes. Also, many risk factors for poor health outcomes are poverty-related, such as substandard housing, environmental pollution, and lack of health insurance. While poverty contributes to the development of disease, there is likely a two-way relationship whereby illness also directly influences poverty due to lost wages and

productivity, plus prohibitive health care costs for those without health insurance. Furthermore, people living in extreme poverty tend to have more frequent and severe disease complications and make greater demands on the health care system.

Childhood Poverty

Children often disproportionately suffer the effects of poverty. Many children die each year from communicable, maternal, perinatal, or nutritional causes, largely due to the effects of poverty. Countries with high rates of child mortality have been shown to have higher rates of extreme poverty. Diseases such as pneumonia, diarrhea, malaria, measles, and HIV/AIDS have been estimated to account for more than half of global under-5 deaths, much of which is associated with preventable malnutrition. In the United States, the poverty rate for children below 18 years was 17.6%, which is higher than any other age group. Nearly 53% of related children below 6 years living in families with a female householder with no husband present were in poverty, compared with 9.9% of children below 6 living in married-couple families. Moreover, the consequences of childhood poverty have been shown to extend into adulthood and effect health throughout the life course, whereby poor and uninsured children are at greater risk of experiencing health problems such as obesity, heart disease, and asthma.

How Poverty Influences Health

One primary way that poverty influences health is through material deprivation and inability to meet basic human needs. Factors such as low income, lack of education and employment opportunities, poor and/or crowded housing conditions, substandard working conditions, insufficient nutrition, unsafe water and inadequate sanitation, environmental pollution, and limited access to health care all contribute to poor health outcomes. Furthermore, poor countries often have limited or weakened health care systems in the context of inadequate governmental infrastructure, along with shortages in health care workers, to adequately respond to the health needs of the poor.

Additional ways that poverty may influence health is through its influence on social relationships, caring for children, psychological health, and lifestyle factors such as smoking, alcohol, exercise, and diet. Poverty may influence the degree of control that people feel they have over their lives. Experiencing such economic and psychosocial strains can lead to unhealthy lifestyle behaviors. Moreover, the cumulative effects of poverty experienced over the life course have been observed. Socioeconomic deprivation experienced in early life has been linked to increased risk of chronic conditions such as cardiovascular disease, respiratory disease, and some cancers in adulthood.

Measuring Poverty

There has been considerable debate over the best way to measure poverty. Poverty is most often considered economic deprivation, with income or consumption as the primary components of the measurement. A poverty line is classified as an income level falling below some minimum level necessary to meet basic needs. Poverty can be conceptualized in absolute or relative terms that vary across time and societies. For example, the World Bank's poverty definitions of $1 or $2 per day are absolute definitions that adjust for the country-specific prices needed to "buy" one or two dollar's worth of goods. Conversely, an example of a commonly used relative definition of poverty is the use of a threshold set at 50% of a country's median income. This particular measurement will constrain the poverty rate to 50% and may not be rooted in a meaningful threshold for unmet basic needs. Relative poverty measures vary from country to country making it difficult to compare rates across countries.

It has been argued that conventional poverty calculations inadequately define poverty and that estimates should include additional items such as indicators of well-being, education, health, access to services, infrastructure, social exclusion and social capital taxes, job-related expenses, and the value of noncash benefits.

Measuring Poverty in the United States

The U.S. Census Bureau releases poverty statistics for the U.S. population every fall. These statistics are based on thresholds that represent the annual amount of income required to support families of various sizes. Despite some criticism, these thresholds remain the official benchmarks for defining economic deprivation for children and adults in the Unites States. Census-based poverty statistics are used in a variety of ways in epidemiologic research. With an increased interest in multilevel modeling and capturing the effects of neighborhood poverty, aggregated regional

poverty measures within various census-defined geographic boundaries (i.e., county, tract, block group, etc.) have been used to determine how area-level poverty influences health outcomes.

Addressing Poverty and the Health Needs of the Poor

Many strategies have been developed to address the health needs of the poor. Although there are thousands of articles addressing poverty in the health literature, relatively few evidence-based interventions have been targeted to serve the poor. There are several distinct but related approaches to bringing attention to the specific health problems faced by the poor. Some approaches target the alleviation of poverty, some target inequality reduction, and some focus on equity enhancement. While these schools of thought differ in some ways, many have argued for moving beyond the epidemiologic understanding of poverty and health into intervention development and policy making. As such, there has been an increase in intercountry research projects that address poverty and health supported by many entities and with the participation of more than 100 countries. Some specific strategies have included the following: preventative and curative interventions for infectious and nutritional-related diseases in children and adults; improved quality and access to education, particularly among women; improved access to necessary drugs, vaccinations, and health care services; microcredit programs, innovative sustainable health care financing alternatives, social and behavioral interventions; family planning services; prevention and rehabilitation of disability; health education programs; and increased access to clean water and sanitation.

The issues surrounding poverty and health have received global attention. Strategies aimed at reducing extreme poverty and hunger were a focus of the Millennium Development Goals (MDGs) developed during the United Nations Millennium Summit of 2000. In addition to poverty and hunger reduction, the MDGs provide a framework for other poverty-related issues such as increasing education; promoting gender equality; reducing child mortality; improving maternal health; combating HIV/AIDS, tuberculosis, and malaria; ensuring environmental sustainability; and developing a global partnership for development.

—Amy B. Dailey

See also Epidemiology in Developing Countries; Health Disparities; Health Economics; Malnutrition, Measurement of; Socioeconomic Classification

Further Readings

Betson, D. M., & Warlick, J. L. (2006). Measuring poverty. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in social epidemiology* (pp. 112–133). San Francisco: Jossey-Bass.

Chuhan, P. (2006). Poverty and inequality. In V. Bhargava (Ed.), *Global issues for global citizens: An introduction to key development challenges.* Washington, DC: World Bank. Retrieved January 31, 2007, from http://siteresources.worldbank.org/EXTABOUTUS/Resources/Chapter2.pdf.

DeNavas-Walt, C., Proctor, B. D., & Hill, L. C. (2006). *Income, poverty, and health insurance coverage in the United States: 2005* (U.S. Census Bureau, Current Population Reports, P60-231). Washington, DC: Government Printing Office.

Gwatkin, D. R. (2000). Health inequalities and the health of the poor: What do we know? What can we do? *Bulletin of the World Health Organization*, *78*, 3–18.

Shaw, M., Dorling, D., & Davey Smith, G. (1999). Poverty, social exclusion, and minorities. In M. Marmot & R. G. Wilkinson (Eds.), *Social determinants of health* (pp. 211–239). New York: Oxford University Press.

United Nations. (2006). *The millennium development goals report*. Retrieved January 31, 2007, from http://unstats.un.org/unsd/mdg/Resources/Static/Products/Progress2006/MDGReport2006.pdf.

Wagstaff, A. (2002). Poverty and health sector inequalities. *Bulletin of the World Health Organization*, *80*, 97–105.

Preclinical Phase of Disease

The preclinical phase of a disease is the period of time during the natural course of a disease where symptoms are not yet apparent, but the disease is biologically present. The concept of a preclinical phase of a disease is most widely referred to in the topic of screening.

Every disease has a natural history that represents the process of the disease in the absence of an intervention. The preclinical phase is one component of this natural history. It begins at the biologic onset of disease and ends when an individual begins experiencing disease-related symptoms. Thus, this phase is when the disease is present but the individual is asymptomatic. When symptoms appear, the preclinical phase ends, and the clinical phase of the disease begins. The length

of the preclinical phase varies for different diseases and between individuals and depends on how quickly the disease progresses. For example, the preclinical phase of many cancers may be quite long, sometimes several years. Many neurodegenerative diseases such as Alzheimer's disease and Creutzfeldt-Jakob disease also have a long preclinical phase.

The preclinical phase of disease can be divided into the nondetectable component and the detectable preclinical phase (DPCP). The former represents the phase that is not yet detectable by screening methods, whereas the latter represents the time window when screening methods can detect the presence of asymptomatic disease. Diseases must have a DPCP to be a candidate for screening.

One of the aims of screening is to detect disease at an early stage so that an intervention can be implemented early and prevent future morbidity and mortality. However, for this to be achieved, the disease in question must have a DPCP. Although there are several criteria for screening, this is a key one of several criteria that indicate whether a screening program could be useful in combating a particular disease. For a screening test to be successful, the disease in question should have a relatively long DPCP with a relatively high prevalence in the population. If it does not, the individuals will pass through to the clinical phase of disease so quickly that the presence of disease will already have been detected by the presence of symptoms, and a screening test will no longer be of use. For example, colorectal cancer has a relatively long DPCP and thus is an ideal candidate for screening by colonoscopy. This is in contrast to other more aggressive tumors that have a very short DPCP, and this screening is not as effective.

For screening to be effective, the DPCP must contain a critical point where intervention is more effective if started at this earlier point than after clinical symptoms appear. Without this early critical point of intervention during the DPCP, screening is not of benefit as early intervention will not make a difference in the progression of disease.

—Kate Bassil

See also Alzheimer's Disease; Cancer; Screening

Further Readings

Hennekens, C. H., & Buring, J. E. (1987). Screening. In S. L. Mayrent (Ed.), *Epidemiology in medicine* (chap. 13). Boston: Little, Brown.

Predictive and Associative Models

See Regression

Pregnancy Risk Assessment and Monitoring System

The Pregnancy Risk Assessment and Monitoring System (PRAMS) is a surveillance project conducted cooperatively by the Centers for Disease Control and Prevention (CDC) and state health departments, with the goal of reducing adverse birth outcomes and improving the health of mothers and infants. PRAMS collects data about pregnancy and the first few months after birth which supplement that available on birth certificates, including maternal experiences and attitudes during pregnancy, while giving birth, and shortly after giving birth. PRAMS was initiated in 1987, and the first participants were the District of Columbia, Indiana, Maine, Michigan, Oklahoma, and West Virginia. Twenty-nine states plus New York City participated in PRAMS in 2006, and six other states have participated at some point in the past.

Two factors led to the establishment of PRAMS: research that indicated that maternal behaviors during pregnancy could influence infant mortality rates and birth weight and the fact that by the mid-1980s birth outcomes were not improving as expected. In particular, the incidence of low-birth-weight babies had changed little over the past 20 years, and infant mortality rates were not declining as rapidly as they had in previous years. PRAMS provides state-specific data for planning and assessing maternal and infant health because responsibility to develop, implement, and evaluate programs to improve birth outcomes rests primarily with individual state health departments. For this reason, some degree of customization is allowed, including differing sample plans so a state may target groups of women who are perceived to be at high risk for poor birth outcomes in a particular state, and some customization of the types of information collected (although there is a common core of information collected in all states).

Each month, PRAMS data are collected from a stratified sample of women in each participating state who

recently gave birth to a live infant. Sample size for each participating state varies between 1,300 and 3,400 women per year and may include oversampling among populations of particular interest to state health officials. Examples of populations include women who gave birth to low-birth-weight children and specific racial and ethnic groups who are perceived by state officials as being at high risk for poor birth outcomes. PRAMS data collection procedures are standardized to allow comparison of data collected in different states. Most data are collected by a questionnaire mailed to selected women 2 to 4 months after delivery; telephone interviews are used if the women selected do not respond after three attempts to contact them by mail.

The PRAMS questionnaire consists of two parts: *core* questions asked in all states and *standard questions* that are chosen from a list of pretested questions supplied by the CDC or developed by the individual state. Topics covered by the Phase 5 Questionnaire, which will be used in the years 2004 to 2008, include attitudes and feelings about the pregnancy, content and source of prenatal care, use of alcohol and tobacco, physical abuse, pregnancy-related illness, infant health care, contraceptive use, and knowledge of pregnancy-related health issues such as the benefits of folic acid and risks of HIV. Topics included in the Phase 5 standard questionnaire cover a broad range of topics, including fertility treatments, child care, HIV testing, breast-feeding, living arrangements, and physical activity. PRAMS questionnaires for Phase 3 (in use 1996–1999), Phase 4 (2000–2003), and Phase 5 (2004–2008) are currently available in English and Spanish for download from the PRAMS Web site.

The PRAMS data sets also include data drawn from birth certificates, including maternal demographics and pregnancy outcomes. To preserve confidentiality, PRAMS files omit information that could identify a particular birth, such as the birth certificate number and the dates of birth of the infant and its mother.

Researchers must apply for permission to use PRAMS data. Those who wish to use data from multiple states must submit a proposal to the CDC; guidelines for developing such a proposal are available on the PRAMS Web site. Those who wish to use PRAMS data from only one state should contact that state's PRAMS coordinator for permission: Contact information for these individuals is available on the PRAMS Web site.

—Sarah Boslaugh

See also Maternal and Child Health Epidemiology; Preterm Birth; Reproductive Epidemiology

Further Readings

Suellentrop, K., Morrow, B., Williams, L., D'Angelo, D., & Centers for Disease Control and Prevention. (2006). Monitoring progress toward achieving maternal and infant healthy people 2010 objectives: 19 states. Pregnancy Risk Assessment Monitoring System (PRAMS), 2000–2003. *MMWR Surveillance Summaries*, *55*(9), 1–11.

Web Sites

Pregnancy Risk Assessment Monitoring (PRAMS): http://www.cdc.gov/prams.

Preterm Birth

Preterm birth is an adverse outcome of pregnancy in which delivery of a live-born infant occurs before the completion of 37 gestational weeks. Infants born between 32 and 36 gestational weeks are considered moderate preterm births, while those delivered earlier than 32 gestational weeks are classified as very preterm births. This entry reviews the occurrence and public health impact of preterm birth and describes the mechanisms and risk factors associated with preterm birth. It also describes approaches used for the detection and prevention of preterm delivery as well as measurement issues encountered in epidemiological studies.

Public Health Impact

Preterm birth is associated with increased infant and childhood morbidity such as neurodevelopmental deficits and behavioral problems. Several adult diseases, such as diabetes, hypertension, and cardiovascular disease, are more likely to occur among preterm infants. Preterm birth is also associated with increased mortality with two thirds of perinatal deaths occurring among preterm infants. Although preterm delivery is associated with birth defects and other causes of mortality, one third of these deaths have been shown to be directly attributable to preterm delivery.

Preterm births can exact a considerable toll on health care systems since most premature babies

require extensive neonatal and postneonatal medical care. In the United States, the disease burden associated with preterm deliveries was estimated at $26 billion a year. The annual cost of neonatal care alone was estimated at $1 billion for preterm births occurring in Canada (excluding costs associated with long-term medical care).

Mechanisms

Preterm birth can occur via at least four major pathophysiologic pathways that may work independently or simultaneously. These include the following:

1. inflammation and infection associated with maternal and fetal cytokine response ($\sim$40% of preterm births);
2. maternal/fetal stress and the production of placental and fetal-membrane derived corticotropin-releasing hormone, which in turn enhances placental estrogen and stimulates fetal cortisol production ($\sim$25% of preterm births);
3. abruption or decidual hemorrhage with thrombin-induced protease expression and disturbances in uterine tone ($\sim$25% of preterm births); and
4. mechanical stretch due to multifetal pregnancy or polyhydramnios-induced uterine or cervical distention ($\sim$10% of preterm births).

These pathways result in activation of the uterine myometrium which can initiate a preterm delivery through uterine contractions, cervical dilation, and premature rupture of the membranes.

Risk Factors

Preterm delivery is a multifactorial outcome in which the cause is unknown in nearly half of preterm births (i.e., idiopathic). Nonidiopathic preterm births can be classified by clinical subtypes, including spontaneous preterm labor, premature membrane rupture, and induction of labor or cesarean section triggered by maternal or fetal indications (e.g., hypertensive disorders of pregnancy, cervical incompetence). Risk factors for preterm birth identified in epidemiological studies include young and old maternal age, African American race, smoking, alcohol use, drug use, nutritional deficiency, poverty/neighborhood factors, stress/anxiety, inadequate prenatal care, inadequate weight gain during pregnancy, hypertension, uterine bleeding, short interconceptual interval, and previous preterm delivery. Systemic maternal infections such as pneumonia and periodontal disease have been associated with preterm delivery in epidemiological studies. Genital tract infections such as bacterial vaginosis have also been linked with an increased risk of preterm delivery. Associations have also been reported in epidemiological studies between risk of preterm birth and various environmental and occupational exposures, including pesticides, organochlorinated compounds (e.g., 1,1-dichloro-2, 2-bis(*p*-chlorophenyl)ethylene) and air pollutants (e.g., sulfur dioxide and particulate matter $< 10\ \mu$m).

Heredity may play a role in the onset of preterm birth, since certain genetic polymorphisms may increase the risk of preterm delivery. Gene-environment interactions are becoming increasingly important considerations for reproductive and developmental epidemiological studies since genetic factors may also modify associations between risk of preterm birth and environmental pollution or other lifestyle/behavioral factors. Molecular epidemiological studies may also be used to determine the contribution of social deprivation, biological differences, and other factors to the racial disparities observed in preterm birth prevalence.

Occurrence

The prevalence of preterm birth ranges from 5% to 15% and is 5% to 10% in most Western societies. The prevalence of preterm delivery was 7.6% in Canada in 2000 and 12.5% in the United States in 2004. Notable disparities by ethnicity exist in the United States for preterm delivery. In 2004, the prevalence of preterm birth was 17.8% for African Americans compared with 11.5% for non-Hispanic Caucasians.

Temporal trends indicate an increase in preterm births over the past few decades in many countries. Most of the increase in preterm prevalence in the United States has occurred among moderate preterm births, since the prevalence of very preterm birth has remained about 2%. This temporal increase may be a reflection of obstetrical intervention in which many more high-risk fetuses are surviving than in previous years due to medical advances. The increased frequency of multiple gestations due to assisted reproductive technologies (i.e., fertility treatment) in many developed countries may also account for some of these trends.

Estimation Techniques

Preterm birth is a common health endpoint examined in epidemiological studies. The validity of preterm birth as an epidemiological endpoint depends on the accuracy of the gestational age estimation technique, since these estimates are not as precise as other clinical measures of fetal development (e.g., birthweight). Common methods for estimating gestational age include ultrasonography, menstrual dating, and clinician estimate. Ultrasound dating based on various measures (e.g., biparietal diameter, crown-rump length, fetal length, abdominal circumference) is considered the most accurate gestational age estimate technique. Gestational age estimates based on last menstrual dating and clinician estimate are commonly recorded on birth certificate data and used to derive preterm birth endpoints in epidemiological studies. Menstrual dating of gestational age depends on maternal recall of the last menstrual period, which can be subject to considerable measurement error. Clinical estimate of gestational age is another method used to gauge the developmental status of the infant but is also subject to measurement error. Random and systematic errors in estimating gestation can result in under- and overestimation of gestational duration and misclassification of the health endpoint being examined. This has potential clinical implications and may also lead to biased relative risk estimates due to false-positive and false-negative cases in the classification of preterm, term, and post-term births.

Detection and Prevention

Early detection and treatment of preterm labor symptoms and maternal infections are keys to reducing the risk for preterm delivery. Biomarkers and other diagnostic tools such as fetal fibronectin, endovaginal ultrasound, and salivary estriol can be used to predict risk of preterm delivery. Avoidance of drugs, alcohol, cigarettes, and other modifiable risk factors are important preventative measures. Adequate weight gain and proper nutrition are essential to the health of a fetus and should be discussed with health care practitioners during prenatal care visits early in pregnancy. As the use of fertility treatment increases the incidence of preterm birth, there will be an increasing need to expand management and treatment alternatives for early deliveries. Promising new hormone treatments, such as 17-alpha-hydroxprogesterone caproate, may decrease the risk of preterm delivery by preventing shortening of the cervix among high-risk pregnant women.

—J. Michael Wright

See also African American Health Issues; Child and Adolescent Health; Gestational Age; Prevalence; Reproductive Epidemiology

Further Readings

Behrman, R. E., & Stith Butler, A. (Eds.). (2006). *Preterm birth: Causes, consequences and prevention.* Washington, DC: National Academies Press.

Callaghan, W. M., MacDorman, M. F., Rasmussen, S. A., Qin, C., & Lackritz, E. M. (2006). The contribution of preterm birth to infant mortality rates in the United States. *Pediatrics*, *118*(4), 1566–1573.

Crider, K. S., Whitehead, N., & Buss, R. M. (2005). Genetic variation associated with preterm birth: A HuGE review. *Genetics in Medicine*, *7*(9), 593–604.

Green, N. S., Damus, K., Simpson, J. L., Iams, J., Reece, E. A., Hobel, C. J., et al. (2005). Research agenda for preterm birth: Recommendations from the March of Dimes. *American Journal of Obstetrics and Gynecology*, *193*(3), 626–635.

Hanke, W., & Jurewicz, J. (2004). The risk of adverse reproductive and developmental disorders due to occupational pesticide exposure: An overview of current epidemiological evidence. *International Journal of Occupational Medicine and Environmental Health*, *17*(2), 223–243.

Savitz, D. A., Terry, J. W., Jr., Dole, N., Thorp, J. M., Jr., Siega-Riz, A. M., & Herring, A. H. (2002). Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. *American Journal of Obstetrics and Gynecology*, *187*(6), 1660–1666.

Sram, R. J., Binkova, B., Dejmek, J., & Bobak, M. (2005). Ambient air pollution and pregnancy outcomes: A review of the literature. *Environmental Health Perspectives*, *113*(4), 378–382.

Prevalence

In epidemiology, the term *prevalence* quantifies the proportion of a population with disease or a particular condition at a specific point in time (sometimes called point prevalence). Prevalence is used widely in the media and by government agencies, insurance companies, epidemiologists, and health care providers. Often confused with prevalence, *incidence* (described in

detail elsewhere) quantifies *new* cases while prevalence describes *existing* cases.

While the term *prevalence rate* is often used synonymously with *prevalence*, the strict definition restricts prevalence to a proportion, not a rate. The difference is in the denominator: Rates describe risk of disease during a given time *interval* among a *population at risk*, while proportions describe the likelihood of disease at a specific *point* in time among the *population.* The point in time may be a specific calendar date or a time that varies from person to person, such as the onset of menopause or puberty, or discharge from the hospital.

For prevalence, the numerator is the number of existing cases or conditions, and the denominator is the total population or group. For example, the prevalence of type 2 diabetes among children aged 2 to 12 years equals the number of children aged 2 to 12 years with type 2 diabetes divided by the total number of children aged 2 to 12 years.

While incidence helps investigators understand the etiology (or cause) of disease, prevalence is especially useful to health system planners and public health professionals. Knowledge of the disease burden in a population, whether global or local, is essential to securing the resources required to fund special services or health promotion programs. For instance, the director of a nursing home must be able to measure the proportion of seniors with Alzheimer's to plan the appropriate level of services for the residents. Legislators and public health professionals need good population statistics to prioritize funding for health promotion programs, such as obesity and smoking cessation. On a community level, understanding the prevalence of English as a second language would be helpful to school administrators. National- and state-level prevalence of behaviors and diseases are usually calculated using data collected systematically from the population through major health surveys, such as the CDC's Behavioral Risk Factor Surveillance Survey (BRFSS), the National Health Interview Survey (NHIS), and National Health and Nutrition Examination Survey (NHANES).

Understanding the difference between prevalence and incidence allows the epidemiologist to apply the terms correctly, define denominators for measures, and conceptualize the study design best suited to a specific research question. In addition, these terms are related mathematically, a property that can prove useful when moving from a measure of incidence to prevalence or vice versa. When the incidence of disease is stable over time, such as in the absence of epidemics or changes in treatment effectiveness, prevalence is the product of the incidence and the average duration of disease or condition ($P = I \times D$). More complex mathematical relationships exist between incidence and prevalence when these assumptions cannot be met.

—Allison Krug and Louise-Anne McNutt

See also Incidence; Proportion; Rate

Further Readings

Hennekens, D. H., & Buring, J. E. (1987). *Epidemiology in medicine.* Boston: Little, Brown.

Last, J. M. (2001). *A dictionary of epidemiology* (Handbooks sponsored by the IEA and WHO). New York: Oxford University Press.

Prevention: Primary, Secondary, and Tertiary

The term *prevention* refers to planning for and taking action to avoid the occurrence of an undesirable event. As it relates to health, prevention aims to hinder the development of diseases or illnesses or avoid injuries. Prevention can be divided into three levels: primary, secondary, and tertiary. The entry discusses each of these levels of prevention and presents examples applied to both communicable and noncommunicable diseases as they relate to individual as well as community efforts.

The importance of prevention in improving health cannot be overlooked. Though the death rates have been decreasing in the United States, preventive actions could further lower these rates. Approximately one third of Americans live with a chronic disease, and almost 70% of the deaths that occur each year are the result of chronic diseases. Furthermore, approximately one third of all U.S. deaths are related to three modifiable health-damaging behaviors—tobacco use, lack of physical activity, and poor eating habits. Establishing healthy habits and making lifestyle changes, which are critical prevention efforts, can significantly decrease the morbidity and mortality rates of Americans.

Levels of Prevention

The first level of prevention, *primary prevention*, sometimes just referred to as *prevention*, is aimed at stopping any occurrence of disease or illness before the disease process begins or taking measures to avoid injury. Thus, many primary prevention activities focus on health education and health promotion programs that are aimed at changing individuals' health behavior and lifestyle.

Injury and illness cannot always be avoided. Some chronic diseases, such as cancer or heart disease, can develop and cause damage before being detected and treated. In such situations, the sooner medical intervention can occur, the greater the chance of preventing death or limiting disability. *Secondary prevention*, sometimes referred to as *intervention*, is aimed at health screening and detection activities that lead to early diagnosis and prompt treatment of a disease or an injury before the disease becomes advanced or the disability becomes severe.

Tertiary prevention, often referred to as *treatment*, is aimed at retraining, reeducating, and rehabilitating the individual who has already incurred a disability. Tertiary prevention measures are applied after the disease, disability, impairment, or dependency has already occurred.

Application of Prevention Principles

The principles of the various levels of prevention can be applied to both communicable and noncommunicable diseases. Furthermore, these principles can be applied to the actions undertaken by single individuals or entire communities.

Prevention of Communicable Diseases

Stopping the spread of communicable diseases in a population is based on stopping the transmission of the pathogens causing the diseases. Successful application of primary, secondary, and tertiary strategies to communicable diseases, particularly primary prevention, resulted in unprecedented declines in both morbidity and mortality during the 20th century. Examples of primary prevention activities undertaken by individuals to stop the spread of communicable diseases include hand washing, proper cooking of foods, and getting immunized against specific diseases. To these can be added community primary prevention measures, including laws dealing with food handling and safety, chlorination of the water supply, the proper collection and disposal of solid waste, and the control of vectors and rodents.

Secondary preventive actions against communicable diseases for individuals can include the self-diagnosis or diagnosis by a physician and treatment of the disease with either over-the-counter medications or those prescribed by a physician. Secondary prevention measures that communities can use are usually aimed at the spread of the disease once it is present in a group of people. Such activities may include case findings and treatment and the reporting of notifiable diseases (those that physicians, clinics, and hospitals are legally required to report to their local health department). Less commonly, communities may isolate or quarantine those infected or exposed, respectfully.

The tertiary preventive measures for control of communicable disease in individuals usually include convalescence from infection, recovery to health, and return to normal activities. Tertiary prevention measures at the community level are aimed at the recurrence of the disease. An example would be the removal, embalming, and burial of the dead.

Prevention of Noncommunicable Diseases

Unlike communicable diseases that are caused by pathogens, the strategies used to prevent noncommunicable diseases focus on the risk factors associated with a particular disease. Thus, the prevention principles are applied a bit differently to noncommunicable diseases, but as with communicable diseases, they can be applied to both individual and community activities. Primary prevention measures for noncommunicable diseases at the individual level begin with a solid education about health and health practices. With such knowledge, individuals can take the necessary steps to prevent noncommunicable disease such as getting enough exercise, maintaining a healthy body weight, eating properly, wearing safety belts, and avoiding excess exposure to the sun by wearing sunscreen. Community primary prevention measures include providing a safe and healthful environment. Examples may include smoke-free environments, exercise trails, and appropriate lighting in parking lots and on sidewalks to reduce injury and crime.

Secondary prevention measures for individuals for noncommunicable diseases include actions, for example, personal screenings such as self-examination for

cancer of the testes or breasts, or participating in screenings provided by the medical community such as mammograms, Pap tests, or PSA (prostate-specific antigen) tests for cancer. The goal of such screenings is early detection, referral, and prompt treatment to either cure the disease or slow the progress of disease, disability, disorder, or death. Behavior change programs are another example of individual secondary prevention efforts. Smoking cessation, weight loss, stress reduction, or early admission in to a drug prevention program are examples of such behavior change programs.

Community secondary prevention measures include the provisions of mass screenings and case finding for chronic disease and the provision of adequate health care personnel and facilities to conduct such screenings. Examples may include blood pressure screenings provided by the paramedics at the local fire house or a local voluntary health agency partnering with a cancer center offering the various array of cancer screenings.

Tertiary prevention measures for noncommunicable diseases by individuals often require significant lifestyle changes. For a person with diabetes, it may include testing oneself regularly for blood sugar levels, taking prescribed medications or injections of insulin, and faithfully watching one's diet and getting enough exercise. Patient education, after care, support groups, and health counseling are some important community health promotion components of tertiary prevention efforts.

It has been commonly reported that in the United States, 95% of all health care dollars is used for medical care services, while only 5% is used for preventive activities. If the health of the American people is going to improve and health care costs are going to be controlled, a greater emphasis needs to be placed on prevention.

—James F. McKenzie and Denise M. Seabert

See also Community Health; Health Behavior; Notifiable Disease; Quarantine and Isolation; Screening

Further Readings

Centers for Disease Control and Prevention. (2003). *The power of prevention: Reducing the health and economic burden of chronic disease.* Atlanta, GA: U.S. Department of Health and Human Services. Retrieved July 24, 2007, from http://www.cdc.gov/nccdphp/publications/PowerOfPrevention.

McGinnis, J. M., Williams-Russo, P., & Knickman, J. R. (2002). The case for more active attention to health promotion. *Health Affairs*, *21*(2), 78–93.

McKenzie, J. F., Neiger, B. L., & Smeltzer, J. L. (2005). *Planning, implementing, and evaluating health promotion programs: A primer* (4th ed.). San Francisco: Benjamin Cummings.

McKenzie, J. F., Pinger, R. R., & Kotecki, J. E. (2005). *An introduction to community health* (5th ed.). Sudbury, MA: Jones & Bartlett.

Timmreck, T. C. (1998). *An introduction to epidemiology* (2nd ed.). Sudbury, MA: Jones & Bartlett.

Probability Sample

A probability sample is one in which members are chosen from a target population using methods that rely on chance such as random number tables. In probability sampling, all members of the target population have a nonzero probability of being chosen; the probability of selection for each can be calculated. Probability samples are superior to nonprobability samples in that the extent to which the sample varies from the target population can be calculated, and in the absence of other biases, study results may be generalizable to the target populations.

Probability Versus Nonprobability Sampling

Due to limits on resources and time, researchers are rarely able to observe every member of a target population, or the group for which a researcher wishes to generalize the results of a study. Instead, a subset must be chosen. Members may be selected for study using either nonprobability or probability sampling methods. In nonprobability sampling, selection occurs in a nonrandom fashion usually based on availability. Two examples of nonprobability sampling methods are convenience and snow-ball sampling. Because all members of a population chosen via nonprobability sampling methods do not have a nonzero chance of being selected for the study, it is not possible to determine how closely a nonprobability sample resembles the target population and, therefore, results from these studies are not generalizable to the entire target population. However, probability sampling, which involves selection of members from the target population using random selection techniques, produces results that, in

the absence of other biases, are generalizable. In probability sampling, all members of the target population have a nonzero opportunity of being chosen to be in the sample and the probability that any given one will be chosen can be calculated.

Types of Probability Sampling

Simple random sampling, systematic sampling, and stratified or cluster sampling are types of probability sampling. In simple random sampling, members are randomly and independently chosen from a list of all target population members; every member of the target population has an equal chance of being chosen for participation in a study. In systematic sampling, a sample is chosen by selecting the first member from a list at random and then by taking every *k*th member from the population list thereafter. In stratified or cluster sampling, a population is first subdivided into groups that share at least one common characteristic, then a sample is chosen from each stratum using simple random or systematic selection.

Probability Sampling and Sampling Error

The degree to which a sample obtained through probability sampling varies from the target population is measured by estimating the sampling error. Sampling error arises when only part of a population is observed; for any given target population, many alternative sample realizations, or sets of members or individuals selected for a sample, can be obtained by employing a given sampling method, and each may lead to a different summary statistic, such as mean, for the parameter being studied. Theoretically, sampling error is derived from the amount of variation that exists between the summary statistics for all possible realizations. In practice, the standard error of the study sample is used to construct a confidence interval for which with a certain level of confidence the true parameter of interest lies.

Probability Sampling and Bias

The use of probability sampling methods does not guarantee that a sample accurately represents the target population; various biases may affect the results of a study despite the use of probability sampling techniques. For example, sample bias may occur in studies that have low response rates as study participants who choose to respond to a survey may differ significantly from those who choose not to, therefore causing the study summary statistic to vary from that of the target population. Sample size, which is inversely correlated with sampling error, is also important in obtaining precise measures.

—Michelle Kirian

See also Bias; Convenience Sample; Sampling Techniques; Stratified Methods

Further Readings

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). Sample design and standard error. In *Survey methodology* (pp. 93–136). Hoboken, NJ: Wiley.

Scientists and Engineers Statistical Data System. (2001, September). *Sampling errors for SESTAT.* Retrieved March 1, 2007, from http://sestat.nsf.gov/docs/stderr00.html.

Statistics Canada. (2006, August). *Sampling methods.* Retrieved March 1, 2007, from http://www.statcan.ca/english/edu/power/ch13/first13.htm.

Program Evaluation

Evaluation has at its root the word *value.* Program evaluation is the part of the evaluation field that determines the merit or worth of a program. A program is a set of planned activities designed to reach a predetermined goal—for instance, to encourage employees in a company to begin and maintain an exercise program or to discourage teenagers from beginning to smoke. Program evaluation consists of the activities that determine the value, merit, or worth of the program—whether the program activities are making a difference and accomplishing the program's goal (outcomes). The question may be asked, "Why evaluate?" Evaluation is the process that allows decision makers to determine whether programs are making a difference and to identify changes that may be necessary for success. It provides program planners, policymakers, legislators, and other decision-making stakeholders with information on which to base decisions concerning program continuation or closure. Since epidemiologists are often in the roles of determining needs and

influencing policy, understanding policy is important. An example of these activities would be the input of epidemiologists in determining government policy surrounding immunization of school-age children. Program evaluation provided evidence that immunizations make a difference. Epidemiologists used that information to advocate for a policy requiring school-age children to be immunized prior to attending public schools.

Program evaluation is considered a *transdiscipline*. That is, it is an area of inquiry that provides services to many disciplines, using methodologies drawn from social and applied sciences and applied broadly across various disciplines such as social services, industry and business, health care, education, and mental health care, among others. The application of a transdiscipline is not uniform across all contexts; some methods are more useful and applicable in some contexts or situations than in others. Michael Scriven describes transdisciplines as disciplines such as logic and statistics that provide tools for other disciplines such as sociology and psychology. Transdisciplines apply across a broad range of inquiry and creative endeavors, yet maintain disciplinary autonomy.

Another way to view program evaluation is through the concept of *appreciative* inquiry (AI). Hallie Preskill and Anne Coghlan describe appreciative inquiry as a method of inquiry that is participatory, collaborative, and systematic in determining an organization's capacity to develop a positive potential in planning its preferred future. In evaluation, appreciative inquiry is a process that promotes positive change, especially in organizations, to build the capacities of the organization.

In the past 10 to 15 years, program evaluation as a discipline has come into its own. This has been one of the concrete long-term outcomes of the federal mandate for accountability through the Government Performance Results Act (GPRA) of 1993 and, more recently, the Program Assessment Rating Tool (PART) developed by the Office of Management and Budget and released in 2002. PART is a questionnaire used to evaluate federally funded programs in terms of their purpose, design, planning, management, results, and accountability to determine its overall effectiveness. The PART established another layer of accountability at the federal level and ensured that evaluation activities are systematically included in program implementation. Yet to truly understand program evaluation as a field, one needs an understanding of its history, knowledge of the key concepts, and its applications in epidemiology and public health.

History of Program Evaluation

Program evaluation as it is known today evolved over several hundred years. Michael Quinn Patton subscribes to the view that program evaluation was used by Daniel when in the lion's den. However, program evaluation as a systematic form of inquiry has its origins in the 19th century. During the 1800s, the British government appointed commissions that reviewed (i.e., evaluated) educational institutions of the time. These commissions established external boards to inspect schools. In the United States in the mid 1800s, Massachusetts assessed student achievement and used those assessments for school comparisons. The accreditation movement for schools and institutions of higher learning (secondary and postsecondary schools) began in the late 1800s.

The early 1900s brought the review of social service and health programs that addressed problems such as slum conditions and infectious diseases. Development of the educational testing movement also began in the early 1900s, due in large part to the psychometric work of E. L. Thorndike and the intelligence testing work of Alfred Binet and Louis Terman. The use of norm-referenced and then criterion-referenced tests moved evaluation of educational programs forward significantly. Egon Guba and Yvonna Lincoln call this first generation of evaluation the measurement generation.

Simultaneously, social service fields were establishing methods for assessing efficiency of their social programs. The management movement of Fredrick Taylor was of notable significance. Yet these efforts were neither widespread, nor did they have government support. The Great Depression changed all that as there was a proliferation of government-supported entitlement and social services programs such as welfare, health, and urban development. These field-based programs served as living laboratories for the applied social scientists who were attempting to determine the effects of these programs.

During World War II, social scientists established mechanisms by which governmental programs were developed to assist the military and, later, the programs for returning veterans. The results of psychological and personality testing for job placement received much attention. It was also during this time

that skill-based teaching and testing were initiated. This objective-based teaching guided curriculum development, and the extent to which students achieved those objectives was then described by developers. This largely descriptive process identified strengths and weaknesses of the curriculum, rather than the abilities of the students. Ralph Tyler did much to advance the use of objectives expressed in measurable terms. Guba and Lincoln refer to this period as the second generation of evaluation.

The second half of the 20th century nurtured the advancements that led to evaluation as it is known today. The concepts of judgment, merit, and worth became the keystone for evaluation. Many scholars developed various models that employed criteria against which programs were judged (e.g., Robert Stake's countenance model; Daniel Stufflebeam's CIPP (context, input, process, product) model; Malcolm Provus's discrepancy evaluation model; Michael Scriven's goal-free model; and Elliot Eisner's connoisseurship model. Guba and Lincoln describe this as the third generation of evaluation. They developed yet another model, the responsive constructivist evaluation model, which they labeled fourth generation.

Government policies in the latter half of the 20th century supported and even demanded evaluation. The Elementary and Secondary Education Act (ESEA) of 1965 was probably the single piece of legislation most responsible for moving program evaluation forward. This legislation funneled massive amounts of governmental funds to local, state, and regional educational institutions and at the same time required the recipients of the funds to provide the funding agency with an evaluation report detailing the program results supported by federal funds.

By the late 1900s, evaluation had come into its own as a profession. The American Evaluation Association was formed in 1986 through a merger of the Evaluation Research Society and the Evaluation Network. Universities had developed graduate programs for preparing evaluation specialists, professional development institutes provided continuing education for practicing evaluators, standards of practice were developed and approved, and scholars were developing theories of program evaluation.

To ensure effective use of scarce fiscal resources, the federal government passed the GPRA and implemented PART mandating evaluations of programs both funded by and housed within the federal government. Evaluation societies proliferated internationally, with more than 20 societies existing around the world. The International Organization for Cooperation in Evaluation (IOCE) ratified its constitution in 2003. As an organization of various national and international professional evaluation societies, the IOCE mission is promoting cooperation and partnership in evaluation worldwide through the exchange of information, ideas, and resources and promoting a high level of professional standards.

Key Evaluation Concepts

Any discussion about program evaluation will typically involve some, if not all, of the following terms: informal evaluation, formal evaluation, formative evaluation, summative evaluation, needs assessment, process evaluation, internal evaluation, external evaluation, logic modeling, outcome evaluation, qualitative evaluation, and quantitative evaluation. Understanding how these terms form the framework of program evaluation provides the reader with a foundation to understand program evaluation.

Informal and Formal Evaluation

Informal evaluation is the process used by individuals in daily activities to make judgments and to make choices based on those judgments. Consumers use informal evaluation in choosing a brand of cereal or canned vegetables. Teachers make observations of students and form judgments of the student's ability. Typically informal evaluations are unsystematic, based on incomplete evidence. For example, in choosing a cereal, it is unlikely that an individual will have tasted every type of cereal available in the store. In addition, other stores may have different brands. Consequently, informal evaluation typically provides incomplete data on which to base the decision of value or worth. Personal and situational biases also contribute to this inadequacy of informal evaluation. Nevertheless, even though informal evaluations may provide incomplete data, they may be the only evaluation possible in many situations.

Formal evaluation is systematic, planned, and context specific. Typically, specific methods are applied to determining the value or merit of a program. One can consider program evaluation, or evaluation of any object, a continuum from informal to formal. Finding the balance providing between unstructured evaluation

and rigorous but excessive formal evaluation is the challenge facing evaluators.

Formative and Summative Evaluation

Formative and *summative evaluation* are terms that Michael Scriven coined in 1967 to distinguish between evaluations that were conducted for program improvement only (formative) and evaluations that were typically conducted for decision making only (summative). Formative evaluation is the evaluation typically occurring during the program's implementation and often evaluates only a part of a program. Formative evaluation allows for midcourse corrections in program implementation. Summative evaluation is the evaluation typically occurring after the program implementation is completed. It provides information to decision makers for a "go-no go" decision on program continuation.

The audiences for the reports of formative and summative evaluation are typically different. The audiences for formative evaluation reports are the program planners or program staff. The audiences for summative evaluation reports are consumers, funding sources, policymakers, and other decision-making stakeholders.

Needs Assessment

Programs are typically planned to answer questions about a condition currently existing. To garner clear and unambiguous information about the nature of the condition, evaluators typically conduct a "needs assessment." The information gathered helps establish the extent to which a need or problem exists and provides information for making recommendations to address that problem.

Process Evaluation

A process evaluation determines the "how" of a program. It details the delivery, the administrative structure, and the successes encountered as the program is implemented. One way to look at process evaluation is to structure the process evaluation around the following questions: (1) What challenges were encountered? (2) What was done to overcome those challenges? and (3) What lessons were learned from this approach? Challenges may be both positive and negative. Answering these questions helps program planners and decision makers when similar programs are being planned.

Outcome Evaluation

An outcome evaluation documents what changes have occurred as a result of implementing a program. Changes can occur in program participants, in those who interact with the program participants, and in the communities in which program participants live. Changes may occur in knowledge, behavior, or practice or in social, environmental, or economic conditions. Outcome evaluations are often conducted as a part of a comprehensive evaluation that includes a needs assessment, a process evaluation, and an output evaluation.

One often hears the term *outcome evaluation* being used synonymously with *summative evaluation.* Outcome evaluations may be summative, and summative evaluations may not describe outcomes. Huey-Tsyh Chen proposed a typology showing the relationship between formative and summative evaluations and among needs assessment, process, and outcome evaluations. It is clear that needs assessments, process, and outcome evaluations can be either formative or summative, depending on the questions being answered by the evaluation.

Internal Evaluation and External Evaluation

When an evaluation is conducted by an employee of the program, the evaluation is typically considered an internal evaluation. When an evaluation is conducted by an individual who is outside the organization conducting the program, the evaluation is typically considered an external evaluation. Although the use of these terms seems reasonably clear, there may be variations whereby an employee of an organization (internal) is not part of the program being conducted (external) (e.g., when there is a multisite program and an evaluation team unfamiliar with the program being evaluated is sent from corporate headquarters).

It is important to consider the advantages and disadvantages of each position when designing evaluations. The external evaluator typically brings greater credibility and perhaps objectivity to the task as well as greater specialized evaluation knowledge. The internal evaluator typically understands the corporate and programmatic culture. The external evaluator will not know the corporate culture, while the internal evaluator may be burdened by personal and situational biases related to the corporate/programmatic culture.

Blaine Worthen, James Sanders, and Jody Fitzpatrick (2003) propose a 2×2 matrix for the combinations of

formative and summative evaluations and internal and external evaluation. They are formative-internal, formative-external, summative-internal, and summative-external. Most commonly, individual evaluators are either formative-internal (because of their knowledge of the program) or summative-external (because of the perceived objectivity).

Logic Modeling

Logic modeling is simply a map of the program executed in a series of if-then statements. Logic modeling is employed during the program planning stages to detail the long-term expectations (often called impacts) of the program. Program planners describe what conditions will change if the program is successful. They attempt to outline what difference the program will make. Program planners then look at medium-term outcomes (often called intermediate outcomes) and short-term outcomes (often called immediate outcomes). One will also see the term *proximal outcome* used to describe short-term and medium-term outcomes and *distal outcomes* to describe long-term outcomes.

Once these outcomes are identified in a time frame, the question is posed by program planners as to what must be done to reach these outcomes. Specifically, what activities will need to be conducted to which audience with what resources? The activities and audience are typically called outputs, while the resources are typically called inputs.

Outputs consist of the number of sessions, encounters, publications, and so on that the specified number of targeted audience will receive in the time frame specified. Inputs are the budgets, personnel, time, equipment, materials, facilities, and so on that are required to perform the activities to the targeted audiences.

Logic models are often linear, although not necessarily so. As with formative evaluations, midcourse corrections can be made in models, implying that the models are more accurately an iterative rather than a linear activity.

Quantitative Evaluation and Qualitative Evaluation

Evaluation traditionally employed methods from the social and applied sciences, such as the quasi-experimental designs, objective measurement techniques, and statistical analyses. These quantitative methods were employed in the comparisons, providing the evaluator with indicators of validity and reliability for making judgments. Quantitative evaluation drawing from applied social science research provides information about two important criteria found in social science research: *internal validity*, or causality, and *external validity*, or generalizability to other settings and times. It is difficult to secure measures for these criteria using the rich narrative found in qualitative evaluation.

Qualitative evaluation, on the other hand, employs qualitative, or nonnumerical, data. Sources of these data are verbal descriptions of observations, interviews, and individual collective perceptions (such as data gathered from focus groups). One could not reasonably use the same criteria one used with quantitative data to determine the value of the program evaluated using a qualitative evaluation. Instead, one would employ *accuracy*, or the extent obtained data are reflective of the real situation; *utility*, or the extent results serve practical needs; *feasibility*, or the degree of prudence, diplomacy, and reality employed; and *propriety*, or the legal and ethical parameters of an evaluation.

The value of each form is clear. Numeric data are typically more precise, while narrative data are typically more descriptively rich. Using both, called *mixed methods*, is the approach more often used by today's evaluators. In using mixed methods, evaluators draw from across disciplinary boundaries and employ methodologies from such varying disciplines as agronomy, anthropology, sociology, psychology, philosophy, mathematics, history, and economics. Policy analysis often draws from legal frameworks.

These methodological approaches are classified into five evaluation approaches relating to the models employed. These approaches are objective-oriented, management-oriented, consumer-oriented, expertise-oriented, and participant-oriented. Qualitative and quantitative methods are used in varying degrees in each of these, so that these approaches form a continuum from utilitarian evaluation to intuitionist-pluralist evaluation, and quantitative to qualitative evaluation.

Application of Program Evaluation

Often, the incidence and prevalence of an event change as a result of a program or intervention. Program evaluation is the discipline that will aid in determining what change has occurred and assist in identifying the attribution of that change to a specific

intervention. Although program evaluation does not explicitly identify causality, it is the discipline that will provide information for decision making by determining the merit or worth of a program and often provides the tools for determining causality later.

—Molly Engle

See also Economic Evaluation; Qualitative Methods in Epidemiology; Quantitative Methods in Epidemiology; Study Design

Further Readings

Chelimski, E., & Shaddish, W. R. (Eds.). (1997). *Evaluation for the 21st century: A handbook.* Thousand Oaks, CA: Sage.

Chen, H.-T. (1990). *Theory-driven evaluations.* Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding and improving policies and programs.* San Francisco: Jossey-Bass.

Patton, M. Q. (1996). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Preskill, H., & Coghan, A. T. (Eds.). (2003). *Using appreciative inquiry in evaluation.* New Directions for Evaluation, no. 100. San Francisco: Jossey-Bass.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation* (American Educational Research Association Monograph Series on Evaluation, No. 1, pp. 39–83). Chicago: Rand McNally.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Shaddish, W. R., Cook, T. D., & Leviton, L. (1991). *Foundations of program evaluation.* Newbury Park, CA: Sage.

Stufflebeam, D. L. (2001). Evaluation models. In G. T. Gary & J. C. Greene (Eds.), *New directions for evaluation* (Vol. 89, pp. 7–99). San Francisco: Jossey-Bass.

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2004). *Handbook of practical program evaluation* (2nd ed.). San Francisco: Jossey-Bass.

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (2003). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). New York: Allyn & Bacon.

Web Sites

Government Performance and Results Act (1993): http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html.

International Organization for Cooperation in Evaluation: http://ioce.net/index.shtml.

Program Assessment Rating Tool: http://www.whitehouse.gov/omb/part.

Propensity Score

Propensity score adjustment is a method of adjusting for all covariates in an observational case-control study, using scalar matching. In case-control studies, the goal is typically to determine if one group (cases) of subjects has a different outcome than another group (controls). These groups might be defined by which treatment they received or which factor they were exposed to, and the purpose of the study is to determine if the treatments or factors result in different outcomes in the cases than in the controls. For example, we can observe people who smoke (cases) and people who don't smoke (controls) and compare the rates of cancer between the two groups. Because people cannot ethically be assigned to one or the other condition (smoking or nonsmoking) we have to accept the groupings that exist. However, because random assignment to condition was not used, the two groups very likely differ on other factors, which can introduce bias into the study. Historically epidemiologists have dealt with this issue by matching subjects in the case group with subjects in control group based on observed covariates, for example, age category and gender, to attempt to remove the influence of these factors on the outcome by equalizing their distribution in the two groups. However, matching can be performed on only a limited number of covariates before the sample size within each matching group becomes too small for statistical analysis. For instance, if we divided age into four categories (e.g., < 30, 31–50, 51–70, ≥ 71), then matching on age and gender would divide the sample into eight separate subgroups. In addition, by dividing a continuous variable (age) into categories, we are losing some of the information contained in the variable.

Propensity score adjustment overcomes this limitation using scalar matching as follows. Let the data measured on subjects be classified into three sets of variables: X is the set of all covariates to be adjusted for, Y is the group membership (case or control, smoker or nonsmoker in this example), and Z is the outcome (cancer or no cancer in this example). Each subject

observed has the set of variables (X, Y, Z) measured on them. Note that the covariates X may be continuous, categorical, or both, and the outcome variable Z may also be either categorical (e.g., develop cancer or not) or continuous (e.g., number of pounds lost).

In the simplest propensity-score-matching approach, a logistic regression model is fit using all the covariates in X to predict the group membership Y. Note that this analysis excludes the use of the outcome Z in the model fitting. The logistic regression model assigns each subject a predicted log-odds value for belonging to the smoking group (case), whether they smoked or not. Matching cases with controls can then be done based on the log-odds of smoking number (the scalar) calculated for each subject. Matching cases with controls who have a similar log-odds of smoking value results in an adjustment for all the covariates in X.

After matching cases with controls, a comparison of the outcome Z in the matched sets is done using standard statistical methods. If matching is done on a one-to-one basis, with a categorical outcome, as in the smoking example, a paired test of proportion (McNemar's test) can be used to determine if death is more likely to occur in the smokers than in the nonsmokers). Stratified matching can also be done by grouping subjects according to the distribution of the log-odds scalar. For instance, to form five strata in the smoking study, the first group would be the cases and controls whose log-odds of smoking are in the lowest 20th percentile, the next group in the 21st to 40th percentile, and so on. The analysis of the outcome can then be performed on each subgroup separately, using Fisher's exact test, or over all the subgroups, using the Mantel-Haenzel test.

—William D. Shannon

See also Bias; Logistic Regression; Observational Studies; Study Design

Further Readings

Connors, A., Speroff, T., Dawson, N., Thomas, C., Harrell, F. E., Jr., Wagner, D., et al. (1996). The effectiveness of right heart catheterization in the initial care of critically ill patients. *Journal of the American Medical Association*, *276*, 889–897.

Juang, P., Skledar, S. J., Zgheib, N. K., Paterson, D. L., Vergis, E. N., Shannon W. D., et al. (2007). Clinical outcomes of intravenous immune globulin in severe *Clostridium difficile*-induced diarrhea. *American Journal of Infection Control*, *35*(2), 131–137.

Rosenbaum, P. (2002). *Observational studies* (2nd ed.). New York: Springer-Verlag.

Rosenbaum, P., & Rubin, D. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*, 516–524.

Proportion

The proportion is a statistic that is used to describe how much of a population has a particular characteristic or attribute and is usually expressed as a fraction or decimal. The defining characteristic of the proportion, as distinct from the ratio, is that every individual in the numerator of a proportion is also included in the denominator. Consider a population in which each member either has or does not have a specified attribute. The population proportion is the percentage (or rate) of the entire population that has the specified attribute. For examples, we might be interested in the proportion of U.S. adults who have health insurance or in comparing the proportions of prevalence of CF antibody to Para influenza I virus among boys and girls in the age group 5 to 9 years. In the first case, the population consists of all U.S. adults and the specified attribute is "has health insurance." For the second case, the population consists of all boys and girls in the age group of 5 to 9 years.

Frequently, the population under consideration is large, and determining the population proportion by taking a census is therefore usually impractical and often impossible; for instance, imagine trying to interview every U.S. adult for the purpose of ascertaining the proportion that have health insurance. Thus, in practice, we mostly rely on sampling and use the sample data to make inferences about the population proportion.

The sample proportion is the percentage of a sample from the population that has the specified attribute. The sample proportion $\hat{p}$ can be computed by the formula

$$\hat{p} = \frac{x}{n},$$

where x denotes the number of members in the sample that have the specified attribute and n denotes the sample size. For example, a study is undertaken to

compare the rates of prevalence of CF antibody to Para influenza I virus among boys and girls in the age group of 5 to 9 years. Among 113 boys tested, 34 are found to have the antibody; among 139 girls tested, 54 have the antibody. Let p_1 denote the population proportion of boys who have the CF antibody and p_2 the population of girls who have the CF antibody. Then sample proportions are

$$\hat{p}_1 = \frac{34}{113} = 0.301 \text{ and } \hat{p}_2 = \frac{54}{139} = 0.388.$$

We may use these sample proportions, in accordance with statistical theory, to make inferences about the difference of these two population proportions.

—Renjin Tu

See also Confidence Interval; Hypothesis Testing; Inferential and Descriptive Statistics; Ratio

Further Readings

Johnson, R. A., & Bhattachayya, G. K. (2006). *Statistics principles and methods* (5th ed). Hoboken, NJ: Wiley.
Weiss, N. A. (2005). *Introductory statistics* (7th ed.). Boston: Addison-Wesley.

Psychiatric Epidemiology

Psychiatric epidemiology is the study of distribution, determinants, and causes of psychiatric conditions or mental health in human populations. The term *psychiatric epidemiology* was first coined at the 1949 Annual Conference of the Milbank Memorial Fund and was later documented in a Milbank Memorial Fund publication in 1950. Long before then, however, studies of mental health in populations had been conducted. Edward Jarvis, a mid-19th-century physician, described the distribution of "insanity" and "idiocy" and health care utilization in a wide range of facilities in Massachusetts from 1850 through 1855. This period marks the beginning of descriptive epidemiology where focused efforts were being made to describe disease distribution in the population. Not long after, psychiatric research began to use analytical epidemiology techniques as well. With methods still in use today, researchers examined hypotheses using various study designs, such as case-control and cohort studies, aimed to understand the nature, etiology, and prognosis of mental disorders.

Diagnosis

Psychiatric disorders include disturbances of thinking, such as schizophrenia, dementia, and mental retardation; disturbances of feeling, such as bipolar disorder, anxiety, and depression; and disturbances of acting, such as alcohol and drug disorders and antisocial disorders. Important childhood psychiatric disorders include autism, depression, and attention deficit disorders. Psychiatric disorders always involve biological or neurological adaptation of some sort and often include disruption of social life as well. These disorders are among the most disabling in the world, accounting for higher percentages of disability-adjusted life years than most other categories of disorder.

The diagnosis of psychiatric disorder is made almost totally on the basis of observed symptoms and behaviors because, to date, no biomarkers or laboratory tests are conclusive in diagnosis. The most commonly used diagnostic systems in psychiatry are the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) and the International Classification of Diseases (ICD). Since its birth in 1952, DSM has been revised a number of times, with the most recent version being *DSM-IV*. *DSM-V* is expected in the near future. The number of psychiatric disorders listed increased from 159 in *DSM-II*, to 227 in *DSM-III*, to 357 in *DSM-IV*, and more are expected in *DSM-V*. The number of disorders has increased with revisions of the ICD also. As a result of the increasing numbers of diagnoses, as expected, the number of comorbid diagnoses also increased. This makes it more challenging to determine independent etiologies or measure an impact from a single disorder.

The use of standard criteria to define a mental disorder allows measurement of prevalence, related impairments, financial burden, and resulting mortality, and also makes it possible to compare these features across different regions, sex, and ethnic groups, as well as groups defined by other characteristics. But despite the many advances that have been accomplished in classifying mental disorders since the 1950s, case definitions are still controversial. Since the birth of the DSM and ICD, categorical diagnoses have been used for psychiatric disorders. However, many argue that mental disorders are best conceptualized as dimensional and that diagnostic thresholds may not be meaningful for etiologic

determination. A categorical diagnostic decision is made depending on whether a patient meets or fails to meet a series of criteria, whereas a dimensional system acknowledges the continuum of symptom severity that may fall above or below a categorical diagnostic threshold. Most researchers suggest that for nosology (the systematic classification of diseases), the need for retaining categorical distinctions is compelling but that dimensional models may be more useful for clinical treatment, epidemiologic research, and policy development.

Psychiatric screening instruments have served the purpose of measuring symptoms and behaviors dimensionally. Historically, screening instruments were developed before diagnostic schedules/systems were established. Screening instruments are shorter and simpler and can be either filled out by participants themselves or by research staff with minimal training. On the other hand, structured or semistructured diagnostic schedules are more comprehensive in coverage of disorders, but more time-consuming, and require psychiatric professionals or trained interviewers to administer. For example, the Center for Epidemiologic Studies Depression Scale (CES-D), the most commonly used short screener for depression in the U.S. epidemiologic studies, consists of 20 items describing behaviors and feelings such as "I felt fearful" and "I talked less than usual" and takes about 5 to 10 min for most people to complete. A cutoff score of 16 was suggested by epidemiologic studies, meaning that a person whose score is greater than or equal to 16 is considered likely to be clinically depressed. In contrast, a diagnosis of major depressive disorder (MDD) using *DSM-IV* criteria requires a clinical evaluation and at least five of nine symptoms for 2 weeks or more, such as "A significantly reduced level of interest or pleasure in most or all activities" and "Behavior that is agitated or slowed down." Because prevalence of a psychiatric disorder (in this example, depression) is highly influenced by its definition, it is not surprising to see differences in CES-D-determined prevalence and *DSM-IV*-diagnosed prevalence.

Study Designs

Epidemiologic study designs developed to study infectious diseases and chronic diseases are frequently used in psychiatric epidemiologic studies. There are also hybrid studies that modify traditional designs and can be tailored to meet a specific study's needs. We describe below some common epidemiologic study designs frequently used in psychiatric epidemiologic studies, with particular focus on unique challenges in studying mental health using each design.

Incidence and Prevalence

Incidence and prevalence are measures of the extent of disorder in the population. Incidence is a measure of the occurrence of new cases per unit of time, and prevalence is the proportion of cases in the population at a defined time point. When depression is defined as a positive score on the CES-D, its prevalence will probably be higher for the same population than if depression were defined according to *DSM-IV* criteria for an MDD. This is because the CES-D taps a wider range of less severe symptoms. Comparisons of the incidence and prevalence of psychiatric disorders in different populations or geographic regions, require that studies use a consistent definition of depression. Other factors may also reflect the measured incidence and prevalence of depression, including, but not limited to, the awareness of the condition in a population, relative access to the health care system, and cultural acceptance of depressive symptoms.

Many persons with psychiatric disorders do not enter into treatment, and thus, population-based incidence and prevalence studies are preferred to clinic-based studies whenever feasible. In addition, population-based studies are more likely to detect a psychiatric condition (e.g., depressed mood) that has a wide spectrum and is common in a population. Individuals with such a condition might not be seen in clinical settings because the condition is so common and people might not be aware of the need for medical treatment, or the condition might be overshadowed by other more severe disorders.

Although measures of incidence are generally preferred over measures of prevalence in risk factor epidemiology, incidence can be difficult to measure for some psychiatric disorders, because many have an insidious onset. Therefore, it is often difficult to determine the precise time of onset and, consequently, what constitutes a new occurrence of disease. This is particularly true for disorders that begin as early as in utero (before birth). In such cases, prevalence rather than incidence is more commonly used as a measure of psychiatric disorders.

Cohort Studies

Cohort studies provide great potential to study the sequence of exposure and event/outcome, which is

essential to determine causality, and allow researchers to investigate psychiatric comorbidities (e.g., MDD, anxiety symptoms, post-traumatic stress disorders) after an exposure of interest (e.g., traumatic stress). A *prospective cohort study* is most successful when the duration between exposure and detectable outcome is not exceedingly lengthy and the outcome is not a rare event (e.g., depression). Because cohort studies follow participants for a defined time period, an accurate exposure measure can be more easily obtained and is less likely to suffer from recall bias. However, the cost of a prospective study can be substantial because a large cohort and long-term follow-up may be necessary to observe sufficient cases of the disease. A *retrospective cohort design* is often used to study conditions that occur less frequently (e.g., schizophrenia) or when there is a lengthy lag between exposure and the event/outcome (e.g., early childhood exposure and dementia). Although the disorder can be determined at the current time point, making follow-up unnecessary, the accuracy of retrospective exposure measures can be problematic for reasons such as recall bias or missing information on exposure. Psychiatric disorders often have slow onset, over years and decades, which lead many psychiatric epidemiologists to favor the life course paradigm in epidemiology.

Case-Control Studies

Case-control studies are more efficient in terms of cost and time as compared with cohort studies, especially prospective cohort studies. The case-control design is frequently used in psychiatric epidemiologic research because many psychiatric disorders are uncommon in the population. A major challenge in implementing this study design is to select controls who come from the same source population as cases. Controls can come from a variety of sources, such as hospitals or the community. Often in psychiatric epidemiologic research, studies that use cases from a psychiatric clinic use controls from a primary care clinic. However, regardless of whether the source of controls is the clinic or community, it can be difficult to know whether controls come from the same source population as cases. Selection bias is an important threat to validity in case-control studies and can obscure the relationship between exposure and disease by introducing extraneous factors that influence this relationship.

There are ways in which selection bias can be reduced. One way is to use the same eligibility criteria for both groups and to treat both groups similarly. For example, cases should not be probed more than controls for exposure information. Another way is to adopt a nested case-control study strategy using an already existing cohort study. This way, cases and controls come from the same source population. However, a parent cohort study is not always available for a nested case-control study. A third example is to use more than one control group. Making comparisons with control groups that may have different sources of bias can aid in the interpretation of results and/or evaluate the extent of bias. If the results are similar for both groups, the investigator is more confident that the results are not affected by selection bias. If, however, they are different, the reasoning for the difference, potentially selection bias, should be explored. Recall bias is another important source of bias that can frequently occur when conducting case-control studies in psychiatric epidemiology. This is because many putative exposures have to be assessed by recall of the individual, or a relative of the individual, with the disorder. It is possible that such persons search their memories and work harder to "explain" a possible cause of their condition than do comparable controls who have no need of an explanation.

Cross-Sectional Studies

A study with a cross-sectional design collects outcome and exposure data at the same time point. Cross-sectional studies are relatively inexpensive and may often use a survey approach in which large numbers of people fill out questionnaires or answer simple questions regarding their mental health and other factors. A study with a cross-sectional design provides an opportunity to have a relatively large number of cases to better detect statistical differences. A major drawback to this approach is that the temporal relationship between exposure and outcome cannot be confidently determined because both measures are collected at the same time point. Another limitation is that survey approaches that depend on self-report measures of psychiatric disorders may be subject to many biases. For instance, persons with low income or low educational levels may be more likely to have an undiagnosed psychiatric condition, resulting in inaccurate reporting of psychiatric disorders, and thus introduce a confounding factor that may obscure the relationship between psychiatric disorders and the factors of interest to the researcher.

Randomized Controlled Trials (RCTs)

An RCT is an experimental study design in which study participants are randomly assigned to either a treatment or comparison group. An RCT is often used to evaluate treatment or intervention effect rather than to study disease etiology; however, its results can complement etiology studies and lend further credence to a hypothesis. Although the RCT is often considered the gold standard study design, there are drawbacks. One of these is that the eligibility criteria to participate in an RCT may be stringent, allowing only a select group of people into the study. For example, this may occur in treatment studies that exclude sicker people for potential safety reasons. Limiting study participation to select individuals does not, in itself, lead to invalid results but it can make them nongeneralizable to other populations. Similarly, results from an RCT study may not be generalizable to real-life situations because an RCT operates under ideal conditions, which will be unlikely to be repeated outside the trial, so that a treatment or intervention shown to be efficacious under ideal conditions may not prove to be effective when used in a community setting.

Multilevel Studies

Results from many psychiatric epidemiologic research studies have indicated that most psychiatric conditions have complicated causal pathways with risk factors from multiple levels. For environmental determinants, this multilevel hierarchy may include factors measured on many levels, including cellular, neurological, physiological, psychological, family, neighborhood, county, state, and national. Depression is a disorder that has risk factors at multiple levels. Risk factors at the individual level include genes, gender, age, neurotic temperament, and life event stressors; at the family level, risk factors include family history and family cohesion; while at the neighborhood level, risk factors may include community disorganization and economic deprivation. Evaluation of the contribution of factors at different levels requires use of multilevel modeling.

Multisite and Multinational Studies

There is a growing demand for studies involving multiple sites and multiple nations in psychiatric epidemiology because of the low incidence of many psychiatric conditions and the large sample size needed to provide sufficient statistical power for multilevel and multifactorial analyses. An example of a cross-national study is the World Mental Health 2000 Surveys recently conducted in 29 countries. Findings from this study provided important insights into the similarities and differences between psychiatric disorders across the world. However, studies implemented at this scale pose a number of challenges due to differences in administrative systems, cultures, and languages across nations.

Environmental and Genetic Factors

Most psychiatric disorders involve an inherited predisposition interacting with environmental exposures in complex ways. Autism, schizophrenia, and bipolar disorder have the strongest degrees of inheritance, while depression, anxiety disorders, and conduct disorders have moderate or small inherited factors. Environmental factors include prenatal complications, psychological experiences, physical conditions, neighborhood risk, life event stresses, social supports, or toxicant exposure. Although gene-environment interaction is a highly popular topic in psychiatric epidemiological research, only a few substantive findings have been reported. These include the interaction of perinatal risk factors with genetic risk exemplifying the neurodevelopmental model of schizophrenia, and genetic risk and the serotonin transport gene interacting with life stressors of one sort or another, exemplifying the diathesis-stress model for depression. This research is still in its infancy. No evident candidate genes have been identified for most psychiatric conditions, and measures for environmental exposures are still problematic. Furthermore, it can be challenging to define and distinguish whether an effect is genetically or environmentally based. For example, psychosocial events seem to have an influence on the prevalence of depression in a population. Although a psychosocial event, at face value, is an environmental risk factor, the means by which life events have an impact on depression may not be independent from genetically determined vulnerability. In fact, it may be that individuals "choose" and "create" their own environmental exposures due to their inherited genotype.

Age, Gender, and Culture

Age

Mental health should be studied with a life course approach that incorporates elements such as genetic

risk, parental psychopathology, prenatal exposure or complications, socioeconomic status, and environmental and contextual factors. These elements consist of a causal pathway over the life span, with critical periods for disease susceptibility in some instances. These elements may also interact with early exposure and later risk/outcome on top of an individual's genetic predisposition. This can be seen in some childhood conditions and adult-onset disorders where subtle deviances are observed in early brain development, although the full adverse consequences are not manifest until later developmental stages. An example is autism spectrum disorders (ASD). Some young children with ASD do not show "full blown" symptoms until 3 or 4 years of age, but abnormalities in brain development and behaviors may have been present from infancy. Likewise, research has uncovered evidence that childhood motor, language, cognitive, emotional, and behavioral problems are precursors to adult-onset schizophrenia.

Gender

Epidemiologic research in the field of psychiatry has shown that gender is a crucial determinant of some mental disorders. Women have a higher prevalence of mood disorders, anxiety disorders, somatoform disorders, and nonaffective psychosis, while men have higher rates of substance use disorders and antisocial personality disorder. Evidence also suggests that women have higher prevalence of three or more comorbid psychiatric disorders than men. Gender differences in mental disorders are not only observed in adults, they are also seen in some childhood diagnoses. The male-to-female ratio is, approximately, 1.5:1 in mental retardation, 3:1 to 10:1 in ADHD (attention deficit/hyperactivity disorder), and 4:1 in autism spectrum disorders. Reasons suggested for the disparity between males and females include gender differences in exposure to risk factors, symptom reporting, symptom expression and severity, natural history of a disease, service utilization, comorbidity and disability, socioeconomic control and position, and gender stereotypes. Inclusion of gender-related perspectives into psychiatric epidemiological research has important implications for clinical practice and policy making.

Culture

Much effort has been put into developing universal diagnostic systems that set criteria for each psychiatric disorder; however, many culture-specific syndromes and conditions remain and are not classified in either *DSM* or ICD. There are more than two dozen culture-specific syndromes acknowledged and listed in *DSM-IV*, and these syndromes remain closely allied with culture and resist universal classification. For example, *locura* is a severe form of chronic psychosis seen in Latinos; *boufee delirante* is a brief delusional syndrome observed in West Africa and Haiti; and *Latah* is a startle-match-obey syndrome found in Southeast Asia. These culture-specific conditions presumably reflect basic human physiologic processes, constrained or precipitated by cultural contexts, and investigations on the effect of culture will improve our understanding of how environmental factors modify disease symptomatology.

Future Directions and Prospects

Through the application of epidemiologic methods, studies of psychiatric disorders have greatly contributed to informed mental health policies and improved prevention and treatment efforts. With advances in methodology and technology, the field of psychiatric epidemiology will continue to have a great impact on improving the quality of life for many people. One important aspect of more advanced research in future psychiatric epidemiology lies in the replacement of simple case-control comparison of groups with and without specified environmental exposures with more complex designs in which multilevel environmental exposures and genotypic variants may be ascertained and interactions between genetic and environmental risk factors may be investigated. A major challenge faced by today's psychiatric epidemiologists is to keep pace with advances in techniques available for measuring environmental risk factors, especially those that occur as early as before birth (e.g., preconception, prenatal). These challenges are amplified because new causal models need to address variables from various levels and perspectives in order to understand the complex nature of most forms of psychiatric disorders. Another challenge is that gene-environment interaction research is struggling to find well-established candidate genes for most psychiatric conditions, and there is a lack of firm conceptual frameworks for environmental exposures. Although knowledge on both genetic and environmental causal factors is still fragmentary, some progress has been made for a few psychiatric disorders. Most of the challenges described above

can be overcome only if studies are not narrowed by disciplinary orientation. Epidemiology is essentially a collective science; therefore, a multidisciplinary research team that integrates expertise is essential in psychiatric epidemiologic research.

Successfully fighting mental diseases will require research in many disciplines, intervention on every level, and involvement across nations. Research needs to incorporate environmental factors, socioeconomic conditions, gender, culture, individual behaviors, biologic components, and molecular genetics. Interventions need to include social changes, individual behavior changes, and effective treatments and should have political support. Because psychiatric disorders impose a heavy burden on population health in many countries, the World Health Organization (WHO) has initiated collaborative work to lay the foundation to extend the use of instruments across diverse cultures and in different languages. One of the clearest messages is that the application of question and answer techniques can no longer be limited to North America and Europe; the perspective must be global, and the techniques need to be adapted for studies in all parts of the world.

—Li-Ching Lee, Rebecca Harrington, and William W. Eaton

Note: Dr. William Eaton's effort on this work was supported by NIMH grant MH 47447.

See also Alzheimer's Disease; Anxiety Disorders; Autism; Bipolar Disorder; Gene-Environment Interaction; Life Course Approach; Multilevel Modeling; Post-Traumatic Stress Disorder; Schizophrenia; Study Design

Further Readings

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Cannon, T. D., Mednick, S. A., & Parnas, J. (1989). Genetic and perinatal determinants of structural brain deficits in schizophrenia. *Archives of General Psychiatry*, *46*(10), 883–889.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, *301*(5631), 386–389.

Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J. P., et al. (for the WHO, World Mental Health Survey Consortium). (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *Journal of the American Medical Association*, *291*(21), 2581–2590.

Jarvis, E. (1971). *Insanity and idiocy in Massachusetts: Report of the commission on lunacy, 1855.* Cambridge, MA: Harvard University Press.

Kendler, K. S., Thornton, L. M., & Gardner, C. O. (2001). Genetic risk, number of previous depressive episodes, and stressful life events in predicting onset of major depression. *American Journal of Psychiatry*, *158*(4), 582–586.

Kessler, R. C., Haro, J. M., Heeringa, S. G., Pennell, B. E., & Ustun, T. B. (2006). The World Health Organization World Mental Health Survey initiative. *Epidemiologia e Psichiatria Sociale*, *15*(3), 161–166.

Susser, E., Schwartz, S., Morabia, A., & Bromet, E. J. (2006). *Psychiatric epidemiology: Searching for the causes of mental disorders.* New York: Oxford University Press.

Tsuang, M. T., & Tohen, M. (Eds.). (2002). *Textbook in psychiatric epidemiology* (2nd ed.). New York: Wiley.

World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva, Switzerland: Author.

World Health Organization. (2001). *World health reports: Mental health new understanding, new hope*. Geneva, Switzerland: Author.

Publication Bias

Publication bias can result from the selective publication of manuscripts based on the direction and magnitude of results, multiple publication of results, and selective reporting of results within a published study. In particular, research with statistically significant positive results is more likely to be submitted for publication, to be published, and to be published more quickly than research with negative or nonsignificant results. Consequently, published studies on a particular topic might not be representative of all valid studies conducted on the topic, leading to distortion of the scientific record.

Publication bias tends to be greater in clinical research than in public health research, and in observational studies as opposed to randomized studies. Nevertheless, it has been demonstrated across all these types of research. One area where a variety of publication biases have been documented is pharmaceutical industry studies of new drug applications.

The primary sources of publication bias are commonly assumed to be editorial decision making, together with authors' reluctance to submit research with null or negative results—sometimes referred to as the file drawer problem. While research has supported the latter explanation, studies of publication bias in editorial decision making have yielded mixed findings. Less well-recognized sources of publication bias include multiple publication of results and within-study selective reporting among multiple outcomes, exposures, subgroup analyses, and other multiplicities. Although these types of publication bias have until recently received little attention, they are likely to cause even greater bias in the literature than does selective publication.

Publication bias presents a serious threat to the validity of systematic reviews and meta-analyses. Undetected publication bias not only can lead to misleading conclusions but at the same time can also give the impression of unfounded precision of results. A screening method for selective-publication bias in meta-analysis involves correlating observed effect sizes with study design features that are potential risk factors for publication bias, such as sample size. A funnel plot provides an informal graphical method where effect sizes are plotted against sample sizes, while the null hypothesis of no publication bias can be tested using rank correlation approaches such as Kendall's tau or Spearman's rho. Detecting within-study selective reporting presents a greater challenge, unless access is available to a study's original protocol and complete results of all analyses performed.

Several strategies exist for reducing or adjusting for publication bias. *Sampling methods* involve tracking down unpublished manuscripts, sometimes referred to as the grey literature, as well as broader systemic solutions such as requiring prospective registration of clinical trials. *Analytic methods* include the file drawer adjustment strategy, where the number of zero-effect studies needed to eliminate significant findings in a meta-analysis is estimated. More complex analytic approaches involving weighted distribution theory are also available. All analytic methods involve important assumptions, which in many situations can be questionable. Perhaps most important, consumers of meta-analyses and systematic reviews are cautioned to be constructively skeptical in interpreting results.

—Norman A. Constantine

See also Evidence-Based Medicine; Meta-Analysis; Peer Review Process

Further Readings

Chan, A. W., Hrobjartsson, A., Haahr, M. T., Gotzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *Journal of the American Medical Association*, *291*, 2457–2465.

Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *Lancet*, *337*, 867–872.

Melander, H., Ahlqvist-Rastad, J., Meijer, G., & Beermann, B. (2003). Evidence b(i)ased medicine: Selective reporting from studies sponsored by pharmaceutical industry. *British Medical Journal*, *326*, 1171–1173.

Phillips, C. V. (2004). Publication bias in situ. *BMC Medical Research Methodology*, *4*(20).

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, *53*, 207–216.

Public Health, History of

Human concern with health dates back to the earliest writings and civilizations. Excavations of Mohenjo Daro and Harappa in the Indian subcontinent reveal bathrooms and drainage systems more than 4,000 years old. Hygiene is a vital component of many religions, governing the cultural and culinary traditions of numerous societies. In addition, people throughout history have considered illness, especially plagues, a judgment or punishment from god(s).

The great writers, philosophers, and physicians of ancient Greece tell us of the beginnings of public health. Hippocrates in "Airs, Waters and Places" distinguished between endemic and epidemic diseases and the factors affecting them, including climate, soil, water, mode of life, and nutrition. He also discussed the link between health and the environment, suggesting conditions to avoid and others to seek out in the interest of health.

The Romans continued the medical inquest of the Greeks and added to the administration of public health. Not only did they construct impressive baths and sewer systems, but aqueducts and water supply systems were carefully erected and monitored as well. There was a government position dedicated to the maintenance of the water supply and the supervision of public use and another office for oversight of the drainage system. Eventually, individuals also would be charged with assessment of the food supply.

Despite the impressive gains in sanitation made by both the Greeks and the Romans, many of the poorest citizens of both societies, including slaves, lived in deplorable conditions. City water and sewage systems typically did not extend into the poorer neighborhoods, leading to filthy living conditions and higher disease incidence. There is little in Greek literature relating to occupational health, but the Romans recognized increased disease frequency among slaves as well as workers in specific trades, such as miners, blacksmiths, and sulfur workers.

Greek physicians often treated the destitute and wealthy alike, and eventually Rome followed suit, implementing a publicly funded medical service in the second century BCE. In Greece, doctors had offices, but it seems hospitals—and certainly charity hospitals—originated in fourth century Rome.

The Middle Ages, bookended by the Plague of Justinian in 543 and the Black Death in 1348, was a time of frequent and profound epidemics. During these years, citizens lived rural lives within the narrow confines of a city, which often lacked reliable municipal water supplies and sewage disposal systems. Overcrowding was common, and livestock in addition to humans contributed to the waste. These cities generally lacked paved roads and drainage systems. Hygiene and health remained connected to religion, though medical care reverted to more pagan traditions and spiritual cleansings. Around 1200, cities throughout Europe began drafting laws to improve public health. Slaughterhouses were established and the possession of animals was regulated. Dumping waste into rivers was prohibited, roads were paved, and covered drainage systems were constructed. The market, the center of city life, became the impetus for food regulation as it was recognized as a common site for the origination of disease outbreaks. Isolation of patients with communicable diseases developed early in the Middle Ages in response to leprosy. In addition, laws required citizens to report others exhibiting symptoms of disease. The isolation premise extended to plague, and from it the segregation of ships coming into Venice's ports led to the term *quarantine* from the Italian *quarantenaria*, meaning 40 days.

The Renaissance brought great strides in scientific discovery, laying the groundwork for advances in public health. During the Renaissance, two theories on the origin of epidemics prevailed. The first, taken from Hippocrates, held that environmental factors dictated the potential for outbreaks and an individual's susceptibility determined whether he would fall ill. The opposing theory of contagion, championed by Giolamo Fracastoro (1478–1533), gave us our present understanding of infection. Fracastoro believed that microscopic agents were responsible for disease and these agents could be transmitted by direct contact, through the air, or by intermediate fomites (inanimate objects that transmit contagious disease). He and his contemporaries, however, did not imagine these infectious agents to be alive. It was not until Anton von Leeuwenhock (1632–1723) observed the first microscopic organisms that people believed this to be possible. And despite earlier conjecture by some leading scientists, the germ theory of disease did not truly take hold until the late 19th century.

As mercantilism and the conquest for wealth and power swept Europe from the 16th to the 18th centuries, public health was encapsulated in the national interest. The necessity to quantify people and their health became clear. William Petty (1623–1687) coined the term *political arithmetic* and advocated the collection of data on income, education, and health conditions. Gottfried Achenwall introduced the term *statistics* in 1749 to replace "political arithmetic." It was John Graunt (1620–1674), however, who published the first statistical analyses of a population's health, noting associations of a variety of demographic variables with disease. He recognized the imperfections in his data but worked to determine the reliability and errors in it. Graunt produced the first calculations of life expectancy. It was during this time that people began to recognize the need for state-supported programs to prevent early death and societal loss, yet it was not until the 19th century that government was able to enact a true national health plan, and even then, most public health measures continued to be administered locally.

As France led the world into the Enlightenment, public health began in earnest. A humanitarian spirit and the desire for equality led to a social understanding of health. Infant mortality was high on the list of concerns and disparities. The public health movement involved concerned citizens lobbying their government to regulate alcohol and to provide for the safe conditions and fair treatment of all infants and children, whether illegitimate, poor, or disabled. Simultaneously, health education became popular, in line with the Enlightenment tenets of universal education and information dissemination. Despite earlier interest in the relationship of environment, social factors, and disease,

health surveys were first employed during this era. Occupational health received additional attention as well. John Howard (1726–1790) exposed the deplorable conditions in English prisons, rousing public sentiment that led to improved conditions. Mental illness, which carried a severe stigma and was generally treated by confining the affected individual, began to be viewed as a public health problem, especially after physicians demonstrated that kind treatment and a stable, nurturing environment produced better results in the insane than restraints and physical punishment.

Variolation (deliberate infection with smallpox), a common practice originating in China and spreading through the East over the centuries, became popular in the West in the 1700s. Although somewhat effective, the practice could induce severe forms of disease and contributed to epidemics. In 1798, Edward Jenner (1749–1823) used naturally acquired cowpox to inoculate others against smallpox. Within 3 years, more than 100,000 people had been vaccinated in England alone. As early as 1800, publications heralded the impending eradication of smallpox, an event that would be officially achieved in 1980.

As the Industrial Revolution spread, first in England then throughout Europe and eventually in the United States, the health of workers quickly deteriorated and calls for improved public health measures followed. The industrialization process widened gaps in income, causing the number of poor supported by local governments to increase beyond capacity. In 1834, Edwin Chadwick (1800–1890) led the development of England's Poor Law Amendment Act, which withdrew government support from the able-bodied poor in an effort to encourage self-sufficiency. The only assistance offered was placement in workhouses. The administration of this system occurred at the national level, with a hierarchy of regional and local boards below. This market system ideology mobilized the workforce, leading to a significant social change. Factories appeared and the population moved toward industrial centers, creating crowded urban areas and work conditions ripe for the spread of disease. Little, if any, city planning occurred as builders rushed to provide enough housing for the influx of workers. Meanwhile, the wealthy, who could afford transportation into the city, moved to suburban or rural areas vacated by the masses. Sanitation systems and public parks were not planned in most cities. Few toilets were available to city-dwellers, and there was no infrastructure for garbage or sewage removal. In 1833, the passage of the Factory Act dealt with working conditions, as well as the poor living conditions of those workers it sought to protect. Throughout the 1830s and 1840s, legislation regulating mines, factories, and child labor were passed in England and Europe.

Disease outbreaks certainly were associated with the poorest, dirtiest parts of cities, but quickly began to affect all social classes. Chadwick understood the poverty-disease cycle and sought statistics to quantify the relationship. Surveys on sanitary conditions resulted in the *Report . . . on and Inquiry into the Sanitary Condition of the Laboring Population of Great Britain* in 1842. The *Report* became a standard for epidemiologic investigation and community health action, and it formed the basis for sanitary reform. Chadwick clearly linked disease and environment and called for city engineers rather than physicians to wage the war on disease outbreaks. The General Board of Health, created by the Public Health Act of 1848, was an attempt at organized government responsibility for the health of its citizens. Though disbanded after a few years, the Board laid the groundwork for public health as we know it. In the United States, Lemuel Shattuck (1793–1859) produced his own *Report on the Sanitary Condition of Massachusetts* in 1850, calling for the establishment of state and local boards of health, increased attention to vital statistics collection, improved health education, and other regulations not standard in his day but now considered part of the basic public health services.

The explosion of vital statistics and survey data collection prompted the publication of several volumes during the mid-1850s. Few, however, employed the same methods, citing the inapplicability of mathematics to health. Adolphe Quetelet (1796–1874) began the work necessary to remedy the perceived incompatibility with his compendium of practical applications of mathematics.

During a cholera outbreak in London in 1848, John Snow (1813–1858), often deemed the Father of Epidemiology, identified a particular water pump as the likely source of the epidemic. Again in 1854, he mapped the reported cholera deaths and associated the clusters with a water supply company that drew its supply downstream of London on the Thames River. Snow hypothesized that cholera transmission was possible through water. In addition, he is generally credited with ending the 1848 outbreak by breaking the handle off the Broad Street Pump, although some historians believed that the epidemic had

already begun to recede by this point. It would be several decades, however, before his hypothesis was proven correct.

In 1866, the New York Metropolitan Health Bill created the Metropolitan Board of Health, reorganized 4 years later into what is today the New York City Health Department. This Board was the foundation for the U.S. public health system. In 1869, Massachusetts used Shattuck's recommendations to create the first effective state health department. Around the same time, efforts to create a National Board of Health failed. In 1878, the authority for port quarantine was bestowed on the Surgeon General of Marine Hospital Services. Eventually, this led to the creation of the U.S. Public Health Service (USPHS).

During the 19th century, two theories relating to communicable disease prevailed. The first was the miasma theory, which held that disease was due to a particular state of the air or environment. The second theory was that a specific contagion was responsible for each disease. In fact, many people believed that some combination of the two was the real explanation: that some contagious agent, whether disease-specific or not, produced disease in combination with social or environmental factors. By the end of the century, the germ theory of disease had been firmly established by Robert Koch, Louis Pasteur, and many others. From 1880 to 1898, the causative agents for a multitude of diseases, from malaria to tuberculosis, plague to typhoid, were identified. Antiseptics became popular in medical care, decreasing morbidity and mortality. Active and passive immunity were established late in the 19th century, and the development of vaccines proceeded nearly as rapidly as the discovery of pathogenic organisms. The U.S. Marine Hospital established one of the first bacteriologic laboratories in the world in 1887. Although the United States was not the site of most scientific discovery in the era, it was the leader in public health application of new knowledge.

Armed with increasingly more effective weapons against disease, public health's mission throughout most of the 20th century continued to be preventing and controlling communicable disease. Public health remained largely a local enterprise until the social change following the Depression, when people needed, and thus allowed, government intervention and subsidy. Throughout the 1900s public health achievements such as water fluoridation, mass immunizations, motor vehicle safety, occupational safety, food supply safety and fortification, improved maternal and child health, family planning, antismoking campaigns, prevention of heart disease and stroke, and of course, control of infectious diseases have led to substantially reduced morbidity and mortality. Public health has been credited with a 25-year increase in life span over the course of the 20th century. The establishment of agencies such as the Centers for Disease Control and Prevention in 1946 (born out of the Office of Malaria Control as the Communicable Disease Center—part of the USPHS) and the World Health Organization in 1948 (the United Nations' dedicated health agency) have allowed for the advancement of public health by establishing centralized agencies to which people can turn for information and assistance.

The definition of public health was also largely established during the 20th century by individuals such as C. E. A. Winslow and through groundbreaking works such as the series of reports by the Institute of Medicine (IOM) dedicated to the field. IOM's 1988 report *The Future of Public Health* clearly defined public health as "assuring conditions in which people can be healthy." It also delineated steps needed to improve a fractured public health infrastructure, and unequivocally determined the three core functions of public health: assessment, policy development, and assurance. In 2002, *Who Will Keep the Public Healthy* established requirements for the training of the public health workforce, and *The Future of the Public's Health in the 21st Century* translated the 1988 recommendations into practice while embracing the Healthy People 2010 initiative of "healthy people in healthy communities."

Public health continues to evolve, although some of the ancient concerns remain. At the dawn of the 21st century, when the industrialized world seemed to be close to conquering major infectious diseases, HIV/AIDS emerged as a deadly contagious disease with no known cure. It makes the infected person vulnerable to diseases not generally of concern to the uninfected population, such as pneumocystis pneumonia (pneumocystis carinii or PCP). Antibiotic resistance has also made the apparent victory over common infections less certain. High rates of nosocomial infections are disconcerting, and as in the field, the prevalence of antibiotic resistant organisms in hospitals is growing. Medical care and insurance in the United States continue to cost more than most people can afford, and as the population ages, the

federal government will face increasing fiscal demands. Bioterrorism and natural disasters have required planning for mass immunization, prophylaxis, evacuation, and treatment.

The future of public health will be busy indeed, but the number of trained workers is increasing to meet the need. New schools of public health are being established, and undergraduates at some institutions can take coursework and complete degrees in public health. Strong partnerships between government, private, and nonprofit agencies exist, and public health on an international scale is becoming more integrated. Laboratory science continues to make discoveries that allow public health improvements in disease treatment and prevention, and as we continue to build our understanding of the human genome, the public health implications will continue to expand. Public health comes from a varied and tumultuous past, and its future lies in continuing to form interdisciplinary alliances while focusing on its core disciplines and functions to assure a continuity of practice in a world of change.

—Erin L. DeFries

See also Emerging Infections; Epidemiology, History of; Graunt, John; Snow, John

Further Readings

Rosen, G. (1993). *A history of public health.* Baltimore: Johns Hopkins University Press.

Turnock, B. J. (2004). *Public health: What it is and how it works* (3rd ed.). Sudbury, MA: Jones & Bartlett.

Web Sites

Centers for Disease Control, information and history: http://www.cdc.gov/od/oc/media/timeline.htm and http://www.cdc.gov/od/oc/media/tengpha.htm.

Institute of Medicine reports: http://www.iom.edu.

World Health Organization, information and history: http://www.who.int/about/en.

Public Health Agency of Canada

The Public Health Agency of Canada (PHAC) was created in September 2004, after the SARS outbreak of 2003, due to concerns about the capacity of the Canadian public health system to anticipate and respond effectively to public health threats. The agency's role is to help build an effective public health system in Canada, which allows Canadians to achieve better health and well-being, while protecting them from threats to their health security. The PHAC has three main areas of responsibility: preventing and responding to outbreaks of infectious disease and other public health emergencies, preventing chronic disease and injury, and promoting good health.

The agency is directed by the Chief Public Health Officer, who reports to the Minister of Health. The Chief Public Health Officer fills a dual role of overseeing the daily operations of the PHAC and advising the Minister on public health matters. The agency is headquartered in Winnipeg and also has an office in Ottawa, as well as regional offices across Canada. Related organizations in the Canadian government's Health Portfolio include Health Canada, the Canadian Institutes of Health Research, and the Hazardous Materials Information Review Commission.

Due to the inherent difficulties in eliciting the requisite level of collaboration among federal, territorial, provincial, and local governments, the agency was specifically designed to encourage collaboration between these entities. All relevant stakeholders were included in the development of a national public health strategy, to serve as a framework for the agency's efforts. A new Pan-Canadian Public Health Network was established in 2005 to formalize communication links between public health experts and officials from all jurisdictions and to facilitate a nationwide approach to public health policy, planning, and implementation.

The PHAC has a mandate to lead federal efforts and mobilize action throughout Canada to prevent disease and injury, and to promote and protect national and international public health, through the following activities:

1. Anticipate, prepare for, respond to, and recover from threats to public health.
2. Carry out surveillance; monitor, research, investigate, and report on diseases, injuries, other preventable health risks and their determinants, and the general state of public health in Canada and internationally.
3. Use the best available evidence and tools to advise and support public health stakeholders nationally and internationally in their work to enhance the health of their communities.
4. Provide public health information, advice, and leadership.

5. Build and sustain a public health network with stakeholders.

The Public Health Agency of Canada publishes annual performance reports, and annual reports on plans and priorities; these are available on the Web site.

Organizational Structure

The PHAC has four main branches; each of the branches includes several agencies and centers. The branches are organized as follows.

Infectious Disease and Emergency Preparedness (IDEP) Branch

1. The Centre for Infectious Disease Prevention and Control is responsible for decreasing the transmission of infectious diseases and improving the health status of those infected via programs in surveillance and risk assessment. Program areas include the following: foodborne, zoonotic, and environmentally acquired infections; immunization; respiratory infections; community acquired infections (hepatitis C, sexually transmitted infections, and tuberculosis); blood safety surveillance and health care acquired infections; HIV/AIDS.

2. The Centre for Emergency Preparedness and Response (CEPR) is Canada's central coordinating point for public health security. Its responsibilities include developing and maintaining national emergency response plans for the Agency, monitoring outbreaks and global disease events, assessing public health risks during emergencies, laboratory safety and security, quarantine issues and travel health advisories, and bioterrorism and emergency health services.

3. The National Microbiology Laboratory (NML) consists of four programs, supported by a Division of Core Services, which includes DNA sequencing, Animal Resources, and a Central Laboratory for Decontamination and Wash-up Services. The four programs are as follows:

- Bacteriology and Enterics, focusing on bacterial diseases such as tuberculosis and meningitis, food- and waterborne pathogens, and infections affecting the nervous system
- Host Genetics and Prion Disease, dealing with transmissible spongiform encephalopathies
- Viral Diagnostics, for a range of viral diseases
- Zoonotic Diseases and Special Pathogens, dealing with viral, bacterial, and rickettsial diseases transmitted to humans from other species

4. The Laboratory for Foodborne Zoonoses (LFZ) provides scientific evidence and advice on human illnesses that arise from the interface between humans, animals, and the environment, with emphasis on intestinal-disease-causing agents.

5. The Pandemic Preparedness Secretariat (PPS) was established in March 2006 to coordinate and facilitate pandemic preparedness and response activities nationwide and internationally, such as those related to avian and pandemic influenza.

Health Promotion and Chronic Disease Prevention (HPCDP) Branch

1. The Centre for Chronic Disease Prevention and Control (CCDPC). The activities of CCDPC focus on facilitating the development of prevention, screening, and early detection programs for chronic diseases; providing project funding to community and support groups; developing national strategies for the management and control of chronic diseases; maintaining an integrated surveillance system to assist in developing chronic disease policy; providing a stimulus for international links in the area of chronic disease prevention and control.

2. The Centre for Health Promotion (CHP) is responsible for implementing policies and programs that enhance the conditions within which healthy development occurs. Programs include healthy child development, active living, families, aging, lifestyles, public information, education, and issues related to rural health.

3. The Transfer Payment Services and Accountability Division promotes excellence in management practices via initiatives on performance measurement and evaluation and the management of grants and contributions. It manages the Population Health Fund and provides administrative services for several grants and funding programs.

Public Health Practice and Regional Operations (PHPRO) Branch

The Office of Public Health Practice (OPHP) was created to support and improve the public health

infrastructure necessary for effective public health practice. Its priority issues are information and knowledge systems, the public health workforce, and public health law and information policy. Regional offices throughout Canada carry out the Agency's mandate by engaging in program delivery, research, policy analysis, community capacity building, and public and professional education.

Strategic Policy, Communications, and Corporate Services (SPCCS) Branch

The Strategic Policy Directorate gathers and synthesizes key policy information, cultivates partnerships, and provides evidence-based policy advice. The branch also includes the Communications Directorate, the Finance and Administration Directorate, the Human Resources Directorate, the Information Management and Information Technology Directorate, and the Audit Services Division.

—Judith Marie Bezy

See also Bioterrorism; Centers for Disease Control and Prevention; Governmental Role in Public Health; Public Health Surveillance; U.S. Public Health Service

Further Readings

Public Health Agency of Canada. (2005). *Performance report for the period ending March 31, 2005* (BT31-4/98-2005). Ottawa, Ontario, Canada: Canadian Government Publishing.

Public Health Agency of Canada. (2006, September 26). *Report on plans and priorities 2006–2007.* Retrieved July 24, 2007, from http://www.tbs-sct.gc.ca/rpp/0607/phac-aspc/phac-aspc_e.asp.

Wilson, K. (2004). A Canadian agency for public health: Could it work? *Canadian Medical Association Journal*, *170*(2), 222–223.

Web Sites

Public Health Agency of Canada: http://www.phac-aspc.gc.ca.

Public Health Nursing

Public health nursing is a specialty within nursing whose primary focus is on the health care of communities and populations rather than individuals, families, or groups. The goal of public health nursing is to prevent disease and preserve, promote, and protect health for the community, a focus that allies it closely with the concerns of epidemiology and public health in general. This entry reviews the history of public health nursing and describes the varied settings and functions of work by public health nurses.

The primary emphasis in public health nursing is on populations that live in the community, as opposed to individuals or families. In public health nursing, problems are defined (assessments/diagnoses) and solutions (interventions) implemented for or within a defined population or subpopulation as opposed to diagnoses, interventions, and treatments carried out at the individual level.

In contrast, community-based nursing is setting specific, whereby care is provided for "sick" individuals and families where they live, work, and attend school. Emphasis is on acute and chronic care and the provision of comprehensive, coordinated, and continuous care. Nurses who work in the community may be generalists or specialists in adult, geriatric, pediatric, maternal-child, or psychiatric mental health nursing. Community health nursing practice focuses on the health of individuals, families, and groups and how their health status affects the community.

History

Public health nursing evolved in the United States in the late 19th and early 20th centuries. Lillian Wald emerged as a leader in the field because of her pioneering work in public health nursing in New York City. Growing up in Rochester, New York, Wald worked as a nurse tending to immigrant families on the Lower East Side of New York. Her experiences provided evidence that many injustices existed in society with differences in health care for those individuals who were able to pay for care versus those who were poor and unable to pay. Wald could not tolerate situations where poor people had no access to health care. With the support of others, Wald moved to the Lower East Side of New York and began campaigning for health-promoting social policies to improve environmental and social conditions affecting health. As author of *The House on Henry Street*, Wald described her work as a public health nurse as well as the development of payment by life insurance companies for nursing services. The Henry Street Settlement, established in New York City, is an example of

a settlement house or neighborhood center that serves as a center for health care and social welfare programs. At Henry Street Settlement, nurses took care of the sick in their homes and tended to the overall population of low-income people in the community. Wald believed that the beginning efforts at Henry Street Settlement needed to be associated with an official health agency. The establishment of rural health nursing services through the American Red Cross was led by Wald, and it addressed public health issues such as tuberculosis, pneumonia, and typhoid fever in areas outside large cities.

Public health nursing was recognized as a legitimate specialty within public health early in the development of the profession and remains an important part of public health practice today. In 1872, the American Public Health Association (APHA) was established to facilitate interdisciplinary efforts and promote the practice of public hygiene. In 1923, the Public Health Nursing Section was formed within APHA to provide a national forum for discussion of strategies for public health nurses within the context of the larger public health organization. By 1981, APHA affirmed the importance of public health nursing and defined it as a specialty that brings together knowledge from the public health sciences and nursing to improve the health of the community.

The Work of Public Health Nurses

A variety of settings and a diversity of perspectives are available to nurses interested in working in public health. Nurses employed at local, state, and federal agencies integrate community involvement and knowledge about populations with a clinical understanding of the health and illness experiences of individuals and families in the population. Nurses work in partnership with other public health staff that include physicians, nutritionists, health educators, epidemiologists, and outreach workers.

The work of public health nursing includes population-based assessment, policy development, and assurance processes that are systematic and comprehensive. Regardless of setting, the role of the public health nurse focuses on the prevention of illness, injury, or disability, and on the promotion and maintenance of the health of populations. Examples of what public health nurses can accomplish include providing preventive services to high-risk populations; establishing programs and services to meet special needs; recommending clinical care and other services to individuals and their families in clinics, homes, and the community; providing referrals through community links; participating in community provider coalitions and meetings to educate others and identify service center for community populations; and providing clinical surveillance and identification of communicable disease.

—James A. Fain

See also American Public Health Association; Community Health; Health Care Delivery; Health Disparities; Public Health, History of

Further Readings

American Public Health Association. (1981). *The definition and role of public health nursing in the delivery of health care: A statement of the public health nursing section*. Washington, DC: Author.

Backer, B. A. (1993). Lillian Wald: Connecting caring with action. *Nursing & Health Care*, *14*, 122–128.

Stanhope, M., & Lancaster, J. (2006). *Foundations of nursing in the community: Community-oriented practice*. St. Louis, MO: Mosby Elsevier.

Public Health Surveillance

Effective public health practice relies on current, relevant information on which to base actions. The information base that serves this core function is called public health surveillance—it is often called information for action. This entry describes the development of public health surveillance as well as information about some surveillance systems themselves.

History

Although the use of information in health decision making can be found as early as Hippocrates, the modern origins of public health surveillance are usually dated to the late 18th century by which time there were organized health authorities and an accepted classification system of diseases. William Farr's analysis of death certificates in England and Wales in the mid-19th century is recognized as one of the first functional surveillance systems. In the United States, as elsewhere, there was growing application of surveillance tools to infectious disease. In colonial times,

Rhode Island required innkeepers to report what we now know to be infectious diseases, and shortly thereafter reporting of cholera, yellow fever, and smallpox was codified. It was only in 1850 that national reporting of deaths was required in the United States, and it was not until 1874 that Massachusetts created a voluntary communicable disease reporting system using postcards submitted by physicians. Michigan initiated compulsory reporting in 1881. Voluntary national reporting of communicable diseases followed, but it was only after the influenza pandemic of 1918 to 1919 that all states began reporting. It is worth noting that to this day, with the exception of conditions required by international treaty (cholera, smallpox, plague, and yellow fever), reporting by states remains voluntary since the Constitution does not delegate authority over health to the federal government. Since 1961, these data have been published weekly by the Centers for Disease Control and Prevention, a federal agency, in the *Morbidity and Mortality Weekly Report* (MMWR).

Early surveillance systems monitored illness in individuals. A major conceptual shift occurred in the context of the first widespread vaccination campaign for polio (poliomyelitis). States reported small number of cases of polio among patients who received the new vaccine. Very rapidly a system for daily reporting of cases was established with rapid follow-up of cases. This led to identification of a single lot of vaccine contaminated with live virus as the source of the cases. This lot was removed from the market, and the vaccination program was able to continue. A potentially devastating disaster had been averted. Here was a clear demonstration of the value of surveillance. Alexander Langmuir led the movement to evolve the concept from monitoring disease in individuals to monitoring the health of populations and established the current definition of surveillance as the ongoing collection, analysis, and dissemination of those who need the information to take action.

In response to public health's attention to a broader range of health conditions and determinants, contemporary public health surveillance systems now embrace a broad array of health conditions, including behavioral risk factors, chronic disease, intentional and unintentional injuries, worksite diseases and injuries, birth defects, medical safety, adverse effect of drugs and vaccines, and quality of health care.

Uses of Surveillance Systems

Surveillance systems are intrinsically action oriented. Although they may capitalize on a variety of data systems, they differ from those data systems in having specific purposes that warrant their ongoing use. Surveillance systems measure health outcomes, such as cases of influenza, or important markers of health outcomes that drive programmatic action, such as obesity or tobacco smoking. This information can be used to detect outbreaks of disease or epidemics; understand the burden of disease; facilitate planning and resource allocation; understand the natural history of diseases and injuries and who in the population is affected; evaluate the impact of public health programs; and monitor changes in infectious organisms, such as antibiotic resistance.

Characteristics of Surveillance Systems

Public health resources are always limited, so surveillance systems are targeted at public health issues based on a number of factors. The public health burden of the condition or the potential public health burden should be substantial based on mortality, frequency, severity, and economic impact. In addition, the condition should be preventable or controllable.

Case definitions are specific criteria used to report and count cases. Case definitions may be straightforward, such as death from diabetes as reported on a death certificate, or based on a set of symptoms as used in syndromic surveillance, or combinations of clinical signs, symptoms, and laboratory data as is often the case for infectious diseases. Case definitions may not accurately capture all cases, but the sensitivity (the degree to which cases are identified) and specificity (the degree to which noncases are not included) need to be known so that the reliability of the data is fully understood. For example, although deaths among people with diabetes may be counted based on death certificates, it is important to realize that approximately half of the people with diabetes who die do not have diabetes recorded on their death certificates. By the same token, using symptoms of influenza will capture many cases that are actually due to other diseases. Thus, laboratory confirmation of at least a subset of cases is important to establish the type of influenza as well as the proportion of influenza-like illness actually due to influenza.

The timeliness required may vary widely. For many infectious diseases, interventions are needed rapidly because of the need to identify outbreaks and prevent further spread. For influenza, several different strategies have emerged to meet the varying surveillance needs. Thus, absence from school is a good measure of the scope of disease, since it is by far the most common reason for large-scale absenteeism during the influenza season. Sentinel physicians are used to identify early cases and to provide specimens for identification of the specific virus. A system that reports pneumonia and influenza deaths weekly from most large cities provides a measure of severity and is also used to assess when influenza exceeds expected rates. All these provide information very rapidly. For many conditions, however, daily, weekly, or even monthly surveillance would be excessive, since rates do not change rapidly. Thus, for many chronic diseases or risk factors, surveillance is conducted at less frequent intervals, often annually.

The data for surveillance may come from case reports provided directly to health departments, from surveys (e.g., the Behavioral Risk Factor Surveillance System), vital records (birth and death certificates), administrative data (e.g., medical claims data), or other sources. Data need to be collected at a sufficient level of detail to meet the public health need, but since there is usually a trade-off between feasibility and detail, efforts are made to keep the data requirements to a minimum. In addition to the condition, basic descriptors of person (age, race, sex) and place are usually collected. Additional information is often needed to provide understanding of risk characteristics to facilitate public health action. Active surveillance systems involve solicited, direct collection of data, whereas passive systems rely on spontaneous reporting. Surveillance of infectious diseases is typically achieved through passive reporting.

Tabulation and analysis of the data need to be completed in a regular and timely fashion, usually with a core set of basic descriptive analyses. More detailed analyses are conducted as needed. Descriptive analyses usually include counts of cases or events based on person (age, race, and sex of cases), location (e.g., by state or county), and time (e.g., by date or trends in number of cases over time).

The results are provided to those who need to act on the results. To facilitate this process, interpretation is often provided along with the analyses themselves. Users of surveillance information are diverse. National, state, and local health departments are the traditional users, but there are many others. For example, infectious disease practitioners rely on antimicrobial resistance patterns in their hospitals in making antibiotic choices; the Food and Drug Administration monitors safety of drugs, biologics, and devices; employers may use surveillance of the quality of managed care organizations, for example, HEDIS (Healthcare Effectiveness Data and Information Set), in making decisions about selection of health insurance plans to offer employees; and cancer researchers may use information about cancer survival trends to generate research hypotheses.

Public health surveillance serves as the nervous system for public health, providing the information to guide action. It is a dynamic process, responding to the changing needs of society and public health. Unlike most research-generated data, surveillance systems must be highly efficient and the strengths and limitations of the data must be well understood so that users can respond appropriately. Modern information systems should enhance the capacity of public health surveillance to meet future demands.

—*Steven Teutsch*

See also Behavioral Risk Factor Surveillance System; Birth Certificate; Centers for Disease Control and Prevention; Death Certificate; Governmental Role in Public Health

Further Readings

Centers for Disease Control and Prevention. (2001). Updated guidelines for evaluating surveillance systems: Recommendations from the guidelines working group. *Morbidity and Mortality Weekly Report*, *50*(RR13), 1–35.

Teutsch, S. M., & Churchill, R. E. (Eds.). (2000). *Principles and practice of public health surveillance* (2nd ed.). New York: Oxford University Press.

P Value

A *p* value expresses the probability that a given statistical result is due to chance. *p* Values are automatically produced for many statistical procedures by

analysis packages such as SAS, SPSS, and Stata, and they are commonly cited in epidemiological and medical research as evidence that results should be considered significant. For instance, a clinical trial concerning the effect of two different drugs on a medical outcome would almost certainly cite one or more *p* values as evidence of the influence, or lack thereof, of the drugs on the outcome. In the example above, perhaps we would expect that if the two drugs are no different in their effect, we would expect the outcome to be the same in each group. Therefore, we would expect the difference between the two groups to be zero, and would be interested in determining whether the actual difference we found could be attributed to chance or whether it is likely to indicate true difference between the groups.

Generally speaking, the greater the difference between the expected and observed results, the smaller the *p* value and, therefore, the less likely it is that the difference is due to chance (random variation). As the likelihood of chance explaining the difference diminishes, so does its plausibility—giving way to an alternative explanation: that the difference is not due to chance, the difference is due to the expected value being wrong.

We will illustrate the concept of *p* value using the simple example of 10 tosses of a coin. Without saying more, it is reasonable to believe the coin is fair: There is an equal chance of either heads or tails on each toss. Following this line of reasoning, it is logical to expect 5 heads from 10 tosses, so the *expected value* of heads is 5. On 10 tosses, however, a variety of outcomes are not at all remarkable. For example, 6 heads (H) and 4 tails (T); 5H and 5T; 4H and 6T do not surprise or call into question the reasonable presumption of a fair coin. Further departures from the 5H and 5T, however, credibly cause doubt, increasing doubt with increasing departure.

Consider the result 8H and 2T. Such an outcome is not expected and causes doubt about the fairness of the coin (or process). Certainly, chance could have produced the outcome, but it is unlikely. The *p* value is the measure of that chance. The probability that such a result is due to chance is derived from the binomial distribution and is .0439. This calculation assumes that the coin is fair, that is, that the probability of heads on each toss is .5. Therefore, the *p* value calculation is a conditional probability with the condition being that the expected value is true. Such an assumption is important since we are determining the probability of the observed difference (departure) from the expected value if the expected value was correct in the first place.

In statistics, we are usually concerned with the probability not of achieving a particular result but of obtaining results at least as extreme as our result. In our example, the expected result is 5H and 5T, so we consider deviations from that expectation to be more extreme as they are less likely. So 9H and 1T is even less likely than 8H and 2T (the probability of 9H and 1T, given a fair coin, is .0098), and 10H and 0T is yet more extreme (with a probability of .0010). To calculate the probability of a result at least as extreme as 8T and 2H, we add together these probabilities: $.0439 + .0098 + .0010 + .0547$.

A final comment to this illustration: The calculation of .0547 is based on a one-sided *p* value. In other words, it only considers the probability of getting 8 or more heads in 10 tosses of a fair coin and ignores the probability of getting 8 or more tails, which would be equally as extreme. If we would consider a deviation toward either more heads or more tails to be a significant result, then the *p* value should be a two-sided calculation incorporating all the outcomes consistent with observation. In our illustration, the set of outcomes would number six: 10H and 0T; 9H and 1T; 8H and 2T; 2H and 8T; 1H and 9T; 0H and 10T. The probability would be the sum of these six independent outcomes: $.0010 + .0098 + .0439 + .0439 + .0098 + .0010 = .1094$. (The fact that this is exactly two times the one-sided *p* value is because the binomial distribution is symmetrical when the probability of a success is .5, the expected probability of an H when assuming a fair coin.) Thus, there is approximately 11% chance (.1094) that 10 fair tosses of a fair coin would produce any one of the six "extreme" values: 10H and 0T; 9H and 1T; 8H and 2T; 2H and 8T; 1H and 9T; 0H and 10T.

The *p* value can be calculated for any difference between an observed and an expected value provided that the probability distribution is known. Modern statistical analysis software calculates *p* values automatically for many different types of statistical tests, and *p* values are typically reported in epidemiological literature.

p Values are excellent metrics if properly used, because they incorporate information about sample size, sampling variation, and differences from expectation into one convenient number yielding a measure of chance for observations differing from expectations.

For example, if an epidemiologist is studying whether a suspected risk factor is associated with a disease, he or she may derive an odds ratio (for a case-control study) or a relative risk or risk ratio (*RR*, for a cohort study) for the risk factor. Until sufficient evidence indicates otherwise, the epidemiologist will assume an expected value of 1.00: that there is no increased (or decreased) risk of the disease with exposure to the risk factor. After collecting data from his cohort study, however, he discovers a *RR* of 1.25, which indicates a slightly elevated risk of the disease for those exposed to the risk factor. This of course could simply be chance. Let us assume that the correct *p* value calculation based on the probability distribution of a *RR* produces a value of .04. This result is a one-sided *p* value consistent with the epidemiologist's suspicion that the risk factor was harmful (antagonistic) rather than protective. Had he suspected an effect, but with no expectation toward a protective or antagonistic effect, he may advocate the use of a two-sided *p* value.

The value .04 takes into account the sample size, the probability distribution of *RR*, and the variability of the observations recorded from the cohort study. Taken together, this evidence suggests that if there is no effect from the risk factor in truth, then there is only a 4% ($p = .04$) chance of obtaining a *RR* of 1.25 or higher for a randomly selected data set of the same size from this cohort.

In practice, this simply means that it is unlikely that the $RR = 1.25$ is due to chance. Therefore, a more plausible explanation for the departure from $RR = 1.00$ is that the risk factor is actually associated with the disease. Typically, when a *p* value is less than .05, an epidemiologist will set aside chance as the explanation in favor of a real association. The value of .05 is based on a long history of scientific precedent memorialized in the concept of Type I error and its presumptive measure, $\alpha = .05$. As a formal rule, when the *p* value is less than α, then reject the null hypothesis (usually the expected value). Otherwise, the null hypothesis is tenable.

Because the *p* value is a practical metric summarizing much of the evidence and avoids the formal language of hypothesis testing, it is a popular way to report results. Unfortunately, the *p* value is taken to be more than what it actually is. One problem is that the *p* value may be heavily influenced by sample size: It is a truism in epidemiology that if you have a large enough sample you will likely find significant results. For this reason, a distinction is often made between statistical and clinical significance—that is, between a difference that is merely improbable in a statistical sense versus one that is large enough to make a difference in a person's health. For the same reason, some journals require the presentation of confidence intervals, which give an indication of the magnitude of differences found. Second, a *p* value is purely a statistical calculation, and its usefulness is limited to reliability and has no bearing on validity.

—Mark Gerard Haug

See also Confidence Interval; Significance Testing; Type I and Type II Errors

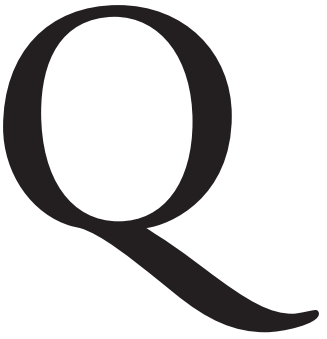
Further Readings

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 391–408). Thousand Oaks, CA: Sage.

Poole, C. (2001). Low *p*-values or narrow confidence intervals: Which are more durable? *Epidemiology*, *12*, 291–294.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: CRC Press.

Tietjen, G. L. (1986). *A topical dictionary of statistics.* London: Chapman & Hall.



Qualitative Methods in Epidemiology

Qualitative methods (QM) are used to explore a wide range of experiences in epidemiology and public health. These methods examine the depths of experience to identify why or how complex events happen, and they are particularly useful for exploring new and complicated topics. Characteristics of QM studies generally include field contact and are intended to provide a holistic perspective. If an estimate of the magnitude of a problem is needed, QM will not be useful. Generalizability from the purposive sample (which samples the topic of interest) to the general population is not a goal of QM research.

The goals of QM are usually exploratory and descriptive, with the aim of understanding and describing a phenomenon and focusing on perceptions of the "lived experience" from the perspective of the research respondent. For example, after determining whether or how well a prevention program works using quantitative methods, a researcher may turn to QM to examine how the program works and what aspects of the program the research participants and the staff believe are and are not working. A QM approach is particularly necessary in participatory research, where giving voice to vulnerable populations is often a particular concern.

QM approaches are derived from various philosophies that inform each of the steps in the research process: framing the research purpose, collecting data, analyzing, and interpreting. Data are generally collected using interviews, focus groups, existing documents, and observation. The most common methods of QM are grounded theory, phenomenology, ethnography, and case study, with the latter two using quantitative data in addition to qualitative data.

Grounded theory examines a process, usually a psychosocial process of change such as adapting to a new diagnosis of a health problem. A first sample of data is collected using "purposive sampling," which refers to sampling to cover the topic of interest. Next, an initial analysis is done, and it then informs the further collection of data, which is followed by analysis of the new data. This process is called "constant comparison." When new data yield no additional information, "data saturation" has been reached, and data collection stops. Analysis involves coding the meaning for each segment of the document, aggregating the codes into themes and then, usually, formulating a core category. The context, conditions, covariances, and consequences are explored in formulating how the core category relates all the themes together. Symbolic interaction theory informs this approach.

Phenomenology focuses on the essential core meaning of a phenomenon, with primary importance given to the social meanings people ascribe to the phenomenon. Recognizing that interpretation is an intermediate step between recognition and behavior, the researcher uses reflective description and meditation to identify the essence. Models or theories are not the goal of this approach. This approach arose from the philosophical traditions of Edmund Husserl and Martin Heidegger.

Ethnography aims to describe a group or a culture, focusing on the routines and usual lives of the people as observed in behaviors and information from records. As with the other QM, ethnography emphasizes a holistic perspective, the multiple realities of the different respondents, and the embeddedness of data in the specific context. Structure and function, symbol and ritual, and micro- and macrolevels of data are concepts that guide this approach. Analysis often involves identification of patterns of thoughts and actions, making flowcharts of major concepts, and the use of simple nonparametric statistics and scales.

Case study is an approach that first establishes bounds for a unit such as a school or a health system. A thorough description is produced of the activities in the actual setting(s) using multiple sources of information rich in social, historical, and economical context. A single case would be of interest because it belongs to a set of cases or is exceptional in some way that is of interest. Issues are selected to help focus the data collection. Patterns and conclusions are developed.

Recently, more attention has been given to the advantages of combining QM with quantitative methods, sometimes called "mixed methods" or "integrated methods." The techniques need to be planned to support each other, either sequentially or concurrently. For a new area of study, a common process is to use focus groups or interviews to identify the domains and the natural vocabulary needed in questionnaire development. QM can be used along with questionnaires to expand, explain, or reinforce the quantitative results.

To be useful, research must be credible and be perceived as appropriately rigorous, whether it is descriptive or posits a causal model. The concepts of reliability and validity are not applied in QM as they usually are in quantitative studies. Rigor of the results has been examined in terms of credibility, transferability, dependability, and confirmability. More recently, some attention has been focused on the research process to ensure rigor. Verification involves checking data, confirming the data links to theory or themes, inclusion of a clear audit trail, and the logic connection between purpose (goal), process (sampling and analysis), and product (knowledge).

Ethical issues are of primary concern in QM, which often involve close contact with research participants. It is important to plan for "leaving the field"; respondents may have considered the researcher a confidant, and they will need closure. In focus groups, the researcher cannot guarantee that the group members will maintain the conversation in confidence. Also, in a group setting, a member may disclose more than he or she intended to; the group leader needs to carefully monitor to prevent this occurrence.

Computer programs can assist with the organization of data files and can reduce some burden for the researcher. For medium to large projects, any of the several available are worth becoming familiar with. Quantitative data, such as demographics and videos, can also be linked to the qualitative data for analysis.

—Martha Ann Carey

See also Community-Based Participatory Research; Ethics in Human Subjects Research; Interview Techniques; Measurement; Survey Research Methods

Further Readings

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitatve research.* Chicago: Aldine.

Miles, M., & Huberman, M. (1994). *Qualitative data analysis: An expanded handbook* (2nd ed.). Thousand Oaks, CA: Sage.

Morse, J., & Field, P. (1995). *Qualitative research methods for health professionals* (2nd ed.). Thousand Oaks, CA: Sage.

Quality of Life, Quantification of

The term *quality of life*, as used in health research and policy, refers primarily to the quantification of the cost of being in a less-than-perfect health state. Such quantification is motivated largely by economic analyses, which require comparison of all costs and benefits of a policy or situation, including not only mortality and resource costs but also morbidity. Without such quantification, it is impossible to assess the net benefits of policies that affect health or to make formal analyses comparing the costs and benefits of different treatments. Various methods exist for eliciting quality of life quantifications and calculating quality-adjusted life years (QALYs). However, all of them have major drawbacks, and thus while it is always possible to

generate the needed numbers, it is difficult to defend them as accurate in most contexts.

Quality of life is a concept that permeates modern and ancient philosophy. It invokes two interrelated but fundamentally different meanings, creating some practical confusion. One sense of the term captures concepts related to happiness, satisfaction, and freedom from pain, while another refers to the worthiness of people's lives. In modern health science, researchers are interested in the former concerns, and particularly with assigning cardinal values to different states of well-being. However, subtle influences of the latter interpretation sometimes confuse measures and understanding.

In the post-Enlightenment world, where everyone's life is considered to have worth and everyone's well-being is considered a legitimate concern, there is near universal agreement that improving people's quality of life is a worthwhile endeavor. Making people happy (comfortable, functional), and not just long-lived, is seen as a goal of health care, health policy, and health research.

The motivation for quantifying quality of life in health science is largely driven by the limits of the ordinal concept—increased longevity or health or happiness is better—in economic analysis (which is the study and assessment of trade-offs) of health care. While it is straightforward to quantify the number of lives (or life years) saved for a given expenditure on mortality-reducing interventions, it is more difficult to quantify the trade-off when the benefits are a reduction in morbidity rather than mortality. Comparison of the value of interventions that reduce morbidity to those that reduce mortality, and analysis of interventions that substantially affect both, is impossible without a common metric. That metric is also useful for descriptive epidemiology and other research, apart from economic analysis.

A naive and inappropriate measure of the cost of morbidity is the loss of productivity (often measured in terms of lost wages). This implicitly invokes the ''worthiness'' sense of *quality*, equating the value of someone's life to what they produce. Lost wages are often the basis of payouts from insurance contracts or policies designed to mimic insurance (e.g., the settlements paid to survivors of victims of the 9/11 attacks). Such measures have a legitimate economic basis (roughly speaking, rational insurance contracts should replace what can be replaced with money, but not pay for those things which cannot be replaced). But they should not be mistaken for the appropriate value to incorporate into decision making. Individual productivity is a reasonable approximation for the value of someone's life in some sociopolitical systems (e.g., primitive cultures where survival of the community is in question, or modern highly communitarian systems such as fascism or communism), but in modern Western traditions, there is general agreement that someone suffering poor health causes much greater cost than merely the wages lost, as does dying prematurely.

QALYs and Related Measures

In response to these challenges, the concept of *quality-adjusted life years* (QALYs) based on quality of life (QoL) scores was developed. QoL is a score on a scale where perfect health is assigned the value of 1.0 and being dead is assigned the value of 0.0 for a particular (actual or hypothetical) state of health. In this scale, 1 is the maximum possible value, though it is possible to suffer misery that would be valued at less than 0. QALYs are derived from this score by multiplying by the number of person-years spent in a particular state of health. The premise is that loss of a given number of QALYs is equally bad, whether it is, for example, premature death by 10 years of one person who had been in perfect health, or 50 people who suffer a QoL drop from 1.0 to 0.8 for 1 year.

With QALYs, it is possible to compare the cost-effectiveness of an intervention that prevents mortality with one that increases quality of life without changing life expectancy. It is also possible to add or subtract health benefits and mortality effects, such as subtracting the morbidity cost of an unpleasant intervention (e.g., chemotherapy or giving up smoking) from the mortality benefit. The cost-effectiveness measure, dollars (or another common metric) per life-years saved, can be easily reconfigured as the (presumptively equivalent) cost in dollars per QALY gained.

Technically, the QALY measure has these computational properties only if QoL is properly measured as *von Neumann-Morgenstern utility* (typically referred to as simply *utility*). The fundamental property is that if there is a good state with utility h, and a bad state with utility d, then by definition an intermediate state has utility x if the individual is indifferent between the intermediate state, or facing a gamble in which he has a $(h-x)/(h-d)$ probability of being in the good state and a $(x-d)/(h-d)$ chance of being

in the bad state. That definition is fairly nonproblematic when applied to monetary wealth: Someone might be just indifferent between getting $50,000 for sure versus a .5 chance of getting $150,000 with a .5 chance of nothing, in which case the utility of having an additional $50,000 is exactly halfway between his current utility and the utility he would have with an additional $150,000. But applying this concept to health states (thus, gaining the mathematical conveniences it allows) requires asking the question, "given a choice between living in a particular poor-health state, or a gamble, with an $x\%$ chance of dropping dead and a $(1-x)\%$ chance of being perfectly healthy what value of x would make you just indifferent between the prospects?"

Direct Measures of Individual Quality of Life

In the above question, the answer given for x is, by definition, that person's QoL score for the health state in question. (To put this in terms of the previous probabilities, recall that death is typically assigned the score $d=0$, while perfect health is assigned $h=1$.) This can be directly assessed by asking some variation of the above question, an approach known as the *standard gamble method.* Not surprisingly, people find it difficult to answer such questions.

In response to that difficulty, a more popular measurement method is the rating scale, in which subjects are simply asked to state a QoL number (or, equivalently, to mark a point on a line segment, a method called a *visual analog scale*) representing a particular health state. While study subjects generally find this much easier, it is not clear what the measure really represents, since there are no real units (such as the probability in the gamble). Another alternative is the *time trade-off* method, wherein subjects state how much time lived in perfect health would be equivalent to a longer given time lived in a diminished health state. This has the advantage of being clearly defined but is complicated by subjects introducing an unknown level of discounting and valuing living at different ages differently.

For methods that do not produce true utility scores, there is no mathematical justification for treating the resulting QALYs as equivalent to healthy life years, though this is often done in practice. This practice continues despite empirically demonstrated problems, such as respondents spreading their scores across a substantial part of the range (0, 1) for even minor morbidities. When such results are converted to QALYs, they imply that many morbidities count as a much greater fraction of dying than we would generally believe is plausible.

Standard gamble surveys typically produce QoL scores that are very close to 1.0, even for major morbidities such as blindness or loss of a limb. Taken at face value, these measures suggest that the quantified quality of life loss from diseases is usually very small compared with the loss from dying. (The exceptions tend to be depression and other major mental illness, morbidities that severely impair communication and social interaction, and unrelenting physical pain.) These results are rather more plausible than many of those generated by non-gamble-based measures. However, eliciting responses, already difficult for lifelong ailments, becomes prohibitively difficult when measuring a temporary condition. It is nearly impossible to make sense of a question such as, "If you had the choice of a gamble where you might be dead for 5 days and then alive again, or having the flu for 5 days," let alone the answer. But the alternative of using a normal standard gamble question and assuming that temporary and permanent conditions produce the same welfare loss per unit time (e.g., that having the flu for 5 days is 1/1,000th as bad as having it for 5,000 days) cannot be justified.

Further complicating the problem is that the answer to a QoL question depends a great deal on who is asked. While there should be genuine heterogeneity among people (an athlete is more bothered by tendonitis, a guitarist is likely to value loss of his left little finger more than most people, etc.), there are problematic systematic patterns. Most people assess the quality of life loss from losing their eyesight or hearing as very large, perhaps a QoL of 0.5, while blind or deaf people often assess their QoL as better than 0.9. There is no clear basis for choosing one of these over the other when we want to calculate, say, the QALYs lost due to carotene-deficiency-induced blindness in India, which would be a useful number for prioritizing interventions.

Whatever question is used to elicit QoL, a fundamental problem remains: Results cannot be validated other than by asking another hypothetical question. It is very rare for someone to be offered an actual gamble (e.g., the choice about health improving but possibly fatal surgery), and no equivalent exists for other elicitation methods. Despite the very different

answers that result from different questions and populations, if QoL quantification is to be used, it is necessary to just assume one measure is right without being able to justify that assumption.

Partial Solutions to the Measurement Challenge

Asking hypothetical questions about potentially fatal rolls of the dice, or even rating scales or time trade-offs, is far beyond the experience and comfort zone of most people, so it is seldom clear what the answers mean. People are more comfortable thinking about hypothetical *willingness to pay* (WTP): How much money they would be willing to part with to get a particular benefit. While such contemplations usually involve simple consumer goods, it is not a major cognitive leap to substitute relief from a minor morbidity. Economists have substantial experience measuring hypothetical WTP and have refined the study methodology, though unless there is actually a real market, such surveys still suffer from the lack of external validation of the hypothetical question.

Most people can assess how much they would pay to get rid of their tendonitis or avoid a case of the flu, while they cannot easily answer questions that measure how many QALYs the condition costs them. Validation studies that compare QoL survey results with WTP results find substantial divergence (which cannot show that either result is correct, but suggests that the QoL responses do not correspond to a measure that appears more robust). WTP measures of morbidity can be entered into a cost-benefit comparison to calculate net benefits. It is not, however, possible to use WTP in cost-effectiveness calculations based on dollars per QALY (an approach that offers no real advantage over cost-benefit comparisons, but is more popular in the medical literature).

For major morbidities, the WTP is less useful. It is difficult to interpret an asserted willingness to pay $5 million to avoid becoming paralyzed when it comes from someone whose lifetime earnings will only be $2 million. But capping the cost of morbidity at the subject's available wealth would clearly be inappropriate. Thus, WTP may be a more promising measure for minor morbidities and benefits that are akin to those from consumer goods, while QoL remains necessary for quantifying major morbidity.

An alternative solution to the measurement problem is to focus on the major motivation for calculating QALYs—their use in economic policy analyses—and simply declare what QoLs will be used to make economic decisions without actually assuming they are genuine utility scores. This is effectively what is done when an expert group is asked to assess QoLs on behalf of others (often clinicians on behalf of their patients). While those doing such studies perhaps do not realize it, they are declaring answers to questions such as, "How many cases of major cerebral palsy should be considered to be as bad an outcome as one neonatal death?" If the subjects are aware of the role they are playing, this is a valid way of generating QALY measures, not because it is based on individual utility, but because it is based on a studied decision about society's trade-offs.

Health in the Context of Overall Well-Being

Capturing the effects of different QoL scores in policy analysis is a major improvement over measuring only mortality. Inclusion of mental health effects on QoL is particularly important (some studies have found that total QALY loss from mental health problems exceeds that from any physical disease). But diagnosable conditions capture only a small part of psychological well-being.

Analyses that consider only mortality, morbidity, and expenditures—which is to say, most studies of costs and benefits of health care or public health interventions, especially those coming from the "health promotion" tradition—ignore many factors that affect quality of life. This is suggestive again of defining quality as someone's worthiness or contribution to society, since medically defined morbidity is typically related to losses of productivity, but less so general happiness. The bias toward treating quality of life as reflecting only morbidity or productivity is reflected in the condemnation of some unhealthy exposures (e.g., recreational drugs and junk food) that make people happy but may decrease their productivity, and the relative silence on other exposures (e.g., driving and overwork) that are just as dangerous to physical health but may increase productivity.

Negative responses to public health interventions (e.g., "If I do everything they say, I will live to 100, but I won't want to") reflect the popular view that increasing longevity, or diagnosable-morbidity-adjusted QoL, may conflict with other preferences. Failure to consider all sources of individual well-being leads to

such cruel absurdities as limiting access to pain-relief medication or forcing psychiatric patients to give up cigarettes (which they relish, and often get substantial symptomatic relief from) so that they can live a longer but less enjoyable or more troubled life.

Quantified quality of life presents the advantages and pitfalls for analysis that any quantification does. It allows a formal measure, for whatever purpose, of an otherwise vague construct, in particular allowing it to be used in economic analysis. However, quantified values tend to overshadow everything that is left unquantified, including uncertainty about the quantification and other values, and thus inaccurate or incomplete quantification can lead to incorrect analyses that are imbued with a false impression of precision.

—Carl V. Phillips

See also Disability Epidemiology; Economic Evaluation; Ethics in Health Care; EuroQoL EQ-5D Questionnaire; Quality of Well-Being Scale (QWB)

Further Readings

Addington-Hall, J., & Kalra, L. (2002). Measuring quality of life: Who should measure quality of life? *British Medical Journal*, *32*, 1417–1420.

Bala, M., Wood, L., Zarkin, G., Norton, E., Gafni, A., & O'Brien, B. (1998). Valuing outcomes in health care: A comparison of willingness to pay and quality-adjusted life-years. *Journal of Clinical Epidemiology*, *51*, 667–676.

Drummond, M., O'Brien, B., Stoddart, G., & Torrence, G. (1997). *Methods for the economic evaluation of health care programmes*. Oxford, UK: Oxford University Press.

Phillips, C., & Thompson, G. (2001). What is a QALY? *Hayward Medical Communications*, *1*, 6.

Quality of Well-Being Scale

The Quality of Well-Being Scale (QWB) is a generic, preference-based measure of health-related quality of life (HRQOL). It has been extensively validated, and its psychometric properties are well established. A self-administered version of the QWB (QWB-SA) has been developed and validated in response to limitations of the QWB, and it is easier to administer in most research and clinical assessment protocols. The questionnaire assesses the presence or absence of symptoms and functioning on specific days prior to administration. The measure produces a single score that ranges from 0 (death) to 1.0 (optimal HRQOL). The score can be integrated with time and mortality to calculate quality-adjusted life years (QALYs) and conduct cost-effectiveness analysis. To place each case on the continuum between death and optimum functioning, the measure uses mean preference weights from a community sample. QWB scores are most commonly used to describe the HRQOL of larger groups or samples and to inform epidemiological research and public health policy. They may be of less value for assessing individual health status.

Health-Related Quality of Life

HRQOL describes a comprehensive picture of health and overall well-being. HRQOL measures differ from one another along several dimensions, including generic versus disease-specific measures and psychometrically based versus preference-based measures. The QWB is a generic measure in that it was designed to be used with any adult population and any health condition, including healthy individuals. The QWB is a preference based measure and was not developed to assess statistically independent domains of HRQOL. It is preference based, meaning that it is scored on the basis of mean health consumer preferences or utilities for the health states. These preferences or utilities are the ratings of observable health states using a continuum anchored by death and optimum health.

Quality of Well-Being

The QWB was developed in the 1970s using theory from the general health policy model. This model includes several components, including mortality (death) and morbidity (HRQOL). The theory proposes that symptoms and disabilities are important for two reasons: First, illness may cause life expectancy to be shortened and, second, illness may make life less desirable at times prior to death. In assessing the impact of a health intervention, the model requires data on both a possible change in mortality as well as a change in HRQOL. In addition to mortality and morbidity, the general health policy model incorporates preference for observed health states (utility) and duration of stay in health states. Preferences or utility for health states are typically measured using economic principles that ask individuals to prioritize or place values on a wide variety of health states involving both symptoms and functioning. The health

preferences or utilities are placed on a preference continuum for the desirability of various health states, giving a "quality" rating on an interval scale ranging from 0 = *death* to 1.0 = *completely well*.

Calculation

Once a value is obtained that describes the level of morbidity or wellness in a sample using a measure such as the QWB, the score can be multiplied by the amount of time at that level of wellness to calculate QALYs. A QALY is defined as the equivalent of a completely well year of life, or a year of life with optimal functioning and no health problems or symptoms. Consider, for example, a person who has a set of symptoms and is in a state of functioning that is rated by community peers as 0.5 on a 0.0 to 1.0 scale. If the person remains in that state for 1 year, he or she would have lost the equivalent of 1/2 of 1 year of life. Thus, a person limited in activities who requires a cane or walker to get around the community would be hypothetically rated at 0.50. If he or she remained in that state for an entire year, the individual would lose the equivalent of one-half year of life. However, a person who has the flu may also be rated as 0.50. In this case, the illness might only last 3 days and the total loss in QALYs might be $3/365 \times 0.50$, which is equal to 0.004 QALYs. This may not appear as significant an outcome as noted for the disabled person. But suppose that 5,000 people in a community get the flu. The well years lost would then be $5{,}000 \times 0.004$, which is equal to 20 years of perfect health in one person. An important feature of the system is that it is completely generic. It can be used to compare small health consequences that affect a large number of people with large health consequences that affect a small number of people. The quality-adjusted life expectancy is the current life expectancy adjusted for diminished quality of life associated with dysfunctional states and the duration of stay in each state.

The calculation of QALYs is required for conducting cost-utility analysis, which is simply a cost-effectiveness analysis that uses QALYs as its unit measure of health benefit. The QWB was the first assessment instrument developed for the primary purpose of calculating QALYs and conducting cost-effectiveness analysis. Prior to the existence of generic, preference-based measures, many different outcomes were used to represent the effectiveness side of cost-effectiveness analyses. Generic, preference-based measures and QALYs have become the recommended standard for cost-effectiveness analyses because they provide a common metric for comparing results across studies and populations.

In the original QWB, respondents report whether or not each of 27 groups of symptoms were experienced on each of the 6 days prior to the assessment. Functioning was assessed by questions about the presence of functional limitations over the previous 6 days, within three separate domains (mobility, physical activity, and social activity). Unlike measures that ask about general time frames such as "the past 4 weeks" or "the previous month," the QWB asks whether specific symptoms or functional limitations did or did not occur on a given day. Each symptom complex and functional limitation is weighted using preferences obtained from the ratings of 856 people randomly sampled from the general population. The four domain scores (three functioning, one symptom) are subtracted from 1.0 to create a total score that provides an expression of well-being that ranges from 0 = *death* to 1.0 = *asymptomatic optimal functioning*. References on the validation of the instrument are available from the University of California San Diego Health Outcomes Assessment Program (UCSD-HOAP). The questionnaire must be administered by a trained interviewer because it employs a somewhat complex branching system of questions and probes. The original questionnaire takes an average of about 15 min to complete. The authors believe that the administration time and complexity of the original measure, which require a trained interviewer, have resulted in its underutilization.

Self-Administered Quality of Well-Being

In 1996, a self-administered version of the questionnaire was developed to address some of the limitations of the original version. The QWB-SA improves on the original version in a number of ways. First, the administration of the questionnaire no longer requires a trained interviewer and can be completed in less than 10 min. Second, the assessment of symptoms follows a clinically useful review of systems model rather than clustering symptoms based on preference weights. Third, a wider variety of symptoms are included in the QWB-SA, making it more comprehensive and improving the assessment of mental health.

Preference weights for the QWB-SA were obtained with a new sample, and studies were conducted and

published comparing the new and old versions. The QWB-SA and QWB were highly correlated and the test-retest reliability is high. The measure is not designed to be internally consistent because the factors it measures (symptoms and functioning) are interdependent. QWB-SA scores tend to be slightly lower than QWB scores, primarily because mental health symptoms are assessed in much greater detail and are more likely to contribute to decreased scores.

The format for the QWB-SA includes five sections. The first part assesses the presence/absence of 19 chronic symptoms or problems (e.g., blindness, speech problems). The question format does not assess each of the previous 3 days (as in the rest of the questionnaire) with the expectation that these chronic conditions do not vary much over the 3-day assessment period. These chronic symptoms are followed by 25 acute (or more transient) physical symptoms (e.g., headache, coughing, pain), and 14 mental health symptoms and behaviors (e.g., sadness, anxiety, irritation). The remaining sections of the QWB-SA are similar to the QWB and include assessment of mobility (including use of transportation), physical activity (e.g., walking and bending over), and social activity, including completion of role expectations (e.g., work, school, or home).

The period assessed by the QWB-SA is shorter than in the QWB. The QWB asked patients about symptoms and function "over the past 6 days" prior to the day of administration, whereas the QWB-SA questions refer to the 3 days prior to the day of administration. This change was designed to reduce respondents' recall bias without decreasing the instrument's ability to assess over a period of time and resulted in a more rapid administration. The impact on the overall quality of life score of using only the last 3 days was examined by dropping information from Days 4, 5, and 6 and recalculating QWB scores based only on the past 3 days. No significant differences in scores were found.

When compared with other generic, preference-based measures of HRQOL, the QWB-SA remains longer and slightly more time-consuming because its assessment of symptoms and functioning is more comprehensive. However, the more detailed assessment of symptoms and functioning may result in greater sensitivity to change in some populations. The QWB-SA asks about the presence or absence of specific complaints on specific days to reduce the influence of memory, or severity ratings such as pain intensity, that require personal interpretation. In addition, the distribution of QWB-SA scores in most studies is close to normal, suggesting that ceiling or floor effects are less common than with other HRQOL measures.

Both the QWB and QWB-SA are available free of charge to users from nonprofit organizations. A small fee is charged to for-profit users. Information on copyright agreements and user manuals are available at www.medicine.ucsd.edu/fpm/hoap.

—Erik J. Groessl and Robert M. Kaplan

See also Confounding; Ecological Fallacy

Further Readings

Groessl, E. J., & Cronan, T. A. (2000). A cost analysis of self-management programs for people with chronic illness. *American Journal of Community Psychology*, *28*(4), 455–480.

Groessl, E. J., Kaplan, R. M., Barrett-Connor, E., & Ganiats, T. G. (2004). Body mass index and quality of well-being in a community of older adults. *American Journal of Preventive Medicine*, *26*(2), 126–129.

Groessl, E. J., Kaplan, R. M., & Cronan, T. A. (2003). Quality of well-being in older people with osteoarthritis. *Arthritis and Rheumatism*, *49*(1), 23–28.

Kaplan, R. M., & Anderson, J. (1996). The general health policy model: An integrated approach. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed., pp. 309–322). Philadelphia: Lippincott-Raven.

Kaplan, R. M., Sieber, W. J., & Ganiats, T. G. (1997). The quality of well-being scale: Comparison of the interviewer-administered version with a self-administered questionnaire. *Psychology & Health*, *12*, 783–791.

Web Sites

University of California San Diego Health Outcomes Assessment Program Web site: http://www.medicine.ucsd.edu/fpm/hoap/index.html.

Quantitative Methods in Epidemiology

See Categorical Data, Analysis of

Quarantine and Isolation

Quarantine and isolation are two different ways to limit the spread of certain infectious diseases by

reducing contact between individuals at risk of spreading infectious disease and the rest of the population. The types of diseases for which quarantine and isolation are useful public health measures are those that involve direct transmission of infection by close contact (e.g., aerosol or droplet transmission). There must also be detectable symptoms that allow individuals who have been infected to be distinguished from those who have not. The aims of quarantine and isolation can vary and include stopping local spread of disease, global eradication of a disease, or simply slowing down the progress of an epidemic to gain time in which to vaccinate or administer drugs.

Fundamentally, the difference between quarantine and isolation depends on whether the individual has a confirmed infected/infectious status (depending on the specific disease, determining infection may be easier than determining infectiousness, which is the ability to transmit the disease). If an individual's infected/infectious status is confirmed, they are *isolated*, which means removed to an environment designed to prevent them from spreading the infection to other individuals. These environments can range from an individual's own home to a highly secure medical facility. Individuals can also receive treatment while in isolation, with the health workers taking precautions against compromising the isolation (e.g., physical barriers or vaccination). Isolated individuals remain in isolation until they are no longer considered to be at risk of spreading infection (established, e.g., by a serology test or a clinical assessment). Occasionally, isolation is enforced by law, as was done with tuberculosis (TB) in New York City during the 1990s. In this case, the aims included the prevention of rapid emergence of drug-resistant forms of TB, caused in part by patients not completing their drug treatment. At around the same time (1986–1993), Cuban residents who were HIV-positive were isolated in sanitariums, though this controversial policy evolved so that patients had a choice of how and where to be treated.

Individuals may be *quarantined* if they are considered at risk of having been exposed to an infectious disease (from an infected individual or another source) but do not display symptoms of the disease. The "at risk" assessment can be made by the individual who may have been exposed or by a third party (e.g., doctor, public health official), and it can be based on contact tracing (determining an individual's recent close contacts by interview or questionnaire), or on the individual's having been to a certain region where the infectious disease is endemic or epidemic. The quarantine conditions can be as strict as those for isolation, but they are often based more on clinical observation and can be as simple as self-reporting and staying at home. If an individual develops symptoms, then he or she meets the criteria for isolation. The length of time that an individual is quarantined for is related to the specific infectious disease; in fact, the origin of the word *quarantine* comes from the 40 days that people arriving by ship had to remain on their ships before coming to land in case they had been exposed to the plague but had not yet become symptomatic. Time required to be spent in quarantine should relate to the incubation period of a particular disease, that is, the time between infection and the onset of detectable symptoms. Mathematical modeling has shown how this quarantine period can best be set and modified based on updated information about the incubation period, which is especially useful for emerging infectious diseases where little epidemiological data is known.

Theoretical work has demonstrated that the success of quarantine and isolation in controlling infectious diseases is strongly linked to both the proportion of presymptomatic transmission and the inherent transmissibility of the etiological agent. Although isolation is probably always a desirable public health measure, quarantine is more controversial. Mass quarantine can inflict significant social, psychological, and economic costs without resulting in the detection of many infected individuals. However, quarantine can be enforced by law, and indeed, during the 2003 severe acute respiratory syndrome (SARS) epidemic, governments added this syndrome to the list of diseases for which individuals can be quarantined. Probabilistic models have been developed to determine the conditions under which quarantine is expected to be useful. Results demonstrate that the number of infections averted (per initially infected individual) through the use of quarantine is expected to be very low provided that isolation is effective, but it increases abruptly and at an accelerating rate as the effectiveness of isolation diminishes. When isolation is ineffective, the use of quarantine will be most beneficial when there is significant asymptomatic transmission and if the asymptomatic period is neither very long nor very short. In these cases, quarantine and isolation can be effectively combined to halt the spread of infection, where each on its own would be insufficient. Both

quarantine and isolation become effective when contact tracing is efficient (i.e., accurate and speedy). Quarantine and isolation can be used in conjunction with other public health measures such as vaccination and antiviral drugs, as has been recommended by the World Health Organization (WHO) when faced by an influenza pandemic.

—Andrew Park and Troy Day

See also Influenza; Outbreak Investigation; Severe Acute Respiratory Syndrome (SARS); Tuberculosis

Further Readings

Day, T. (2004). Predicting quarantine failure rates. *Emerging Infectious Diseases*, *10*, 487–488.

Day, T., Park, A., Madras, N., Gumel, A., & Wu, J. (2006). When is quarantine a useful control strategy for emerging infectious diseases? *American Journal of Epidemiology*, *163*, 479–485.

Fraser, C., Riley, S., Anderson, R. M., Ferguson, & N. M. (2004). Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 6146–6151.

James, L., Shindo, N., Cutter, J., Ma, S., & Chew, S. K. (2006). Public health measures implemented during the SARS outbreak in Singapore, 2003. *Public Health*, *120*, 20–26.

Quasi Experiments

Quasi experiments, like all experiments, manipulate treatments to discover causal effects (quasi experiments are sometimes referred to as nonrandomized experiments or observational studies). However, these experiments differ from randomized experiments in that units are not randomly assigned to conditions. Quasi experiments are often used when it is not possible to randomize ethically or feasibly. Therefore, units may be assigned to conditions using a variety of nonrandomized techniques, such as permitting units to self-select into conditions or assigning them based on need or some other criterion. Unfortunately, quasi experiments may not yield the unbiased estimates that randomized experiments yield because quasi experiments can neither reliably rule out alternative explanations for the effects nor create error terms that are orthogonal to treatment. To improve causal inferences in quasi experiments, however, researchers can use a combination of design features, practical logic, and statistical analysis. Although researchers had been using quasi-experimental designs long before 1963, it was then that Donald Campbell and Julian Stanley coined the term *quasi experiment.* The theories, practices, and assumptions about these designs were further developed over the next 40 years by Campbell and his colleagues.

Validity and Threats to Validity

In 1963, Campbell and Stanley created a validity typology, including *threats to validity*, to provide a logical and objective way to evaluate the quality of causal inferences made using quasi-experimental designs. The threats are common reasons why researchers may be incorrect about the causal inferences they draw from any cause-probing study, including randomized and quasi experiments. Originally, Campbell and Stanley described only two types of validity: internal validity and external validity. Thomas Cook and Campbell later added statistical conclusion validity and construct validity. We define the validity types shortly. Of the four types of validity, internal validity is the most crucial to the ability to make causal claims from quasi experiments. *Internal validity* concerns the validity of inferences that the relationship between two variables *A* and *B* is causal from *A* to *B*. The act of randomization helps reduce the plausibility of many threats to internal validity. Lacking randomization, quasi experiments have to pay particular attention to these threats:

- *Ambiguous temporal precedence:* the inability to determine which variable occurred first, thereby preventing the researcher from knowing which variable is the cause and which is the effect.
- *Selection:* systematic differences between unit characteristics in each condition that could affect the outcome.
- *History:* events that occur simultaneously with the treatment that could affect the outcome.
- *Maturation:* a natural development over time that could affect the outcome.
- *Regression:* when units are selected for their extreme scores, they may have less extreme scores on other measures, including later posttests, making it appear as if an effect occurred.
- *Attrition:* when units who drop out of the one condition are systematically different in their responses than those who drop out of other conditions.

- *Testing:* repeatedly exposing units to a test may affect their performance on subsequent tests, appearing as if a treatment effect occurred.
- *Instrumentation:* changes over time or conditions in the instrument used to measure responses may make it appear as if an effect occurred.
- *Additive and interactive threats to internal validity:* the impact of a threat can be compounded by, or may depend on the level of, another threat.

The other three types of validity also affect causal conclusions about the treatment and outcome, but they do not necessarily affect quasi experiments more than any other type of experiment. *Statistical conclusion validity* addresses inferences about whether and how much the presumed cause and effect covary. Examples of threats to statistical conclusion validity are low statistical power and violation of statistical assumptions. *Construct validity* addresses inferences about higher-order constructs that research operations represent. Examples of threats to construct validity include reactivity to the experimental situation (units respond as they want to be perceived rather than to the intended treatment) and treatment diffusion (the control group learns about and uses the treatment); note that in both these cases, a question is raised about whether the researchers are actually measuring or manipulating what they intended or claimed. *External validity* addresses inferences about whether a causal relationship holds over variation in persons, settings, treatment variables, and measurement variables. Examples of threats to external validity include interactions of the causal treatment with units or setting, so that the observed causal relationship might not hold in new units or settings.

Basic Types of Quasi Experiments

While there are many variations of quasi-experimental designs, basic designs include, but are not limited to, (a) *one-group posttest only designs*, in which only one group is given a treatment and is then observed for effects using one posttest observation; (b) *nonequivalent control group designs*, in which the outcomes of two or more treatment or control conditions are studied, but the experimenter does not control assignment to conditions; (c) *regression discontinuity designs*, in which the experimenter uses a cutoff score from a continuous variable to determine assignment to treatment and comparison conditions, and an effect is observed if the regression line of the assignment variable on outcome for the treatment group is discontinuous from that of the comparison group at the point of the cutoff; and (d) *interrupted time-series designs*, in which many (ideally, 100 or more) consecutive observations over time are available on an outcome, and treatment is introduced in the midst of those observations to determine its impact on the outcome as evidenced by a disruption in the time series after treatment; and (e) *single-group or single-case designs*, in which one group or unit is repeatedly observed over time (more than twice, but fewer than in a time series) while the scheduling and dose of treatment are manipulated to demonstrate that treatment affects outcome.

The causal logic of threats to validity can also be applied to two other classes of designs that are not quasi experiments because the cause is not manipulated, as it is in the previous five designs. These are (f) *case-control designs*, in which a group with an outcome of interest is compared with a group without that outcome to see how they differ retrospectively in exposure to possible causes; and (g) *correlational designs*, in which observations on possible treatments and outcomes are observed simultaneously to see if they are related. These designs often cannot ensure that the cause precedes the effect, making it more difficult to make causal inferences than in quasi experiments.

Design Features

To prevent a threat from occurring or to diagnose its presence and impact on study results, researchers can manipulate certain features within a design, thereby improving the validity of casual inferences made using quasi experiments. These design features include (a) adding observations over time before (pretests) or after (posttests) treatment to examine trends over time; (b) adding more than one treatment or comparison group to serve as a source of inference about the counterfactual (what would have occurred to the treatment group if they had not received the treatment); (c) varying the type of treatment, such as removing or varying a treatment; and (d) using nonrandomized assignment methods that the researcher can control or adjust, such as using a regression discontinuity design or matching. All quasi experiments are combinations of these design features, chosen to diagnose or minimize the plausibility of threats to validity in a particular context.

New designs are added to the basic repertoire of designs using these elements. For example, by adding pretest observations to a posttest-only nonequivalent control group design, existing pretest differences between the treatment and control groups can better be measured and accounted for, which helps reduce effects of selection. Likewise, adding a comparison group to a time-series analysis can assess threats such as history. If the outcome for the comparison group varies over time in the same pattern as the treatment outcome, history is a likely threat.

Examples

Martin Atherton examined the effectiveness of a Web-based, interactive intervention aimed to improve self-management of asthma, called MyAsthma™, on the quality of life for asthma suffers. High volume users of the intervention (those visiting the Web site 17 or more times) reported a better quality of life than low volume users over a 6-month period of time. The researcher compared quality of life for users before (pretest) and after (posttest) the intervention. Using a pretest was critical for this study, since the low volume users had higher pretest scores on all measures of quality of life than the high volume users. Without accounting for the pretest scores, it would have been more difficult to assess the effectiveness of the treatment between the high and low volume users. This study might be thought of as a nonequivalent comparison group design with pretests and posttests.

In 1999, as part of the Complying with the Minimum Drinking Age project (CMDA), law enforcement agencies from several communities across the midwestern United States began periodic enforcement checks in which minors attempted to purchase alcoholic beverages from local establishments. A total of 116 observations were collected once every 2 weeks over 4.5 years from both on-premise (i.e., restaurant and bars) and off-premise (i.e., grocery and liquor stores) sites. Alexander Wagenaar, Traci Toomey, and Darin Erickson found a 17% decrease in alcohol sales to minors immediately after the law enforcement checks among both sites, although long-term effects of these checks varied. The off-premise venues eventually returned to their previous rates of illegal sales. However, the on-premise venues were better at reducing alcohol sales to minors, demonstrating a long-term decrease of 8.2% in illegal sales. Although several of the establishments in the control group were threatened by treatment diffusion (law enforcement agencies began checks in these establishments beyond the researchers' control), those establishments in which law enforcement agencies did not check for illegal sales did not decrease their alcohol sales to minors over time. This study was an interrupted time-series quasi experiment using a nonequivalent control group.

Statistical Adjustments

While Campbell emphasized the importance of good design in quasi experiments, many other researchers sought to resolve problems in making causal inferences from quasi experiments through statistical adjustments. One such method uses *propensity scores*, the conditional probability that a unit will be in a treatment condition given a set of observed covariates. These scores can then be used to balance treatment and control units on predictor variables through matching, stratifying, covariate adjustment, or weighting. Another method, *selection bias modeling*, attempts to remove hidden bias that occurs when unobserved covariates influence treatment effects by modeling the selection process. A third method uses structural equation modeling to study causal relationships in quasi experiments by modeling latent variables to reduce bias caused by unreliable measures. While these statistical adjustments have been shown to reduce some of the bias present in quasi experiments, each of these methods has its limitations that prevents it from accounting for all the sources of biased estimates. Therefore, it is often more effective to obtain less biased estimates through good designs than elaborate statistics; and these statistics always perform better when the study was better designed at its start.

Conclusions

Quasi experiments may never rule out threats to internal validity as well as randomized experiments; however, improving the designs can reduce or control for these threats, making causal conclusions more valid for quasi experiments than they would be otherwise. The strategy is to begin with a basic design that is most appropriate for the research question and most feasible given the practical constraints on the study. Then add design features to address particular plausible threats to validity that may exist.

While certain conditions within field studies may hinder the feasibility of using more sophisticated quasi-experimental designs, it is important to recognize the limitations of designs that are used. In some cases, statistical adjustments can be used to improve treatment estimates; however, even then, causal inferences from quasi experiments should be made with caution.

—Margaret H. Clark and William R. Shadish

See also Bias; Causation and Causal Inference; Randomization; Study Design

Further Readings

Atherton, M. (2000). Outcome measures of efficacy associated with a web-enabled asthma self-management programme: Findings from a quasi-experiment. *Disease Management and Health Outcomes*, *8*(4), 233–242.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand-McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand-McNally.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153–161.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Wagenaar, A. C., Toomey, T. L., & Erickson, D. J. (2005). *Addiction*, *100*, 335–345.

Questionnaire Design

Questionnaires are one of many ways to elicit information. Researchers who design their own questionnaires need to take steps to ensure that they are validly measuring whatever it is they seek to measure. A strong, well-designed questionnaire starts with the conceptualization of the problem and ends with the visual clarity of the presentation. Poor questionnaire design not only leaves a researcher with incomplete and/or inaccurate information but also wastes the time of the individuals who complete the questionnaire.

Conceptualizing the Problem: What Questions Need to Be Asked?

Before specific questions are developed, it is imperative that a researcher identify what problem he or she is trying to understand as well as consider potential explanations for that problem. This process of conceptualization may use established theoretical frameworks typically based on prior research or, if no known framework exists to the researchers knowledge, a generative process of exploring all possible explanations should be pursued. Because this process is often difficult and time-consuming, it is sometimes omitted. However, failure to conceptualize all the possible ''whys'' that may explain a problem will result in the probable exclusion of important questions in the final questionnaire. It is often not until the conclusion of the study that these omissions become apparent, and it is then too late to remedy them.

To illustrate the process of conceptualizing a problem, consider the researcher interested in understanding smoking behavior by female adolescents. In the medical, biological, and social science literature, there are a number of possible explanations for why a young girl begins to smoke. For example, there are arguments identifying biological, familial, emotional, and social exposures influencing smoking initiation. While a particular researcher may be interested in understanding the impact of parental smoking on an adolescent girl's decision to smoke, going through the generative process of conceptualizing and examining other explanations alerts the researcher to consider including questions about age, psychological well-being, and the environment in which the girl lives. Even if they are not the focus of the study, including these factors will further refine the researcher's ability to understand how parental smoking operates as a risk factor.

Operationalizing the Measures

Once a researcher has identified the concepts to be measured, the next step is to determine the specific ways to measure them. This is referred to as operationalization. To measure a concept accurately, a researcher must ask questions whose answers will provide useful information about the concepts of interest. The questions must be phrased in such a way that the respondent understands what is being asked and can

provide a reasonable reply. To ensure clarity in questions, certain types of questions should be avoided, such as ''doubled-barreled'' questions (which ask more than one question at a time), long questions, and questions that use language that may confuse the respondent. For example, using the problem of female adolescent smoking, a question such as, ''When you are feeling overwhelmed, do you tend to want a cigarette?'' is doubled-barreled—a ''No'' could mean that the respondent does not smoke when overwhelmed or that the respondent does not get overwhelmed. Additionally, the word ''overwhelmed'' may not be understood by all adolescent respondents.

Another important consideration when designing a question is sensitivity. If the researcher asks a question that a person would feel uncomfortable answering, then the researcher risks not only missing responses but also potentially alienating the respondent so that he or she is reluctant to honestly answer any further questions. While some of this can be avoided by considering how the questionnaire is delivered (e.g., face-to-face vs. telephone), it can also be avoided by learning how others have asked these questions in the past and by considering factors such as the culture, age, and religion of the respondent. For example, to ask a question such as, ''Do you think smoking after sexual intercourse is common?'' may not bother some respondents, whereas it may be very uncomfortable for others.

It is also important to consider recall when asking questions. The ability of respondents to remember whatever is being asked may be a universal problem, that is, one that all respondents would have difficulty with. Or, it may be a problem unique to one set of respondents but not for another. For example, few individuals who smoke could probably recall what their relationship was like with their parents when they bought their first pack of cigarettes. Furthermore, and possibly more problematic, individuals who always have an easy or always have a strained relationship with their parents may have an easier guess at this question, while for those whose parental relationship are sometimes easy and sometimes strained, remembering the specific condition at a given moment in the past could be difficult. This introduces an element of bias into the subjects' responses.

Two other considerations when designing a question relate less to the wording of the question itself but are also important to consider. First, for closed-ended questions (those with delineated response options), the response options must be clear and mutually exclusive. For example, a question about marital status should include not the categories ''single'' and ''divorced'' (because they could both apply to the same person), but the categories ''single,'' ''never married'' and ''currently divorced.'' Second, often a concept cannot be fully measured using only one question. For example, to understand if an adolescent female were a smoker, it would be important to find out not only if she currently smoked but also how much and for how long.

Formatting the Questionnaire

Once the specific questions are determined, it is important that a researcher consider the format of the questionnaire carefully. A common accepted practice is to begin and end a questionnaire with easy to answer questions such as demographics (gender, age, number of siblings). On a similar note, if more sensitive questions are to be asked, these questions are usually placed later in the survey, so that the respondent is comfortable with the survey process by the time he or she reaches those questions, thus increasing the likelihood that the respondent will answer them honestly.

Each new section of questions should be clearly distinguished and, when needed, specific instructions should be provided for each section. This is particularly so if the responses to the questions are in a new format (e.g., ''for the next set of questions, please circle all that apply'') or if they require reference to a particular time period (e.g., ''for the next set of questions, think back to when you started ninth grade'').

It is a useful tool to present a series of questions to measure one concept or a similar set of ideas in a similar format (e.g., ''for the next 10 questions, please read each scenario and circle True or Untrue as it applies to you''). These series questions are comfortable for respondents to answer and allow for more questions to be asked efficiently. However, depending on the topic, the length of the questionnaire, and characteristics of the respondents, grouped questions can be problematic, as some will disregard the specifics of each question and answer all questions in a set with the same response, for instance, by always choosing the first or last category. Often, reverse-ordered questions are included in these series to encourage more careful reading of each question. For example, rather than asking, ''I think kids should

be allowed to smoke outside school grounds'' and ''I think kids should be allowed to smoke if given permission by their parent/guardian,'' this second question could be reversed to read ''Kids should not be able to get permission from a parent/guardian to smoke.''

Skip patterns provide the opportunity to ask only certain people certain questions, potentially providing improved efficiency in a questionnaire. For example, following a question that asks, ''Do you currently smoke'' respondents who answer ''Yes'' can then be referred to one set of follow-up questions while respondents who answer ''No'' could be referred to another set. While this saves time and improves cooperation by not subjecting respondents to a number of irrelevant questions, it can potentially be confusing for respondents if directions are not completely clear. Skip patterns are easily implemented in computer-assisted interviewing and are commonly used in that context.

Implementation Options

Questionnaires can be administered in a number of formats each with its own strengths and limitations. Telephone interviews may provide a sense of anonymity and make it easy to reach people who are geographically spread out, and the presence of the interviewer allows for clarification or elaboration of questions when needed. However, many households do not have telephones and this fact is related to other factors such as age, race, income, and disability status, thus introducing an element of bias into phone surveys. In addition, many phone numbers are not accessible (e.g., unlisted or disconnected), and cell phone numbers are often not included in lists of potential respondent households, introducing another element of bias as many households now have only cell phone service. When potential respondents are reached by telephone, response rate is moderate.

Mailed questionnaires can provide a strong sense of anonymity (if a return envelope only contains a return address) and can reach individuals who are geographically spread out. Mailed surveys are potentially less expensive in that less ''interviewer time'' is needed. However, mailed surveys provide no opportunity to clarify or elaborate on responses, and the response rate (return of completed questionnaires) is typically lower than that of telephone or personal interviews.

Face-to-face interviews can create a sense of comfort that may assist in improving response rate and comfort when answering questions. Ambiguities can be clarified, and elaboration of responses can be obtained. However, it can be costly to do face-to-face interviews with a large sample, particularly if they are geographically spread out.

Computers have been engaged to assist in questionnaire implementation in a variety of ways. Computers have been used to assist during telephone interviews in order to allow the researcher to work through skip patterns with ease and to directly enter responses in order to reduce data entry mistakes. Computers using voice-activated response have also been engaged to conduct the interview over the telephone, saving person-time due to unanswered calls. Computers have also been engaged as a means of questioning respondents directly through Web-based surveys and free-standing touch screen computer kiosks. Computers have also been used in face-to-face interviews to allow subjects to answer particularly sensitive questions, such as those regarding sexual behavior, by entering their responses directly into the computer rather than by responding to the interviewer.

A final implementation option that is crucial to developing a thoughtful questionnaire is pilot testing, preferably on a sample of individuals similar to those who are the target population for the survey. Pilot testing a questionnaire is the best way to determine if questions are clear and written in such a way as to elicit appropriate responses, not misleading or offensive, and to determine if the questionnaire is easy to follow. Any difficulties encountered during pilot testing should be remedied by revisions in the instrument, interviewer training, and so on, before the full survey is implemented. By looking at responses and discussing issues with a pilot sample, questionnaires can be further refined to increase their ability to ultimately measure what the researcher hopes to understand.

—Eve Waltermaurer

See also Psychometrics; Reliability; Survey Research Methods

Further Readings

McNutt, L. A., Waltermaurer, E., McCauley, J., Campbell, J., & Ford, D. (2005). Rationale for and development of

the computerized intimate partner screen for primary care. *Family Violence Prevention and Health Practice*, *3*, 1–12. Retrieved June 18, 2007, from http://endabuse.org/health/ejournal/archive/1-3/McNutt.pdf.

Mitchell, A. A., Cottler, L. B., & Shapiro, S. (1986). Effect of questionnaire design on recall of drug exposure in pregnancy. *American Journal of Epidemiology*, *123*, 670–676.

Wacholder, S., Carroll, R. J., Pee, D., & Gail, M. H. (1994). The partial questionnaire design for case-control studies. *Statistics in Medicine*, *13*, 623–634.

R

Race and Ethnicity, Measurement Issues With

Race and ethnicity are controversial variables in epidemiological studies. Most of the controversy comes from the misuse of these variables as risk factors and from issues concerning validity and consistency of data over time and territory. Substantial inconsistencies in the categorization of race and ethnicity can be found in the literature. For these reasons, some journals have written policies and published glossaries to better define these variables. However, revisions of criteria are often required due to the dynamics of social and demographic change, such as migrations, globalization, and other cultural movements that may change the perception of group identity.

Epidemiological studies may use race and ethnicity variables in several situations. In the sampling process, these variables may be used to determine whether the true diversity of the total population is being represented by the sample and to audit the randomization process. For example, National Institutes of Health (NIH) requires the assessment of these variables to ensure that the traditionally understudied minorities are sufficiently represented. However, the validity of using race and ethnicity as causal explanatory variables or risk factors is very questionable. The detection of statistical differences between racial or ethnic groups should be considered as a starting point to better understand the true underlying genetic, environmental, or socioeconomic risk factors.

Once the relevance of the use of race and ethnicity is established, the measurement of these variables needs to be planned, validated, and analyzed with caution. The main issues are the difficulty in separating the concept of race from the concept of ethnicity, the nonequivalence of data collection methods, and the mutability of use and meaning of terminology.

Race Versus Ethnicity

In the simplest terms, ethnicity can be defined as a socially constructed method of categorization of human beings, while race can be defined as a biologically constructed method. Race takes into account the physical characteristics of the population, such as skin color, that are marked by traces transmissible by descendant. In contrast, ethnicity emphasizes cultural characteristics that lead to a sense of group membership, such as language, religion, traditions, and/or territorial identity.

One of the main problems in accurately measuring race and ethnicity is the fact that these concepts are not always distinguishable. The reason for this is the lack of a clear boundary between perceptions of race and perceptions of ethnicity. For example, while all races can be found within the group Hispanic/Latino, this group is included as a racial category in some questionnaires. Nevertheless, when race and ethnicity are collected in a single question in the questionnaire, and depending on how the question is stated, inconsistency over time may be observed due to ambiguous membership. The Office of Management and Budget

(OMB) sets standards for classification of race and ethnicity in federal data. In 1997, the standard revision provided two options—collecting race/ethnicity in one combined question or in two separate questions, one for race and one for ethnicity—but stated that to allow flexibility and to ensure data quality, separate questions are preferred.

Nonequivalence of Data Collection Methods

Analyses of epidemiological studies may require the combination of data from multiple sources. It is important to make sure that the methods used to collect the information and the categories used are compatible. Special attention should be given to who provided the information (self-reported or by an observer), how the question was stated (allowing multiple answers or not), and what categories were available.

Who Provided the Information?

While self-identification of race and ethnicity is fairly common, many studies also use information on race/ethnicity provided by an observer such as a health care provider or a direct interviewer. The complexity of the concepts of race and ethnicity and the conflicts between the social perceptions of these variables and the individual's self-identity generate differences in the data collected based on who is giving the information. Self-reported race/ethnicity is considered superior by many organizations. According to the OMB standards, self-reporting or self-identification is the preferred method for collecting data on these variables. The Centers for Disease Control and Prevention, in *Use of Race and Ethnicity in Public Health Surveillance*, discontinued the use of the observer-reported method. Independently of who provided the information, the variation in individual self-perception and social perception caused by different backgrounds, beliefs, and countries of origin contributes to limiting the quality of these variables.

How Was the Question Stated?

A second issue concerns the way the question was stated; for example, did it allow for multiple answers? Ideally, each variable—race and ethnicity—should be exclusive and exhaustive, meaning that within the variable, each subject belongs to exactly one category. That is not always an easy task because of the overlap of races and the heterogeneity of some ethnic groups. Because respondents may identify themselves as multiracial, much inconsistency in answers will be generated. In practice, the researchers may find some of these respondents still choosing more than one category. Because of these possibilities, questions that allow multiple answers are preferred.

What Categories Were Available?

Ideally, the categories should reflect the respondents' self-perception. For example, the terms *Latino* and *Hispanic* are not exchangeable. The reason for this is the variation in respondent preferences for one term or another, depending, for example, on their country of origin. For this reason, the *American Medical Association Manual of Style* recommends that whenever possible, a more specific term (such as Mexican American, Latin American, etc.) be used. Another example is the definition of race for native Hawaiians. The descendants of the original native inhabitants of the State of Hawaii were considered by OBM as Asian or Pacific Islander. However, that was not their self-perception. They perceived themselves as belonging to the category American Indians and Alaska Natives. As a result, the OBM reviewed the racial categories, and Asian or Pacific Islander category was divided into two categories: Asian and Native Hawaiian or Other Pacific Islander. The standards now have five categories of races: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White.

A study by Ulrike Boehmer and colleagues compared information on race/ethnicity from clinical files of a large sample of outpatients with their respective surveyed race/ethnicity. These data sets differed in who provided the information (self-reported vs. nonself-reported), how the question was stated (multiple answers vs. single answer), and what categories were available ("American Indian" vs. "American Indian or Alaska Native"; "Asian" vs. "Asian, Native Hawaiian or Pacific Islander"; "Black" vs. "Black or African American"; "Hispanic" vs. "Spanish, Hispanic, or Latino," and the availability of "Unknown"). Results from this study showed that self-reported whites had the fewest "unknown" in their files and the fewest misclassifications. In contrast,

self-reported Asians had the most "unknown" and self-reported American Indians the most occurrences of misclassifications in their files. These results demonstrate how different methods can generate differences in the classification of race/ethnicity.

Terminology Changes Over Time and Territory

Classifications of race have changed over time. For example, the terms *mulatto*, *quadroon*, and *octoroon* were used during the 19th century to describe individuals with one half, one quarter, and one eighth of black ancestry, respectively. Later, this terminology was abandoned and the term *black* was used to refer to persons with any black ancestry. Interesting enough, currently the term *mulatto* (or *mulato* in Spanish or Portuguese) is commonly used by several countries in Latin America. This illustrates why translation of race and ethnicity descriptions from the original language used in a questionnaire can lead to additional measurement error.

Recommendations

The appropriate study design to measure race and ethnicity will depend on the research question under investigation. Independently of how terms are defined officially, the respondent may have a different perception of what category of race or ethnicity he or she belongs to. Pretests of the questionnaires/forms and audits for quality assurance are important to verify if the perception of the respondent is captured by the correct category. Different ethnic groups may have different compositions, though sometimes with one race predominating. However, given the underlying complexity of each variable, the combination of race and ethnicity in one question is not always a good idea.

—Ana W. Capuano

See also African American Health Issues; Asian American/Pacific Islander Health Issues; Health Disparities; Latino Health Issues; Race Bridging

Further Readings

Bhopal, R. (2004). Glossary of terms relating to ethnicity and race: For reflection and debate. *Journal of Epidemiology and Community Health*, *58*(6), 441–445.

Boehmer, U., Kressin, N. R., Berlowitz, D. R., Christiansen, C. L., Kazis, L. E., & Jones, J. A. (2002). Self-reported vs. administrative race/ethnicity data and study results. *American Journal of Public Health*, *92*(9), 1471–1472.

Centers for Disease Control and Prevention. (1993, June). *Use of race and ethnicity in public health surveillance summary of the CDC/ATSDR workshop.* Retrieved March 2, 2007, from http://www.cdc.gov/mmwr/PDF/rr/rr4210.pdf.

Kaplan, J. B., & Bennett, T. (2003). Use of race and ethnicity in biomedical publication. *Journal of American Medical Association*, *289*(20), 2709–2716.

Lin, S. S., & Kelsey, J. L. (2000). Use of race and ethnicity in epidemiologic research: Concepts, methodological issues, and suggestions for research. *Epidemiologic Reviews*, *22*(2), 187–202.

Office of Management and Budget. (1997). *Revisions to the standards for the classification of federal data on race and ethnicity.* Retrieved March 2, 2007, from http://www.whitehouse.gov/omb/fedreg/1997standards.html.

Winker, M. A. (2006). Race and ethnicity in medical research: Requirements meet reality. *Journal of Law Medical Ethics*, *34*(3), 520–525.

Race Bridging

Race bridging refers to making data collected using one set of race categories consistent with data collected using a different set of race categories, to permit estimation and comparison of race-specific statistics at a point in time or over time. More specifically, race bridging is a method used to make multiple-race and single-race data collection systems sufficiently comparable with permit estimation and analysis of race-specific statistics such as birth and death rates. This entry provides an overview of the origins of race bridging and race-bridging methods and focuses on race bridging to estimate single-race population counts, as this has been the primary use of race bridging to date.

Background

The need for race bridging arose when the Office of Management and Budget (OMB) issued revised standards in 1997 for the collection, tabulation, and presentation of data on race and Hispanic origin within the federal statistical system. These standards replaced the 1977 OMB standards. The revised standards increased the minimum set of race categories from

four (American Indian or Alaska Native [AIAN], Asian or Pacific Islander [AIP], Black, and White) to five (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White). In addition, the revised standards require federal data collection programs to allow respondents to select more than one race category when responding to a query on their racial identity. This means that under the revised standards, there are potentially 31 race groups (five single-race and 26 multiple-race groups), depending on whether a respondent selects one, two, three, four, or all five of the race categories. Because of the addition of the multiple-race groups, race data collected under the revised standards are not comparable with race data collected under the 1977 standards.

The question on race on the 2000 census was based on the revised OMB standards and so allowed respondents to select more than one race category. As a result, the race data on the 2000 census are not comparable with historical race data (e.g., previous censuses, administrative records, surveys, population estimates) or with data on other data systems that have not yet transitioned to the 1997 standards. As many data systems use population estimates to create rates, this left many data users unable to compute current statistics and to track changes over time. One such example is the problem faced by the National Center for Health Statistics (NCHS) in computing birth and death rates for 2000 and beyond and measuring and tracking changes in these vital events. As of 2004, most states had not revised the race question on their birth or death certificates and were still collecting race data using the 1977 race categories. Thus, the calculation of post-2000 race-specific birth and death rates (which use birth and death counts in the numerator and population estimates in the denominator) requires population estimates with the 1977 race categories.

OMB-Proposed Bridging Methods

Recognizing the need to make race data collected under the 1997 standards comparable with race data collected under the 1977 standards, the OMB proposed a number of bridging methods. The proposed methods fall into two broad categories, whole allocation methods and fractional allocation methods. Whole allocation methods assign each multiple-race respondent to only one of the possible single-race categories. Fractional allocation methods divide each multiple-race respondent into parts and assign a part to each possible single-race category. The proposed methods include the following:

- *Smallest Group.* Assigns responses with two or more racial categories to the category, other than white, with the smallest single-race count.
- *Largest Group Other Than White.* Assigns responses with two or more racial categories to the category, other than white, with the largest single-race count.
- *Largest Group.* Assigns responses with two or more racial categories to the category with the largest single-race count.
- *Plurality.* Assigns responses based on data from the National Health Interview Survey (NHIS). Since 1982, the NHIS has permitted respondents to select more than one race and has asked them to indicate with which race they identify most closely (primary race). For each multiple-race group, the proportion selecting each race category as the primary race is calculated. Plurality assigns all responses in a particular multiple-race group to the category with the highest proportion.
- *Equal Fractions.* Assigns multiple-race responses in equal fractions to each single-race category identified.
- *NHIS Fractions.* Assigns responses by fractions to each racial category identified, where the fractions equal the NHIS proportions described in the plurality method above.

Regression Bridging Method

NCHS's National Health Interview Survey (NHIS) provides a unique bridging data source as discussed above. NCHS, in collaboration with the Census Bureau, developed a regression bridging methodology that used information about NHIS multiple-race respondents to obtain single-race population estimates. Schenker and Parker (2003) demonstrated that the regression bridging approach can provide better-bridged estimates than the other proposed bridging methods.

The regression bridging methodology used NHIS data for 1997 to 2000 and involved fitting individual logistic and multilogit models for the larger multiple race groups and a composite multilogit model for the smaller multiple race groups. The models included demographic covariates such as age, sex, and Hispanic origin and county-level contextual variables such as region, urbanization level, percentage in single-race categories, and percent multiple-race population.

Each model estimated the probability that members of the multiple-race group would select each possible single-race category. The probabilities obtained from the bridging models were specific for sex, Hispanic origin, single year of age, and county of residence.

The bridging probabilities derived by NCHS from the regression models have been applied by the Census Bureau to county population estimates beginning in 2000 for 31 races to produce county population estimates for four races. The resulting bridged-race population estimates are available for public use. During the transition period, before all or most birth and death data are available in the multiple-race format, NCHS is also using the bridging probabilities to bridge multiple-race responses on birth and death certificates to single-race responses.

Variance of Bridged-Race Population Estimates

Population estimates generally are assumed to be fixed and do not contribute to the variance of rates. However, this is not true for bridged-race population estimates. Nathaniel Schenker (2003) has developed a methodology to compute variances for bridged-race population estimates.

Race Bridging and the Census Quality Survey

The Census Quality Survey (CQS) was conducted by the Census Bureau in 2001 to produce a data file that could be used to bridge between multiple- and single-race distributions. The sample consisted largely of households that reported at least one multiple-race person in Census 2000 (90% of the initial sample). The CQS respondents were asked at one point in time to ''mark one race'' and at another point in time to ''mark one or more races.''

NCHS has used the CQS to develop new bridging models that incorporate additional county-level variables from Census 2000. The outcome of this research will inform decisions concerning selection of bridging probabilities (NHIS or CQS) for future bridging of multiple-race data.

Impact of Bridging

Bridging has the greatest impact on estimates for the AIAN and API populations because a large proportion of each of these populations reports multiple races. Bridging has a small impact on estimates for the black population and negligible impact on estimates for the white population.

—Deborah D. Ingram

See also Birth Certificate; Death Certificate; Logistic Regression; Mortality Rates; National Health Interview Survey

Further Readings

Bentley, M., Mattingly, T., Hough, C., & Bennett, C. (2005). *Census Quality Survey to evaluate responses to the Census 2000 question on race: An introduction to the data*. Retrieved June 6, 2006, from http://www2.census.gov/census_2000/datasets/CQS/B.3.pdf.

Ingram, D. D., Parker, J. D., Schenker, N., Weed, J. A., Hamilton, B., Arias, E., et al. (2003). United States Census 2000 population with bridged race categories. National Center for Health Statistics [Electronic version]. *Vital Health Statistics 2*(135). Retrieved June 6, 2006, from http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm.

National Center for Health Statistics. (2005). *U.S. Census populations with bridged race categories*. Retrieved June 6, 2006, from http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm.

Office of Management and Budget. (1997). *Revisions to the standards for the classification of Federal data on race and ethnicity* (Federal Register 62FR58781–58790). Retrieved June 6, 2006, from http://www.whitehouse.gov/omb/inforeg/r&e_app-a-update.pdf.

Office of Management and Budget. (2000). *Provisional guidance on the implementation of the 1997 standards for the collection of Federal data on race and ethnicity.* Retrieved June 6, 2006, from http://www.Whitehouse.gov/omb/inforeg/r&e_guidance2000update.pdf.

Parker, J. D., Schenker, N., Ingram, D. D, Weed, J. A., Heck, K. E., & Madans, J. H. (2004). Bridging between two standards for collecting information on race and ethnicity and application to Census 2000 and vital rates. *Public Health Reports*, *119*, 192–205.

Schenker, N. (2003). Assessing variability due to race bridging: Application to Census counts and vital rates for the Year 2000. *Journal of the American Statistical Association*, *98*, 818–828.

Schenker, N., & Parker, J. D. (2003). From single-race reporting to multiple-race reporting: Using imputation methods to bridge the transition. *Statistics in Medicine*, *22*, 1571–1587.

Symons-Smith, A., & Ingram, D. D. (2006, March 30 to April 1). *Bridging the gap between the old and new race categories: Comparing population estimates prepared*

using two different sets of bridging factors. Population Association of America meetings, Los Angeles, CA.

Radiation

In the context of epidemiology, it is useful to divide radiation into two types: ionizing and nonionizing. Ionizing radiation contains sufficient energy to remove electrons from atoms or molecules, leaving positively charged particles known as ions. X rays, neutrons, alpha particles, beta particles, and gamma rays are forms of ionizing radiation. Nonionizing radiation does not contain sufficient energy to remove electrons from their atoms: types of nonionizing radiation include radiowaves and microwaves. Ionizing radiation is known to be harmful to human tissue in some dosages and can cause damage to DNA. Although some people believe that human health can be harmed by nonionizing radiation emitted by electronic products, such as the radiofrequency radiation used by cell phones, this has not been established scientifically.

Everyone is exposed to small amounts of ionizing radiation, often referred to as "background radiation," from the sun, rocks, water, soil, and so on. For this reason, it is critical to calculate the amount of exposure to radiation when evaluating whether radiation poses a threat to health, because while low levels may be apparently harmless, high levels of exposure can cause serious health effects, including skin burns, hair loss, nausea, birth defects, and death. Exposure to high levels of radiation is also associated with increased risk of certain types of cancer. Apart from accidents such as the Chernobyl nuclear power plant explosion, most radiation exposure results from occupational exposures or from medical applications such as X rays and radiopharmaceuticals.

History

Wilhelm Roentgen discovered artificial radioactivity in 1895 with his observation that emissions from a Crookes tube (a glass vacuum tube with a high-voltage electric current flowing through it) caused a paper coated with fluorescent material to glow. He put this discovery to use by taking an "X ray" of his wife's hand by placing the hand on a photographic plate and exposing it using the Crookes tube: The developed plate revealed the bones of the hand. In 1896, Henri Becquerel discovered the existence of natural radioactivity, which he demonstrated by exposing a photographic plate wrapped in black paper by laying crystals of a uranium compound on top of the paper. The exposed plate displayed emanations from the uranium that were similar to the X rays discovered by Roentgen.

Many uses were found for both natural and artificial radiation, but unfortunately, the consequences of human exposure to radiation were not immediately understood. One of the worst examples of occupational radiation poisoning involved young women who painted dials on watch faces using radioactive paint. The first dial painter to die of radium poisoning was a young woman who had been working at U.S. Radium in New Jersey for only 3 years; her death in 1922 was followed by that of a number of her coworkers. All the early deaths involved necrosis of the jawbone (the painters used their lips to maintain a fine point on the brush) and rampant infections; others died of anemia, bone cancer, or multiple myelomas. Working with artificial radiation also proved dangerous: for instance, Clarence Dally, chief assistant to Thomas Edison, repeatedly exposed his hands to X rays in the course of his experimental work. After a few years, Dally began to suffer burns and hair loss, followed by ulcers and cancerous sores, and ultimately had both arms amputated. Radiologists, who in the early years of their profession were exposed to high levels of radiation on a daily basis, suffered higher rates of cancer, infertility, and birth defects than the general public, and hand amputations were common among that occupational group.

Health Effects

The health effects of radiation are generally related to the type and amount of exposure. There are two broad categories of health effects: stochastic and nonstochastic. Stochastic health effects are associated with long-term, low-level exposure to radiation; increased exposure increases the probability of these effects, but does not influence their type or severity. Cancer is an example of a stochastic health effect: Ionizing radiation may break chemical bonds in atoms and molecules in the body and thus disrupt the control processes that regulate cell growth. Ionizing radiation can also cause changes in the DNA that may lead to genetic and teratogenic mutations.

Nonstochastic effects are caused by exposure to high levels of radiation and are more severe if the exposure is greater. The term *acute* is often used to characterize this type of exposure and the subsequent health effects. Nonstochastic effects include burns and radiation sickness; symptoms of the latter include nausea, weakness, hair loss, and diminished organ function. Cancer patients being treated with radiation typically receive high doses for a short period of time and often experience acute radiation effects.

Regulation

Most regulations regarding permissible exposure are based on the ''linear no-threshold theory'' that states that there is no totally safe amount of radiation exposure and that the danger increases directly with the dosage. This theory has been challenged, in particular by some scientists who believe that low doses of radiation may be beneficial, but is still reflected in, for instance, standards set by the U.S. Federal Government. The term *radiation dose* refers to the amount of radiation absorbed in the body and is measured in a unit called the *rem* (roentgen equivalent man).

In the United States, radiation safety policies are set by the Environmental Protection Agency (EPA), while execution of these policies are assigned to different agencies. For instance, the U.S. Nuclear Regulatory Commission (USNRC) regulates nuclear power plants and the disposal of radioactive waste, the Mine Safety and Health Administration regulates the exposure of miners to radon and gamma rays, and the Food and Drug Administration develops standards for radioactive material concentrations in food, devices that emit ionizing radiation, and medical devices used in radiation therapy. The 1999 USNRC regulations set dose limits at 0.1 rem/year for the general public, 5.0 rem/year for persons with occupational exposure, and 0.5 rem/year for pregnant women.

Chernobyl

In 1986, explosions within a reactor at the nuclear power plant in Chernobyl, Ukraine, led to large releases of radioactive materials into the atmosphere. These materials were deposited all over Europe, particularly in Belarus, Ukraine, and the Russian Federation. This accident provided a unique natural experiment in the effects of radiation exposure on human health. The WHO conducted a series of meetings in the years 2003 to 2005 to review scientific evidence on health effects of the Chernobyl accident and compare it with results from studies of other situations involving high radiation exposure, such as survivors of the atomic bombs dropped on Japan during World War II.

The WHO concluded that the only type of cancer that clearly increased after the Chernobyl accident and that could be directly attributed to radiation exposure from that accident was thyroid cancer. A large increase in thyroid cancer was found among people who lived in the most contaminated areas who were children or adolescents at the time of the accident. This was due to the radioactive iodine released from the reactor, which was deposited in pastures where the cows grazed. Children who consumed the milk produced by theses cows would get affected. A general iodine deficiency in the local diet exacerbated this problem. An increase in leukemia was found among the Chernobyl liquidators (people involved in containing and cleaning up the radioactive debris, many of whom received acute doses of radiation) but not among residents of the contaminated areas.

Increased mortality is expected over the lifetime of people exposed to radiation from the Chernobyl accident, but it is too early to test those predictions against actual mortality rates among residents of the contaminated areas. Among liquidators, 134 were diagnosed with acute radiation sickness (ARS) and 28 died due to ARS in 1986. No effects on fertility or adverse pregnancy outcomes were found that could be attributed to the Chernobyl accident.

—Sarah Boslaugh

See also Birth Defects; Cancer; Environmental and Occupational Epidemiology; Natural Experiment

Further Readings

Agency for Toxic Substances and Disease Registry. (1999). *Toxicological profile for ionizing radiation.* Atlanta, GA: Author. Retrieved December 13, 2006, from http://www.atsdr.cdc.gov/toxprofiles/tp149.html.

Nadakavukaren, A. (2000). *Our global environment: A health perspective* (5th ed.). Prospect Heights, IL: Waveland Press.

World Health Organization. (2006). *Health effects of the Chernobyl accident: An overview* (Fact Sheet No. 303).

Geneva, Switzerland: Author. Retrieved December 13, 2006, from http://www.who.int/mediacentre/factsheets/fs303/en/index.html.

Random-Digit Dialing

Random-digit dialing (RDD) is a method used to select participants for telephone surveys and for related purposes such as selecting control group subjects in case-control studies. The basis of RDD is the random generation of telephone numbers that are used to contact potential survey respondents or study participants. Several major U.S. Federal Government public health surveillance projects use RDD, including the Behavioral Risk Factor Surveillance System (BRFSS) and the National Immunization Survey (NIS). RDD does not require the use of telephone directories and has the advantage of including as potential respondents households with unlisted numbers or who have recently moved or changed phone service; failure to include these types of households can seriously bias the sample. However, RDD has the disadvantage that many of the numbers generated may not be in use or may be nonresidential leading to wasted time and effort.

RDD can be a cost-effective method of selecting subjects in an area where telephone ownership is nearly universal. However, it shares with all telephone-based survey methods the disadvantage that households that do not have telephones generally differ systematically from those who do (e.g., in terms of income, education, and other measures of social capital) and these differences can introduce bias into a study. This can be a major concern in some geographical areas; for instance, in parts of the rural Southern United States, as many as 40% of renter households do not have a telephone. In addition, calculating response rates may be more difficult in RDD surveys than in surveys that used a published telephone directory as a sampling frame.

List-assisted RDD can increase the efficiency of the sampling process. The basis of list-assisted RDD is limiting the randomly generated numbers to groups of numbers, known as 100-blocks, which are known to be in use and contain a high proportion of residential numbers. Each telephone number in the United States is made up of 10 digits—the area code (first three digits), the prefix (the next three digits), and the suffix (the last four digits). The first eight digits are sometimes collectively called 100-blocks because they define sets of 100 telephone numbers with the same first eight digits. Lists of these 100-blocks for the geographical area to be sampled, as well as lists of working phone numbers, may be purchased by firms that specialize in providing this information. Comparing the randomly generated numbers to a list of known business numbers and eliminating those that do not also have a residential listing can further improve efficiency, as can use of a machine to detect the dial tone that precedes the "number not in service message" and eliminating these numbers from the sample.

The increasing popularity of cell phones, in particular the increase in households that do not also have a "land line" (traditional phone) has introduced several other issues. Because cell phone numbers have not traditionally been included in telephone surveys, households with only a cell phone (about 7% of U.S. households in 2005) are excluded from the possibility of participation. In addition, concerns such as safety (a person could answer his or her cell phone while driving, which could lead to an accident), cost to the respondent (because cell phones contacts often include a charge for receiving incoming calls), and low yield (because cell phones are disproportionately owned by children and adolescents) are issues that must be dealt with.

—Sarah Boslaugh

See also Bias; Health Disparities; Social Capital and Health; Study Design; Survey Research Methods

Further Readings

Glaser, S. L., & Stearns, C. B. (2002). Reliability of random digit dialing calls to enumerate an adult female population. *American Journal of Epidemiology*, *155*(10), 972–975.

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2006). Addressed-based versus random-digit-dial surveys: Comparison of key health and risk indicators. *American Journal of Epidemiology*, *164*(10), 10191025.

Voigt, L. F., Davis, S., & Heuser, L. (1992). Random digit dialing: The potential effect on sample characteristics of the conversion of nonresidential telephone numbers. *American Journal of Epidemiology*, *136*(11), 1393–1399.

Randomization

Randomization is a term used in clinical trials to denote a scheme for assigning study subjects to treatment groups using methods that are independent of the individual subjects' characteristics. Typically, when randomization is used, each participant has an equal chance of assignment to each study group or treatment group. Many characteristics of study subjects may affect the relationship of treatment and outcome; some of these are known to the researcher in advance, some are not known. By randomization we hope to sort people with these characteristics equally between the treatment groups. Randomization should also yield equal distributions of characteristics that affect the outcome in ways that the researcher did not anticipate.

The effectiveness of randomization is evaluated through comparing the resulting treatment groups on baseline characteristics and demographics. If a subject characteristic is found unequally distributed between treatment groups despite randomization, it should be treated as a potential confounder in the analysis.

The effectiveness of randomization is sensitive to the size of the study population. A large study population will increase the chances that randomization will be successful in yielding equivalent distributions of participant characteristics; a smaller population is more susceptible to unequal assignment to groups through chance.

The unit of randomization may be the individual study subjects, or it may be larger groups. The unit of randomization may be groups such as clinic, hospital, neighborhood, town, city, or other social groupings. In this case, the entire group would be randomly assigned to one treatment group. For example, in a trial of the effectiveness of smoking-cessation messages, entire communities may be randomly assigned to different types of smoking-cessation messages.

Methods for implementing randomization include using a random number generator to assign a number to each subject with a set method of allocating specific numbers to particular treatments. For example, a random number generator is used to assign the numbers 0.1, 0.3, 0.8, 0.4, 0.3 to the first five study subjects and by prior decision those with an odd number are assigned to the active treatment and those with an even number are assigned to the control treatment. Study Subjects 1, 2, and 5 would be assigned to the active treatment, while Subjects 3 and 4 would be assigned to the control treatment.

Some methods used to assign subjects to groups are not as truly random and should be used with caution if at all. For example, if clinic is held 4 days a week, subjects who come in on Monday or Wednesday may be assigned to one treatment, while those who come in on Tuesday or Thursday are assigned to the other treatment. This method may create effective randomization if there is no relationship between the day of the week and other subject characteristics, but the burden of establishing this rests on the researcher.

Nonrandomized trials may have problems with selection bias if patients are assigned to the treatment group according to some characteristic. In an early trial of cardiac care units (CCUs), heart attack patients who were deemed at greater risk were preferentially sent to the CCU rather than the comparison, the standard treatment. In the analysis, the CCU was found to have higher mortality than the standard treatment, but the comparison was skewed by the more serious condition of the patients assigned to the CCU compared with those who received the standard treatment.

—Sydney Pettygrove

See also Bias; Clinical Trials; Confounding

Further Readings

Friedman, L., Furberg, C. D., & DeMets, D. L. (1999). *Fundamentals of clinical trials* (3rd ed.). New York: Springer.

Piantadosi, S. (2005). *Clinical trials: A methodologic perspective* (2nd ed.). New York: Wiley-Interscience.

Random Variable

A variable whose observed values may be considered as outcomes of a stochastic or random experiment is called a *random variable.* The values of such a variable in a particular sample cannot be anticipated with certainty before the sample is gathered. Random variables are commonly classified as qualitative or categorical, discrete, or continuous.

A random variable is defined as a qualitative or categorical variable if its set of possible values do not represent numerical information. For example, gender is a categorical variable. Suppose that among 100 patients, there are 65 females and 35 males, and let X be the sex of a randomly chosen patient among these 100 patients. Then X is a qualitative *random variable*, and the values of X are "Female" and "Male" with 65% and 35% chance to be chosen, respectively. Even if numeric values, such as 0 and 1, are used to code gender in a data set, it remains a categorical variable because the values represent membership in a category rather than a measured quantity.

A random variable is discrete if its set of possible values is countable. If only two values are possible, such as alive versus dead, it may also be called a binomial random variable. For example, a new technique, balloon angioplasty, is being widely used to open clogged heart valves and vessels. The balloon is inserted via a catheter and is inflated, opening the vessel; thus, no surgery is required. Suppose that among untreated people with heart-valve disease, about 50% die within 2 years, and experience with balloon angioplasty suggests that approximately 70% treated with this technique live for more than 2 years. We can define X as the number of patients who will live more than 2 years, among the next five patients treated with balloon angioplasty at a hospital. Then, X constitutes a *discrete random variable*, which can take on the values 0, 1, 2, 3, 4, or 5.

To make an inference about the population from our sample data, we need to know the probability associated with each value of the variable that is called its *probability distribution.* Probability calculations are relatively simple for discrete variables and are often displayed in tabular form, as presented below. The *probability distribution* for a *discrete random variable* X displays the probability $P(x)$ associated with each value of x. This display can be presented as a table, a graph, or a formula. To illustrate, consider the above example; all possible values of X are 0, 1, 2, 3, 4, and 5. The probability distribution is a binomial distribution with $n = 5$ and $p = 0.7$ that can be given by the formula:

$$P(X = k) = \frac{n!}{k!(n-k)!}(0.7)^k(0.3)^{n-k},$$
$$k = 0, 1, 2, \ldots, 5.$$

It may also be displayed as shown in Table 1.

Table 1 Probability Distribution for Discrete Random Variable X

X	$P(x)$
0	0.00243
1	0.02835
2	0.13230
3	0.30870
4	0.36015
5	0.16807

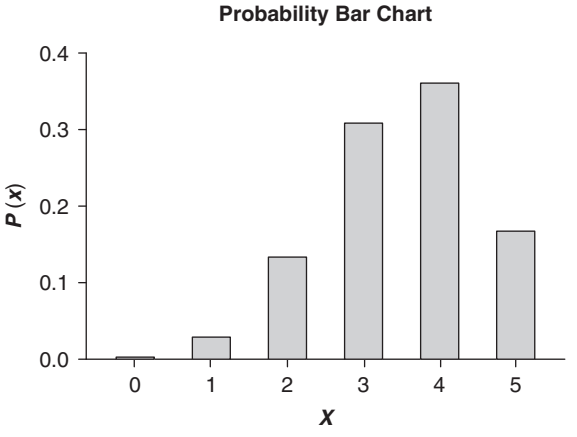


Figure 1 Bar Chart Displaying Probability Distribution for Discrete Random Variable X

Or, it may be presented graphically as a bar chart, as shown in Figure 1.

The properties of discrete random variables are as follows:

- The probability associated with every value of x lies between 0 and 1.
- The sum of the probabilities for all values of x is equal to 1.
- The probabilities are additive, that is, $P(X \geq 4)$ is the same $P(X = 4) + P(X = 5)$.

A random variable is defined as continuous if its set of possible values is an entire interval on the number line, that is, if it can take any value within a range rather than only a discrete set of values such

as was specified in the previous example. Of course, any measuring device has a limited accuracy and therefore a continuous scale may in practice be something of an abstraction. Some examples of continuous random variables are the height of an adult male, the weight of a newborn baby, a patient's body temperature, and the survival time of a patient following a heart attack.

To describe the distribution of a continuous random variable, a probability density function $f(x)$ is used, which has three properties:

1. The total area under the probability density curve is 1.
2. $P\{a \leq X \leq b\} =$ Area under the probability density curve between a and b.
3. $f(x) \geq 0$ for all x.

Unlike the description of a discrete probability distribution, the probability density $f(x)$ does not represent the probability that the random variable will exactly equal the value x. Instead, a probability density function relates the probability of a value falling within the range between a and b, that is, the areas of the curve over that interval. The probability that $X = x$ for a continuous random variable is always equal to 0.

The last statement may need some clarification. In the birthweight example, $P\{X = 8.5\} = 0$ probably seems shocking. Does this mean that no child can have a birth weight of 8.5 lb? No. To understand it, we need to recognize that the accuracy of every measuring device is limited, so that here the number 8.5 is actually indistinguishable from all numbers in an interval surrounding it, say [8.495, 8.505], and the area under the curve between this interval is no longer 0.

There are many useful continuous random variables, each with a specific distribution or set of distributions described by mathematical functions that allow us to compute probabilities regarding specific values. The most famous of these distributions is the normal distribution, also called the bell curve, Z distribution or Gaussian distribution, which plays a special role in statistical theory (through the Central Limit Theorem) as well as in practice.

—Renjin Tu

See also Binomial Variable; Central Limit Theorem; Normal Distribution; Sampling Distribution; Z Score

Further Readings

Devore, J., & Peck, R. (2004). *Statistics: The exploration and analysis of data* (5th ed.). Belmont, CA: Duxbury Press.
Johnson, R. A., & Bhattacharyya, C. K. (2005). *Statistics principles and methods* (5th ed.). New York: Wiley.

Rate

A rate is a measure of change in one quantity with respect to change in another. As used in epidemiology, this typically refers to an incidence rate, where the numerator is the number of new events and the denominator is total person-time at risk. This is one of the key measures of occurrence of disease in populations and gives an estimate of how fast disease or death is happening in a given population.

An example calculation of an incidence rate can be done using data from Table 1. A total of three events occurred, and the total person-time at risk summed over all population members is $30 + 17 + 22 + 11 + 20 = 100$ person-years, giving a rate of 3/100 or 0.03 per year. In calculating a rate, events counted in the numerator should be those occurring among people contributing person-time to the denominator. Likewise, the denominator should include only person-time during which any events experienced by the subject would be counted in the numerator. Sometimes the denominator can be estimated as average population size times follow-up time for a relatively short period of time with stable population level. For example, this is often done for an annual mortality rate in a geographic area, such as a state.

Some properties of incidence rates include the following:

- They range from 0 to infinity.
- Units are (time) $-$ 1, where any unit of time can be used.

Table 1 Data for Sample Calculation of Rate

Person ID#	*Total Years of Follow-Up*	*Event*
01	30	N
02	17	Y
03	22	Y
04	11	N
05	20	Y

- The actual measure depends on the unit of time used in the denominator.

For example, the following are equivalent:

$$1\frac{\text{Event}}{\text{Person}-\text{year}} = 0.083\frac{\text{Events}}{\text{Person}-\text{month}} = 10\frac{\text{Events}}{\text{Person}-\text{decade}}.$$

The same rate may arise through alternate scenarios involving different lengths of follow-up time and population sizes. For example, following 100 people for an average of 1 year each and observing three events would give an incidence rate of 0.03 per year. The same rate of 0.03 per year would also be calculated if three events were observed among only five people followed for an average of 20 years, as shown in the example above.

Incidence rates are occasionally reported in terms of change in a unit other than person-time—for example, motorist fatality rates per person-mile or aviation events per pilot-flight hour. The rates given simply per unit time as opposed to per unit person-time may be referred to as absolute rates.

The term *rate* has sometimes been used in a more general sense to refer to proportions or ratios. The concept of rate as different from risk (a proportion) was elucidated in the 19th century by William Farr. Farr reported vital statistics for England and contrasted cholera with tuberculosis. The former had a higher *rate* of death among patients, because the disease could be quickly fatal; whereas the latter had a higher *risk* of death, since a greater percentage of those falling ill would eventually succumb to the disease. Even so, use of terminology such as *attack rate* and *prevalence rate* for measures that are technically proportions still persists.

Incidence rate is also known as incidence density, person-time rate, and force of morbidity or mortality.

—Keely Cheslack-Postava

See also Incidence; Mortality Rates; Person-Time Units; Proportion; Ratio

Further Readings

Elandt-Johnson, R. C. (1975). Definition of rates: Some remarks on their use and misuse. *American Journal of Epidemiology*, *102*(4), 267–271.

Morabia, A. (2004). Epidemiology: An epistemological perspective. In A. Morabia (Ed.), *A history of epidemiologic methods and concepts* (pp. 3–125). Basel, Switzerland: Birkhauser Verlag.

Rothman, K. J., & Greenland, S. (1998). Measures of disease frequency. In K. J. Rothman & S. Greenland, *Modern epidemiology* (2nd ed., pp. 29–46). Philadelphia: Lippincott Williams & Wilkins.

Ratio

A ratio is an expression of the magnitude of one quantity in relation to another. Ratios are typically expressed by two numbers separated by a colon, for instance, 4:3, read as "four to three" and meaning that there are four units of the first items for every three units of the second item. Ratios do not require that the two numbers have common units and in fact are typically used to express the relationship between two quantities consisting of different units. For instance, the ratio of male to female patients in a hospital might be expressed as 2:1, meaning there were twice as many male patients as female patients or that there were two male patients for every one female patient.

The concept of ratio has to be clearly distinguished from the definitions of proportion and of rate. A ratio is a fraction in which the numerator is not necessarily a part of the denominator or, in other words, in a ratio the numerator is not necessarily included in the population defined by the denominator. In contrast, in a proportion the numerator by definition is included in the denominator. Taking the hospital example again, if the ratio of male to female patients is 2:1, in order to express this as a proportion we must introduce the unit of "patient" (as opposed to male patient and female patient) to be able to make the statement that proportion of male patients among all patients is 66.7% or two thirds; in this case of the proportion, male patients are included in both the numerator and denominator of the fraction. Ratios are distinguished from rates because ratios do not include a measure of time in the denominator.

The main properties of ratios are that they are greater than zero, they may or may not be greater than 100, and may or may not have units. Ratios may also be expressed as percentages. Ratios are commonly used in epidemiology and public health: For instance, the risk ratio, also known as relative risk, is used to

express the risk of a person developing a condition given a particular exposure, relative to those lacking the exposure. Odds ratios similarly express the odds of developing a condition given an exposure, compared with those who do not have the exposure. Both the risk ratio and odds ratio are dimensionless.

Ratios are also used in epidemiology to express availability of services or cases of disease for a particular population. For instance, a commonly reported measure of health care availability is the number of hospitals or hospital beds per 10,000 people, which is calculated by dividing the number of hospitals or beds by the population size and multiplying by 10,000. Obviously, the numerator in these cases are hospitals or hospital beds, and the denominator in both cases are people, so they do not have a common unit. Similar examples include the per capita income: that is, the total income earned during a year by a group of people divided by the number of people (units = dollars per capita); and the mortality (or death) rate: that is, the number of deaths during a specified period divided by the number of persons at risk of dying during this period (units = deaths per 100 people; larger units such as per 10,000 people can be used for rare diseases or when mortality is rare). Note that the terms *ratio* and *rate* are sometimes used interchangeably, particularly when speaking of statistics such as the number of hospitals per 10,000 people. However, many epidemiologists prefer to reserve the term *rate* to refer to numbers expressed per unit time, such as infections per year.

—Carlos Campillo

See also Proportion; Rate

Further Readings

Armitage, P., & Berry, G. (1994). *Statistical methods in medical research*. London: Macmillan.

Daniel, W. W. (2004). *Biostatistics: A foundation for analysis in the health sciences*. New York: Wiley.

Receiver Operating Characteristic (ROC) Curve

The receiver operationg characteristic (ROC) curve is a two-dimensional measure of classification performance depicting the trade-off between sensitivity and specificity. It is used in the analysis of a diagnostic test or screening test that classifies experimental units into two categories such as diseased (D) or nondiseased ($\bar{D}$). Screening and laboratory test results are usually reported as a continuous variable. For example, the risk variable serum concentration of creatine phosphokinase for myocardial infarction (D) is approximately normally distributed varying from less than 100 units/ml to greater than 4,000 units/ml. The serum concentration of creatine phosphokinase for those without myocardial infarction ($\bar{D}$) also has an approximate normal distribution but has a different mean (see Figure 1). Suppose that we dichotomize the serum concentration by some cutpoint so that values above it represent positive ($+$) test results and values below it represent negative ($-$) test results. We may now define the following misclassification rates: *false-positive rate* $P(+|\bar{D})$ is the probability of classifying a noncase as positive, *true-positive rate (sensitivity)* $P(+|D)$ is the probability of classifying a case as positive, *false-negative rate* $P(-|D)$ is the probability of classifying a case as negative, and *true-negative rate (specificity)* $P(-|\bar{D})$ is the probability of classifying a noncase as negative. As shown in the table below, different cutpoints lead to tests with different levels of misclassification rates. For example, when the cutoff value of the serum concentration is chosen to be 5.4, calculation from the two normal distributions gives $P(+|D) = .725747$ (see the shaded area in Figure 1), $P(-|D) = .27425$, $P(-|\bar{D}) = 991802$, and $P(+|\bar{D}) = .008198$ (see the small double-shaded area in Figure 1). Note that if we lowered the cutoff value, we would decrease the false-negative rate, but we would also increase the false-positive rate. Similarly, if we raised the cutoff value, we would decrease the false-positive rate, but we would increase the false-negative rate (see Figure 1).

An ROC curve is obtained by plotting the false-positive rate (1 − specificity) against the true-positive rate (sensitivity) for a series of cutpoints defined by the test (see Figure 2). It shows the trade-off between the true-positive rate and the false-positive rate of a test (any increase in sensitivity will be accompanied by a decrease in specificity and conversely). In statistical terminology, it is the plot of Type I error against the power. This ROC plot is representative of those plotting one conditional distribution function against another found to be useful in epidemiology and other health sciences, which includes plotting the posttest

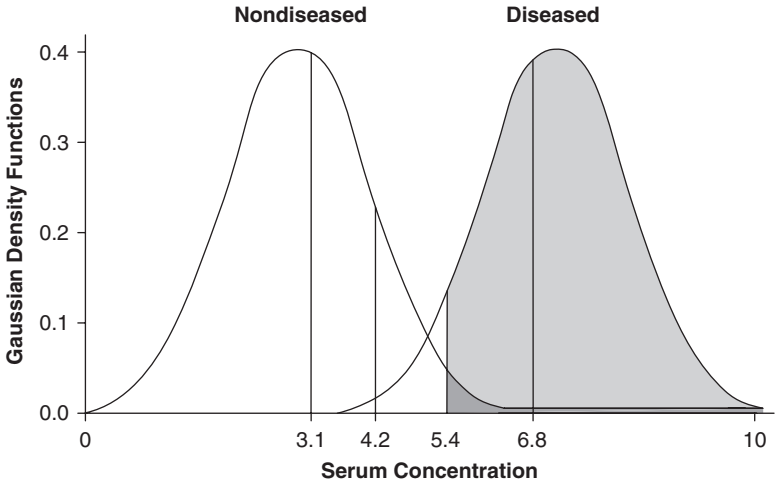


Figure 1 Distribution of Serum Concentration: Diseased Versus Nondiseased

probability of disease given the test is positive against the pretest probability of disease, plotting the positive predictive value against the point prevalence rate, and plotting the total time on test against the distribution function of the duration time. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The closer the curve comes to the 45° diagonal of the ROC space, the less accurate the test. The slope of the tangent line at any point on the ROC curve may be accurately estimated by spline interpolation and differentiation. The slope of the tangent line at a cutpoint gives the likelihood ratio (LR) for that value of the test. So, by choosing the slope of the tangent to the ROC curve to equal the LR that will minimize the total cost of making false-positive and false-negative errors, one can identify the optimal cutoff values. Such LR turns out to be the ratio of the product of the net cost of treating nondiseased patients and the pretest probability of no disease to the product of the net benefit of treating diseased patients and the pretest probability of disease.

The area under the curve (AUC) is a measure of test accuracy, namely, a measure of how well the risk variable discriminates a disease state. If you take a random person from the nondiseased population and obtain a value X for the serum concentration and a random person from the diseased population and get a score of Y, then the area under the ROC curve represents $P\{Y > X\}$. This implies that the more apart the distribution for the diseased is from the distribution for the nondiseased, the more accurate is the test. In other words, the accuracy of the test depends on how well the test separates the group being tested into those with and without the disease in question. AUC = 1, which corresponds to the left and top border of the ROC space, represents a perfect test, AUC = .5, which corresponds to the 45° diagonal of the ROC space, represents a random (hence useless) test, namely that the risk variable is completely independent of disease so that the probability of detecting disease will be the same for those with and without disease (the two distributions—one for the diseased and the other for the nondiseased—of the risk variable completely overlap). Here is a rough guide for classifying the accuracy of a diagnostic test using AUC:

96% to 100% = excellent

90% to 96% = very good

80% to 90% = good

70% to 80% = fair

60% to 70% = poor

50% to 60% = useless

Construction of ROC Curve

Suppose the serum concentrations for both diseased and nondiseased populations are normally distributed with the same variance but different means. They are transformed into $N(3,1)$ and $N(7,1)$ distributions in Figure 1. We first choose a series of cutpoints on the

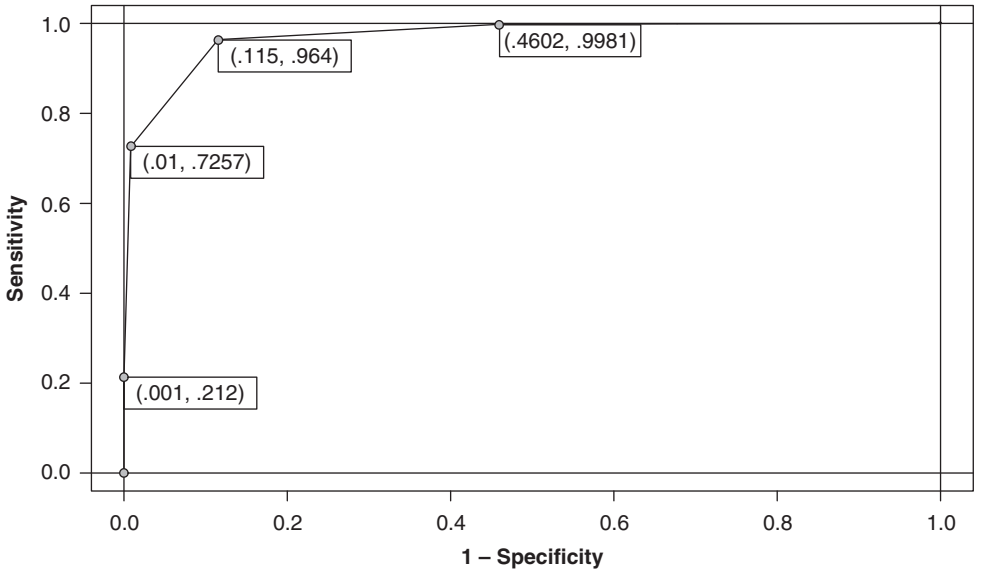


Figure 2 ROC Curve for Serum Creatine Phosphokinase

serum concentration 3.1, 4.2, 5.4, and 6.8 and erect vertical lines at these cutpoints in Figure 1. We then compute the corresponding sensitivity and false-positive rate (1 − specificity) for each cutpoint from the two normal distributions. These are the areas from each vertical cutpoint line to the right tails of the two respective normal curves. The two shaded areas in Figure 1 correspond to the sensitivity = 0.725747 and 1 − specificity = 0.008198 when cutoff value is chosen to be 5.4. The results for the four chosen cutoff values are as shown below:

Serum Concentration	*Sensitivity*	*1 − Specificity*
3.1	0.998134	0.460172
4.2	0.964070	0.115070
5.4	0.725747	0.008198
6.8	0.211855	0.000072

The data given in the last two columns (1 − specificity, sensitivity) are then graphed to obtain the ROC curve with 1 − specificity on the horizontal axis and sensitivity on the vertical axis as shown in Figure 2. These data are also used to compute the area under the ROC curve, where Area = $(1 + .998134) \times (1 - .460172)/2 + (.96407 + .725747) \times (.11507 - .008198)/2 + (.725747 + .211855) \times (.008198 - .000072)/2 + (.211855 \times .000072)/2 = .972$, by the trapezoidal rule, which, according to the criteria stated above, is considered to be excellent. This means that the relative ordering of the serum concentration of creatine phosphokinase has a 97.2% probability of correctly distinguishing a person with myocardial infarction from a normal person. A more accurate estimate of the area may be obtained by cubic spline interpolation and integration. When data on frequencies of various categories of the risk variable defined by the cutpoints for both the diseased and the nondiseased samples are available, sensitivity $P(+|D)$ and specificity $|P(-|\bar{D})$ can be estimated directly from these data as, $\#(+, D)/\#(D)$ and $\#(-, \bar{D})/\#(\bar{D})$, respectively, where $\#(x)$ stands for the number of x. These estimates can then be used to construct the empirical ROC curve by plotting the estimates of 1 − specificity versus the estimates of sensitivity. The resulting empirical curve may then be smoothed by smoothing splines.

Finally, the construction of two-dimensional ROC curve described above can also be generalized to construct the three-dimensional ROC surface just as plane geometry has been generalized to solid geometry.

—John J. Hsieh

See also Life Tables; Likelihood Ratio; Normal Distribution; Screening; Sensitivity and Specificity; Type I and Type II Errors; *Z* Score

Further Readings

Biggerstaff, B. J. (2000). Comparing diagnostic tests: A simple graphic using likelihood ratios. *Statistics in Medicine*, *19*(5), 649–663.

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29–36.

Zweig, M. H., & Campbell G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, *39*(4), 561–577.

Reed, Walter
(1851–1902)

Walter Reed was a surgeon in the U.S. Army who significantly contributed to knowledge of the etiology and epidemiology of yellow fever. Reed's work is significant in that he focused on the means of disease transmission rather than a specific disease agent and, in doing so, greatly reduced infection rates. His yellow fever experiments also established the important role of the "healthy volunteer" in epidemiologic research and contributed greatly to the formalization and documentation of informed consent. Reed was born in Belroi, Virginia, and became a medical officer in the U.S. Army after graduating from the University of Virginia medical school. He remained in the military for the remainder of his life.

During the Spanish-American War, yellow fever killed thousands of soldiers in Cuba—more than died in battle—and continued to threaten troops occupying the island as well as individuals throughout North and South America. For several decades, scientists and local physicians had proposed that yellow fever was mosquito borne, but the insect's exact role was unclear. In 1900, Surgeon General George Sternberg established the Yellow Fever Commission under Reed's direction, and Reed went to Cuba.

Because there was no animal model in which to study yellow fever, identifying the exact mode and source of transmission required humans. Reed and his colleagues designed an experiment in which common house mosquitoes (now known as *Aedes aegypti*) that had fed on yellow fever patients were allowed to bite noninfected individuals. Reed's colleague suggested that the research team serve as the first group of subjects; after two physicians became ill (and one eventually died), Reed decided to forego self-experimentation. Instead, healthy volunteers (primarily soldiers and native Cubans) were recruited and separated into two groups—those who would be bitten and those who would be exposed to soiled bedding from patients (another potential suspect). The theory that yellow fever is transmitted by mosquitoes and not direct contact with an infected individual was confirmed.

Reed's research was recognized by Congress, and his reputation as a heroic researcher and the bravery of his colleagues were celebrated for decades. After conducting malaria research in Cuba, he returned to Washington, D.C., to teach pathology and bacteriology at the Army Medical School and the George Washington University Medical School. Reed's health began to decline following an appendectomy; in 1902, he died of peritonitis and was buried in Arlington National Cemetery. Named in his honor, Walter Reed General Hospital in Washington, D.C., opened on May 1, 1909. In 1951, the hospital was renamed the Walter Reed Army Medical Center, now the premier military medical facility in the eastern United States.

—Emily E. Anderson

See also Ethics in Human Subjects Research; Informed Consent; Insect-Borne Disease; Yellow Fever

Further Readings

Bean, W. B. (1982). *Walter Reed: A biography.* Charlottesville, VA: University of Virginia Press.

Pierce, J. R., & Writer, J. (2005). *Yellow Jack: How yellow fever ravaged America and Walter Reed discovered its deadly secrets*. Hoboken, NJ: Wiley.

Reed, W., & Agramonte, A. (1983). Landmark article. February 16, 1901: The etiology of yellow fever. An additional note. *Journal of the American Medical Association*, *250*(5), 649–658.

Web Sites

Walter Reed Army Medical Center. http://www.wramc.amedd.army.mil.

Regression

Many analyses of epidemiologic data are conducted using statistical methods common to other research

fields, including the social and biologic sciences. In epidemiologic studies, however, the focus of the research question, and thus the methods used to study this question, tend to differ. The combination of human, social, environmental, and biological factors that may be present in epidemiologic studies can lead to a complexity not seen in large randomized trials or when the environment can be controlled. This entry discusses regression methods used in epidemiology and the conceptual framework that underlies these methods.

Predictive Analysis Versus Associative Analysis

Epidemiologic research questions tend to fall into two general categories: (1) What factors best explain or predict the occurrence of another factor (or outcome)? (2) What is the association between exposure(s) and outcome(s)? The analytic methods used for these research questions are known as predictive or associative, respectively.

Regression analysis is used for both analysis of prediction and association. However, the selection of factors included in the model differs based on the type of analysis performed. While most statistics books focus on measures of prediction, epidemiologic studies are primarily concerned with questions of association. Because measuring associations is the most common use of regression analysis in epidemiology, this methodology is the focus of this entry. Analysis for questions of prediction will primarily be discussed to provide a contrast on how analysis differs from that for estimating measures of association.

Prediction

Research questions focused on prediction take two main forms. They may seek to identify any factors that may influence the detection of a health outcome, or they may seek to identify which of the factors are most predictive of development of the health outcome in affected individuals.

An example of the first use of predictive analysis is the identification of victims of intimate partner violence. A study may be done to identify indicators of partner violence victimization for women seen in a primary care setting. In such a study, a group of factors is found to identify victims. These include injuries, multiple nonspecific physical symptoms (e.g., pain, fatigue, headaches, diarrhea), and psychiatric diseases (e.g., depression, anxiety, post-traumatic stress disorder) as well as characteristics of the victim (e.g., young), perpetrator (e.g., young, excessive alcohol use), and relationship (e.g., wife makes more money than husband). Many of these factors are common among women seen in primary care (e.g., young age, depression). The more characteristics a woman has that were identified as predictive in regression, the more likely that she is a victim of partner violence. This analysis had no interest in identifying the "best" predictor, but in understanding what factors, alone or in combination, predict partner violence so that these factors can be communicated to both physicians and patients. Some of the factors so identified may be outcomes rather than causes of partner violence, but this is not a concern when the purpose of the study is to identify potential victims of partner violence rather than make causal statements about it.

The goal of the second type of predictive model is to identify the most important factors that predict the outcome. For example, it is known that infection with Hepatitis C virus is a risk factor for development of liver cancer. However, not all individuals who are infected with Hepatitis C virus develop this cancer. Now that we can test for Hepatitis C, it may help understand what factors best predict liver cancer among those infected. It is then necessary to examine these other factors, which may include coinfection with Hepatitis B virus, gender, viral genotype, liver enzyme level, use of alcohol or tobacco, and environmental and occupational exposures. Using predictive modeling, we can identify which factors best predict development of liver cancer and then monitor the group with these characteristics more carefully to identify early disease and focus care to more aggressively reduce risk of liver cancer. Here, we are interested in identifying factors that are predictive of liver cancer, not how the factors are associated with or cause cancer.

Measures of Association

Questions of association focus on the estimation of the strength of association between an exposure, or exposures, and an outcome. Studying associations in this way helps inform us about the causes of disease, hospitalization, death, and other health-related outcomes. While a strong association does not mean that the exposure caused the outcome, establishing that an

association exists is a critical piece of information to assess potential causality. Examples of studies that focus on association include measurement of the association between location of work within a chemical plant and the risk of developing cancer, between infection with a particular microorganism and development of a clinical disease, and between exposure to airborne dust and development of asthma.

In epidemiology, we are most often focused on identifying causal relationships, that is, determining the association between potential exposures (i.e., risk factors) and an outcome(s). Because other factors can complicate this association, particularly confounders and effect modifiers, regression analysis is a valuable tool to estimate the association when the relationship is complex; that is, confounders or effect modifiers exist.

Regression Analysis to Estimate Measures of Association

Although stratified analysis may be used to examine confounders, this approach quickly becomes problematic when many confounders exist. Many 2×2 tables need to be generated and analyzed, and as the number of tables grows, so does the potential for zero values in the table cells, which can lead to a poor estimate of association strength. Multivariate regression methods may be used to study these associations while taking into account all the potential confounders. Logistic, log-binomial, Poisson, and linear regression are discussed here to provide insight into these methods.

Regression Model Format

Most regression equations model the relationship between an outcome measure and a function (e.g., logit, log) of a linear combination of the independent factors and regression parameters. In studies designed to estimate the measure of association, the independent variables are made up of the exposure(s), potential confounders, and interaction terms for potential effect modifiers. The key difference in analysis between studies of prediction and studies of association is variable selection. For studies measuring associations, classic stepwise regression techniques are not appropriate; rather, variables need to be assessed with regard to their role as potential effect modifiers and confounders.

In a regression analysis, all information about association between exposure and outcome is stored in the slopes (β) of factors that contain the exposure term. Consider the following general combination of independent factors for a study:

1. $\beta_0 + \beta_1 \times E$
2. $\beta_0 + \beta_1 \times E + \beta_2 \times C_1 + \beta_3 \times C_2 + \beta_4 \times C_3$
3. $\beta_0 + \beta_1 \times E + \beta_2 \times C + \beta_3 \times M + \beta_4 \times E \times M$
4. $\beta_0 + \beta_1 \times E + \beta_2 \times C_1 + \beta_3 \times C_2 + \beta_4 \times C_3 + \beta_5 \times M_1 + \beta_6 \times M_2 + \beta_7 \times E \times M_1 + \beta_8 \times E \times M_2$

where

β (beta) is the regression coefficient

E is a dichotomous exposure (1 = exposed, 0 = not exposed)

Cs are potential confounders: C_1 is dichotomous (1 = present, 0 = absent), C_2 is dichotomous (1 = present, 0 = absent), C_3 is continuous

Ms are potential effect modifiers: M_1 is dichotomous (1 = present, 0 = absent), M_2 is continuous

Typically, a model including all factors of interest is created and fit to the data. The exception to this approach is when there is substantial collinearity; that is, the factors overlap a great deal causing mathematical problems in estimating βs.

For a model with a dichotomous exposure factor as the sole independent variable (Model 1), the measure of association between the exposure and outcome is simply the $\exp(\beta_1)$, that is, the odds ratio (in a logistic regression model), relative risk (in a log-binomial model), or rate ratio (in Poisson regression).

For studies where the measure of association may be confounded but no effect modification exists (Model 2), the measure of association is also estimated simply by $\exp(\beta_1)$; however, this estimate is different from the estimate from Model 1, as it is adjusted by the potential confounders during the iterative process used to estimate βs. To determine if the potential confounders are, in fact, confounders in the study, each one is removed, the model is rerun without that factor, and $\exp(\beta_1)$ (i.e., the odds ratio for the exposure) is examined to determine if it has changed compared with its value when the potential confounder was in the model. If $\exp(\beta_1)$ for the exposure is about the same whether the potential confounder is included in the model or not, then it is not

a confounder in this study. If $\exp(\beta_1)$ for the exposure does change between the two models, then the potential confounder is a confounder in this study, and it must remain in the model to remove the effects of confounding associated with it. How much does $\exp(\beta_1)$ need to change to provide evidence of confounding? The answer is "it depends." Some epidemiologists use a 10% rule; that is, if the $\exp(\beta_1)$ changes by 10% or more between the two models, then the factor removed is a confounder. Others use a more subjective rule based on the study measures; that is, determine if the change is meaningful in the interpretation of the association between exposure and outcome. Still others do not believe it is appropriate to remove any potential confounder that was considered based on the literature, even if it is not a confounder for the study.

Usually, if potential effect modifiers exist, they are assessed first. Assessing effect modification is done by evaluating βs for the interaction terms. Epidemiologists focus on βs and not the *p* values for decision making, because the *p* value is affected by factors besides strength of association, such as sample size. As is discussed below, effect modification exists if there is a different association identified between exposure and outcome based on a third factor—for example, when the association between gender and risk of asthma onset is modified by age. For measures of association between an exposure (dichotomous) and outcome, the odds ratio (logistic regression), relative risk (log-binomial), and rate ratio (Poisson regression) is different for young children than for adolescents. In Model 3, with one dichotomous exposure and one dichotomous potential effect modifier, the odds ratio, relative risk, and rate ratios are estimated by

$$\text{Modifier} = 0 \text{ Measure of association} = \exp(\beta_1).$$
$$\text{Modifier} = 1 \text{ Measure of association} = \exp(\beta_1 + \beta_4).$$

It should be noted that βs are adjusted for the other factors in the model (i.e., the confounder and main effect of the modifier). These adjustments are done during the iterative process that is used to estimate βs.

To determine if *M* is actually an effect modifier, we assess if $\beta_4 = 0$. If β_4 is about equal to zero, then there is no effect modification, and the estimated association is about $\exp(\beta_1)$ for each level of *M*. A large *p* value (i.e., one that is not statistically significant) could occur for two reasons: (1) $\beta_4 \sim 0$, that is, no effect modification is evident or (2) there are not sufficient data to assess effect modification (β_4 appears different than 0, but there is a relatively small sample size in some of the modifier levels, so the estimate β_4 lacks precision). Thus, *p* values need to be considered in conjunction with the actual βs to understand if no effect modification exists or if insufficient data are available to adequately assess effect modification.

Logistic Regression

Logistic regression is the most popular regression analysis method used in epidemiology today. Computer programs for widespread use were developed in the 1970s to respond to the data analysis needs of the Framingham Study. Now, user-friendly statistical software for logistic regression is widely available for researchers.

Models and Formulae

At the heart of logistic regression is the odds ratio, which is exactly what the name implies: a ratio of two odds. In fact, logistic regression provides a direct method to compute adjusted odds ratios, adjusting for confounders. In a case-control study, the exposure odds are calculated; that is, the odds of cases being exposed versus the odds of controls being exposed. In a cohort study or randomized trial, the incidence odds are computed; that is, the odds of the outcome among the exposed and the odds of the outcome among the not exposed. Fortunately, the model looks the same and the odds ratios are calculated similarly regardless of the study design. In logistic regression, the probability of a specific outcome, $P(Y = 1)$, is modeled as a function of the factors of interest (i.e., exposure(s), confounder(s), and effect modifier(s)) and regression parameters. In its simplest form, with only one dichotomous exposure factor (*E*, with 1 = exposure and 0 = no exposure), the model is as follows:

$$P(Y = 1|E = 1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}.$$

The odds ratio is then

$$OR = e^{\beta_1}.$$

Regression analysis is typically not conducted with just a dichotomous exposure and outcome, because this can be done more simply using a 2×2 table, but

is useful when a model includes multiple confounders. For instance, a model might include E as the exposure of interest (1 = exposed and 0 = not exposed) and C_1 and C_2 as potential confounders. Thus,

$$P(Y=1)=\frac{\exp(\beta_0+\beta_1E+\beta_2C_1+\beta_3C_2)}{1+\exp(\beta_0+\beta_1E+\beta_2C_1+\beta_3C_2)}.$$

The adjusted odds ratio between exposure and outcome is

$$aOR=e^{\beta_1}.$$

The simplicity of this formula is due to the fact that the adjustment of β, and thus the odds ratio, is done during the iterative process used to estimate βs. Thus, β is adjusted for all the other factors that are included in the model.

Logistic Regression Model Assumptions

The following are the assumptions made with this model:

- The outcome variable is binomial.
- All observations must be independent and identically distributed, or the dependence between observations must be taken into account in the analysis (for instance, by using generalized estimating equations).
- The sample size is large (or an exact program is used, such as LogXact). The suggested guidelines for adequate sample size typically range from 10 to 15 cases per independent variable.
- The factors are linear in the logit scale.
- The model fits the data.

Advantages and Disadvantages

The main advantages of logistic regression are its natural fit to study data from public health and medical research and ease of use due to widely available software. Compared with other types of modeling with categorical outcomes, logistic regression is not prone to issues with the outcome data such as the problem of overdispersion, sometimes seen with Poisson regression, and the problem of estimates existing outside the appropriate boundary parameter, as can happen with log-binomial regression.

However, because the output of logistic regression is the odds ratio, the standard warnings governing the use of this measure apply, most notably the overestimation of the relative risk when the outcome is common, usually defined as greater than 10% for each set of characteristics (i.e., each covariate pattern). When the outcome is mathematically rare, the odds ratio is a good estimate of the relative risk.

Poisson Regression

Poisson regression is most commonly used when data exist as counts of events per a unit of measure (e.g., time, area). Cohort studies and randomized trials that estimate rates because of different follow-up times for participants will typically use Poisson regression to estimate adjusted rate ratios. Also, while data can exist as rates (e.g., number of deaths per month), the use of rates is not required for Poisson regression. Nonrate measures of risk (e.g., number of nurses colonized with methicillin-resistant *Staphylococcus aureus* in hospital wards) also often follow a Poisson distribution and can be modeled using this technique.

Poisson regression can also be used to estimate adjusted relative risks of studies of common outcomes (i.e., when more than 10% of the participants develop the outcome). In this case, all participants must be followed for the same length of time, a requirement for the direct estimation of relative risk. However, in this case a robust approach is needed to compute confidence intervals. This is the approach suggested by Spiegelman and Hertzmark (2005).

Models and Formulae

Poisson regression models the natural logarithm of the expected value of the outcome, $\mu=E(Y)$, as a linear combination of regression parameters and independent factors. For example, a single dichotomous exposure, E, with two confounders, C_1 and C_2, can be expressed as

$$\log(\mu)=\beta_0+\beta_1E+\beta_2C_1+\beta_3C_2.$$

The adjusted rate ratio between exposure and outcome is

$$aRR=e^{\beta_1}.$$

Given a Poisson model, the probability of the dependent variable being equal to a given value (k) can be calculated as

$$P(Y=k)=[(e^{-\mu})\times\mu^k]/k!$$

Model Assumptions

The following are the assumptions made with this model:

- The outcome variable is distributed Poisson.
- All observations must be independent and identically distributed, or the dependence between observations must be taken into account in the analysis (such as using generalized estimating equations).
- The expected value of the dependent variable $E(Y)$ is equal to the variance of the dependent variable $\mathrm{Var}(Y)$.
- The model fits the data.

Checking Model Assumptions

Independence of observations can be maintained through the use of appropriate study design and data collection, ensuring that observations collected are independent. The fit of the regression model is evaluated through analysis of residuals. Overdispersion can be evaluated by checking the ratios of the goodness-of-fit statistics (deviance and Pearson χ^2) to the degrees of freedom for the analysis. Values much greater than 1 may be indicative of overdispersion.

Advantages and Disadvantages

Poisson regression is most commonly used when the data under study exist as individual counts, such as the number of cases of illness in different communities. Poisson regression, with robust estimation for confidence intervals, is also useful to estimate relative risk directly when the outcome is common.

The defining characteristics of the Poisson distribution can lead to an estimation problem. Poisson distributions by definition have their variance equal to their mean. However, a real data set may have variance in the observed data that is greater than the theoretical variance calculated in the regression model, a condition known as overdispersion, which is indicative of inadequate model fit and may indicate the need to use other modeling techniques, such as log-binomial. Using standard Poisson regression with overdispersed data may result in confidence intervals that are too wide. In this case, it is possible to modify the regression model to incorporate robust error variances into the Poisson regression.

Log-Binomial Regression

Log-binomial regression, similar to logistic regression, models a binomial outcome. However, in log-binomial regression the log of the proportion of interest is modeled, as opposed to the log odds or logit. Since the proportion is directly modeled, the final result from log-binomial regression is a direct measure of the relative risk. Log-binomial regression is, therefore, particularly useful for estimating relative risk when the need to control for multiple confounders exists.

Models and Formulae

Log-binomial regression models the log of the outcome under study as a linear combination of regression parameters and independent factors. For example, a single dichotomous exposure, E, with two confounders, C_1 and C_2, can be expressed as

$$\log(Y) = \beta_0 + \beta_1 E + \beta_2 C_1 + \beta_3 C_2.$$

The adjusted relative risk between exposure and outcome is

$$aRR = e^{\beta_1}.$$

Using the above equation as an example, the probability of a positive outcome ($Y = 1$) can be calculated as

$$P(Y = 1 | X = x_i) = \left[e^{(\beta_0 + \beta_1 E + \beta_2 C_1 + \beta_3 C_2)}\right].$$

Model Assumptions

The following are the assumptions made with this model:

- The outcome variable is binomial.
- All observations must be independent and identically distributed, or the dependence between observations must be taken into account in the analysis (such as using generalized estimating equations).
- The sample size is large or exact methods are used.
- The data fit the model.
- The estimates are within the boundaries of the parameter space.

Advantages and Disadvantages

Because the log proportion is used, as opposed to the logit in logistic regression, direct measurements of

risk proportions and relative risk can be made using log-binomial regression.

While odds can range between 0 and ∞, proportions can only range between 0 and 1. When parameter estimates are at or near the boundary of the parameter space, there is a possibility of the estimates exceeding the limits of a proportion with log-binomial modeling. This indicates a failure of the model to fit the data properly within the bounds of log-binomial regression, possibly requiring modeling with a different technique.

Linear Regression

Linear regression differs from the previously described regression modeling techniques in that it is used to model outcomes that are continuous, as opposed to categorical. The measure of association calculated from linear regression is also different, yielding the correlation coefficient as opposed to the relative risk, rate ratio, or odds ratio. The expected value of Y can be directly determined through the straightforward model.

Models and Formulae

Linear regression directly models the expected value of the dependent variable as a linear combination of regression parameters and independent factors. For example, a single exposure, E, with two confounders, C_1 and C_2, can be expressed as

$$E(Y) = \beta_0 = \beta_1 E + \beta_2 C_1 + \beta_3 C_2.$$

Based on least squares analysis, the correlation between independent and dependent variables can be calculated using Pearson's correlation coefficient and R^2. The measure R^2 provides an estimate of the amount of variation in the data that is explained through the regression line and ranges from 0 (none of the variation is explained by the regression line) to 1 (all data points lie exactly on the regression line).

Model Assumptions

The following are the assumptions made with this model:

- The outcome variable is continuous.
- A predicted value of Y can be calculated for each set of x_is based on the regression line, and each of these predicted values of Y has a defined mean and variance.
- All observations must be independent and identically distributed or the dependence between observations must be taken into account in the analysis (such as using generalized estimating equations).
- The relationship between $E(Y|x_i)$ for all x_i is a straight line function.
- The predicted value of Y calculated for any x_i is normally distributed.
- The variance of the predicted value of Y calculated for any x_i is homoscedastic, meaning that the variance of Y for each x_i is the same.

Checking Model Assumptions

Independence of observations can be maintained through the use of appropriate study design and data collection, ensuring that observations collected are independent. Homoscedasticity can be evaluated by plotting the residuals as a function of the independent variable(s) and observing if the spread of the data points does not widen as the independent variables' values increase. The fit of the regression model, including evaluations of normality, is evaluated through analysis of residuals.

Advantages and Disadvantages

Linear regression is a straightforward technique to perform and to interpret. It is very robust in its ability to handle continuous outcome measures as long a there is one mode that does not fall on an extreme of the parameter space. Given a set of regression coefficients, for any combination of x_i, the expected value of Y can be directly calculated. Also, the level of correlation between dependent and independent variables can be directly determined.

The main disadvantage of linear regression is its inability to study categorical outcomes. As many epidemiologic data are in this form (e.g., persons have the outcome of interest or they do not), the application of linear regression is limited. However, for studies with continuous measures as an outcome (e.g., blood pressure), linear regression is an extremely useful technique.

Model Fit: Influence and Outliers

It is always important to understand the fit of the model to the data. Several questions are posed to determine if the model is a "good" one. Does the

model fit the data overall? Does the overall fit change significantly when factors are added or dropped? Are there cases in the study that exert an undue influence on the final model, and thus the final estimate(s) of the measure of association between exposure(s) and outcome? If so, which are they? What is the association without these influential cases? Are there subgroups of cases that the model does not describe well, and if so, which cases?

Assessing the model fit usually has three components: (1) overall fit, (2) influential cases, and (3) outliers. Most researchers spend substantial time selecting terms for a model before assessing the fit of the model to the data. This is a problematic approach because if the first version of the model does not fit the data, decisions based on this model are flawed. Thus, it is important to determine if the first model reasonably fits the data before making decisions on effect modification and confounding and subsequently estimating the measure of association of interest. Overall fit of the model is typically assessed with some form of a chi-square test comparing the observed data and the expected values based on model estimates. If these are similar, then the model is deemed to fit reasonably well. If not, then the model is not a good fit overall and finding a model that does fit the data is the first order of business.

Assessment of the influence of individual cases on the overall outcome in linear regression is performed by dropping one individual case at a time and refitting the model. Cases that produce the largest difference in the parameters of interest (i.e., those relating to the exposure factors) when dropped are identified as the most important influential effects, which require further investigation. The measure of association is also assessed with these individuals included in the analysis and excluded to examine their influence.

In an analysis modeling categorical outcomes with categorical independent factors, it is possible that many individuals have exactly the same values on exposure, confounders, and modifiers. Thus, when assessing influence instead of dropping individuals one at a time, all individuals with the same characteristics are removed and the model refit with the remaining participants. Again, the difference in estimates with and without the group being evaluated is compared and their influence is estimated by the difference in the parameter estimates of interest between the two models. When there are many individuals with the same characteristics, then the group may be influential simply due to its size. This is not a matter of concern. Only when a small group (e.g., 1, 2, or 3 individuals) exerts strong influence on the model is influence considered a problem.

Outliers are individuals that are not well described by the model. That is, the reported measures of association between exposure and outcome are not typical for the outliers. Outliers often provide interesting information about the association and how it varies. For example, it is possible that effect modification by ethnicity exists but that the sample size in a minority group is not sufficient to estimate it with any precision. When fitting a model to the entire population, members of a minority in the study may thus be identified as outliers. Noting for whom the model does not fit is important for the overall interpretation of the study results. Additionally, outliers may provide information that can lead to further study to understand a more complex relationship between exposures and outcomes.

Data Analysis Example

To illustrate the points made above, an example of quantitative data analysis is presented here. Briefly, it is an examination of the number of cesarean sections occurring in two regions of a state to determine the reason for the higher level of births by cesarean section in one region versus the other.

For reference, Table 1 displays the results of the crude exposure-outcome analysis along analyses stratified on three available covariates—number of comorbidities experienced (categorized as 0 to 1 or 2 or more), whether Medicaid benefits were used (yes or no), and age (16 to 17 years, 18 to 34 years, and 35 to 49 years).

Trying to determine associations across many different stratified analyses can be difficult. According to the stratified analysis discussed above, number of comorbidities and use of Medicaid benefits appear to be effect modifiers, but the level of confounding seen by age is not as clear. To further examine the influence of age, Medicaid use, and number of comorbidities, the data were analyzed using multivariate regression modeling. For purposes of illustration, models were generated using logistic regression, log-binomial regression, Poisson regression, and Poisson regression with a robust error variance to account for overdispersion. Table 2 displays the results of modeling with interaction terms included to account for effect modification by number of comorbidities and use of Medicaid benefits, as well

Table 1 Crude and Stratified Tabular Results for Analysis of Association Between Hospital Location in Two Regions of a State and Deliveries by Cesarean Section

	Region A				*Region B*					
	Cesarean		*Vaginal*		*Cesarean*		*Vaginal*			
Delivery Type	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*OR (95% CI)*	*RR (95% CI)*
Crude	6,918	28.1	17,740	71.9	2,717	17.1	13,147	82.9	1.89 (1.80, 1.98)	1.64 (1.57, 1.70)
Age										
16–17	64	15.5	349	84.5	32	7	424	93	2.43 (1.55, 3.80)	2.21 (1.48, 3.30)
18–34	4,856	26.3	13,631	73.7	2,037	16.1	10,598	83.9	1.85 (1.75, 1.96)	1.63 (1.56, 1.71)
35–49	1,998	34.7	3,760	65.3	648	23.4	2,125	76.6	1.74 (1.57, 1.93)	1.48 (1.38, 1.60)
Medicaid Benefits										
No	1,058	20.5	4,104	79.5	1,836	15.9	9,703	84.1	1.36 (1.25, 1.48)	1.29 (1.20, 1.38)
Yes	5,860	30	13,636	70	881	20.4	3,444	79.6	1.68 (1.55, 1.82)	1.48 (1.39, 1.57)
Comorbidities										
2 or more	1,976	27.5	5,222	72.6	1,005	19.5	414	80.5	1.56 (1.43, 1.70)	1.41 (1.31, 1.50)
0 or 1	4,942	28.3	12,518	71.7	1,712	16	9,007	84	2.08 (1.95, 2.21)	1.77 (1.69, 1.86)

Table 2 Comparative Results of Regression Modeling Examining the Association Between Hospital Location in Two Regions of a State and Deliveries by Cesarean Section

		Regression Type			
		Logistic	*Log-Binomial*	*Poisson*	*Poisson**
Medicaid Benefits	*Comorbidities*	*OR (95% CI)*	*RR (95% CI)*	*RR (95% CI)*	*RR (95% CI)*
No	0 or 1	1.83 (1.68, 2.00)	1.58 (1.48, 1.69)	1.58 (1.47, 1.71)	1.58 (1.48, 1.69)
No	2 or more	1.40 (1.25, 1.56)	1.27 (1.16, 1.38)	1.27 (1.15, 1.40)	1.27 (1.17, 1.38)
Yes	0 or 1	1.53 (1.40, 1.68)	1.41 (1.31, 1.53)	1.42 (1.30, 1.54)	1.42 (1.31, 1.53)
Yes	2 or more	1.17 (1.05, 1.30)	1.13 (1.04, 1.23)	1.14 (1.03, 1.25)	1.14 (1.04, 1.24)

Notes: Analyses include controlling for confounding by age. Asterisk (*) indicates Poisson regression with robust error variance.

as controlling for confounding by age. Table 3 displays the results of modeling with interaction terms included to account for effect modification by number of comorbidities and use of Medicaid benefits, with no other covariates controlled for.

Results for log-binomial regression, Poisson regression, and Poisson regression with robust error variance (Poisson* in the table) are very similar to each other. Overestimates of variance seen in Poisson regression are not large, but are accounted for when Poisson regression with robust error variance is used.

The odds ratio values obtained from logistic regression modeling are similar to odds ratios generated from multiple stratified tabular analysis, just as the relative risk values obtained from log-binomial, Poisson, and Poisson with robust error variance

Table 3 Comparative Results of Regression Modeling Examining the Association Between Hospital Location in Two Regions of a State and Deliveries by Cesarean Section

		Regression Type			
		Logistic	*Log-Binomial*	*Poisson*	*Poisson**
Medicaid Benefits	*Comorbidities*	*OR (95% CI)*	*RR (95% CI)*	*RR (95% CI)*	*RR (95% CI)*
No	0 or 1	1.82 (1.67, 1.98)	1.58 (1.47, 1.69)	1.58 (1.46, 1.70)	1.58 (1.47, 1.69)
No	2 or more	1.40 (1.25, 1.56)	1.27 (1.17, 1.38)	1.27 (1.15, 1.40)	1.27 (1.17, 1.38)
Yes	0 or 1	1.50 (1.37, 1.65)	1.39 (1.29, 1.50)	1.39 (1.28, 1.52)	1.39 (1.29, 1.50)
Yes	2 or more	1.15 (1.03, 1.28)	1.12 (1.03, 1.22)	1.12 (1.02, 1.23)	1.12 (1.03, 1.22)

Notes: Analyses do not include controlling for confounding by age. Asterisk (*) indicates Poisson regression with robust error variance.

regression modeling are similar to relative risks generated from multiple stratified tabular analysis. The results suggest that age does not need to be adjusted for in the analysis as the results are similar with and without adjusting for age. It should also be noted that logistic regression provides odds ratio estimates that overestimate relative risks (or prevalence ratios in this case). Thus, logistic regression should not be used for this analysis as the outcome (cesarean section) occurs too frequently for the odds ratio to be a reasonable estimate of the prevalence ratio.

—Robert Bednarczyk and Louise-Anne McNutt

See also Causation and Causal Inference; Effect Modification and Interaction; Logistic Regression; Study Design

Further Readings

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Barros, A. J. D., & Hirakata, V. N. (2003). Alternatives for logistic regression in cross-sectional studies: An empirical comparison of models that directly estimate the prevalence ratio [Electronic version]. *BMC Medical Research Methodology*, *3*(21). Retrieved August 8, 2007, from http://www.biomedcentral.com/content/pdf/1471-2288-3-21.pdf.

Hardin, J. W., & Hilbe, J. M. (2002). *Generalized estimating equations*. Boca Raton, FL: Chapman & Hall.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Kleinbaum, D. G., & Klein, M. (2002). *Logistic regression: A self-learning text* (2nd ed.). New York: Springer.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1998). *Applied regression analysis and other multivariable methods* (3rd ed.). Pacific Grove, CA: Duxbury Press.

McNamee, R. (2005). Regression modeling and other methods to control confounding. *Journal of Occupational and Environmental Medicine*, *62*, 500–506.

McNutt, L. A., Wu, C., Xue, X., & Hafner, J. P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology*, *157*, 940–943.

Rothman, K. J., & Greenland. S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

SAS FAQ. (n.d.). *How to estimate relative risk in SAS using Proc Genmod for common outcomes in cohort studies?* Retrieved March 8, 2007, from http://www.ats.ucla.edu/STAT/SAS/faq/relative_risk.htm.

Spiegelman, D., & Hertzmark, E. (2005). Easy SAS calculations for risk of prevalence ratios and differences. *American Journal of Epidemiology*, *162*(3), 199–200.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical data analysis using the SAS system* (2nd ed.). Cary, NC: SAS Institute.

Wacholder, S. (1986). Binomial regression in GLIM: Estimating risk ratios and risk differences. *American Journal of Epidemiology*, *123*(1), 174184.

Web Sites

LogXact Overview: http://www.cytel.com/Products/LogXact.

Relational Database

By far the most common use for computers during most of their existence has been to create and store

databases. From large databases that manage a bank's account information to the common e-mail program, databases are the engines behind most of the software in use today. In simple terms, a database is computer software that contains organized data. The data are structured to allow a user to search for specific data, reorder the data, and create reports containing specified parts of the data. For the epidemiologist, a relational database can provide a tool to manage and maintain large data sets, create reports, and prepare basic statistical analyses.

In the earliest mainframe computers, databases were complex to create and maintain and remained the province of trained professionals. When the desktop or personal computer appeared in the 1980s, database software was introduced that allowed individuals without a background in computer science to create and use databases. Ashton Tate's dBASE was the first major commercial database software to be widely used on those early desktop computers. With the evolution of the graphic interface within Microsoft's Windows operating system and Apple's Macintosh computer, it became even simpler for noncomputer professionals to create their own database systems. Currently, the two dominant database systems on desktop computers are Microsoft Access and Filemaker Pro. These two programs offer both the database software and integrated tools to create data displays and reports and allow users to write basic computer programs necessary for data management, even for users with little or no experience in database design.

A structure of a database is referred to as a "schema." The basic structure is made up of *records*, each of which contains *fields.* To use the analogy of a patient's medical form, each form containing information about an individual patient is a record. The data on the form are contained within different fields, such as name, address, age, and gender. The database itself is comparable with a file cabinet that contains all the patient records. In this type of database, referred to as a "flat file," all the data are self-contained and could just as well be maintained on a spreadsheet as on a relational database.

The limitation of the flat file is that it may be inconvenient to have all the information about a given patient in a single flat file. For instance, you may want to keep a record of each of a patient's visits, to record specific information about that visit (such as blood pressure, temperature, height, and weight), to add a new field to a single flat file for each of these variables on each visit would be awkward. Neither do you want to reenter basic information about patients, such as their age and insurance company, each time they have an office visit. Another reason to not store all information in a single flat file is that certain information needs to be kept confidential. For instance, you would not want information about a patient's HIV status to be accessible to a staff member who needs to use the patient file only to perform billing operations. By using a *relational* database model, multiple databases, also referred to as "tables," can be linked so that different types of information can be entered into different tables, yet all the information about a single record, for instance, a particular patient, is linked and can be combined to create different reports. For instance, you might have one table that records basic demographic and contact information for each patient (age, home address, etc.) and a second table that records information about individual patient visits. Both would be linked by an identification number unique to a particular patient. This is referred to as a "one-to-many" relationship, because one patient record may be linked to multiple visit records. The visits database can display the patient information from the patient file, and because this information is linked rather than reentered, the likelihood of error is reduced. By extension, other tables can be created containing data about drugs prescribed, treatments administered, insurance payments, and so on. These tables can all be linked to and can use data from the original patient information table.

	A	B	C	D	E
1	**Dr. Lionel Schmerz**				
2					
3	Patient Records				
4					
5	**Patient _ID**	**DOB**	**Gender**	**1st Visit**	
6	1	04-May-56	F	01-Aug-05	
7	2	15-Mar-53	M	03-Sep-05	
8	3	27-Mar-53	M	10-Oct-04	
9	4	17-Jun-90	F	08-May-05	
10	5	20-Dec-63	F	08-Jan-05	
11	6	09-Jun-42	M	09-Jul-04	
12	7	08-Apr-62	F	07-Sep-05	
13	8	17-Mar-59	M	28-Jan-05	
14	9	08-May-62	F	19-Aug-05	
15	10	05-Nov-68	F	08-Dec-05	

Figure 1 Spreadsheet With Patient Records

Note: In this spreadsheet, each row contains data for a specific patient, and each column contains data of a specific type.

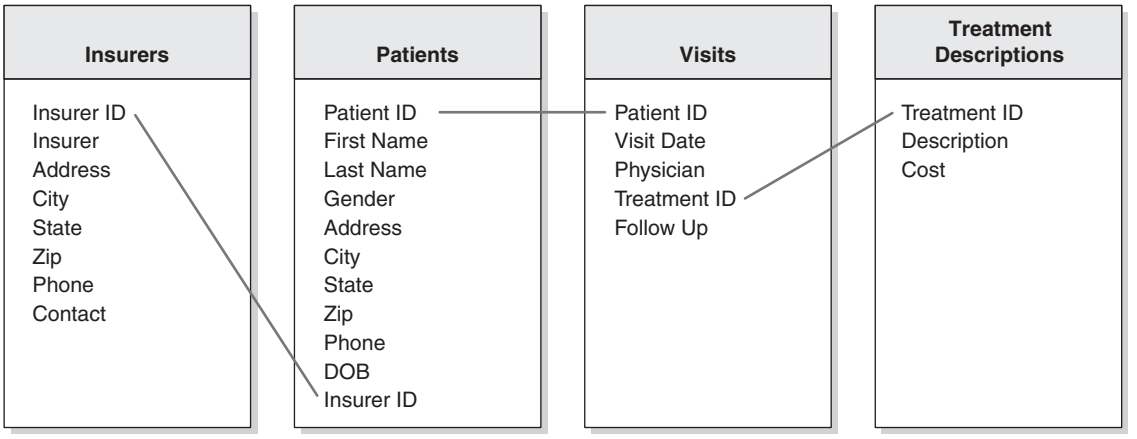


Figure 2 Database Structure

Related tables are linked through the use of a *key field.* The key field contains data unique to an individual record, a patient ID number, for example. When a new record is created in a related table, the key field data are entered in that record and provide the link back to the appropriate record in the first table.

When designing a database, particular attention must be paid to creating a proper structure from the outset. For example, names should always be separated (parsed) and entered in two fields (for first name and last name) rather than as a single name field. With a single name field, it would be impossible to sort (reorder) the data by last name, a serious limitation. Since it is difficult to restructure data after it has been entered, it is important that the structure be correct from the start.

Relational database software has many capabilities beyond just storing and displaying information. For instance, queries may be performed that allow the user to create customized reports, and built-in functions allow the computation of many basic statistics. Charts and form letters can also be produced directly from information stored in the database Access and Filemaker Pro that have features that allow a user to create data displays that incorporate both data and graphics elements. Data entry forms can be created with elements such as drop-down lists or check boxes in a particular field to speed up data entry and minimize entry errors. Relational database software also includes data-processing functions similar to that available in spreadsheets that can calculate and display arithmetic and statistical results.

Database software can import data from a wide variety of sources, so data can be added easily to an existing system. The data can then be searched, reordered (sorted), and extracted for use in other programs. For example, it would be simple to search a database of patient records to identify all those pertaining to women above 50 years of age, and then export those records (with data from multiple tables) for analysis in a dedicated statistics package such as SAS.

Multiuser databases allow multiple users to use the databases simultaneously, so that one user can enter new records while another searches the data and creates reports, all working simultaneously on the same set of tables. Multiuser capabilities can be limited to just the internal network of an office or extended to the entire world via the Internet, greatly facilitating multisite research projects.

Multiuser databases must be configured with certain security issues in mind, especially when they contain confidential information such as patient records or Social Security numbers. Through the use of account names and passwords, each user can be allotted specific privileges so that one user may only view data but not change it, while another can only enter new data but not change existing entries. Security settings can also prevent a specific user from viewing data in certain fields while allowing them to work on others.

—Daniel Peck

See also Data Management; Spreadsheet

Further Readings

Chase, K. (2005). *Access 2003 for starters: The missing manual*. Sebastapol, CA: O'Reilly Media.

Hernandez., M. J. (2003). *Database design for mere mortals: A hands-on guide to relational database design*. Upper Saddle River, NJ: Addison-Wesley.

Hester, N. (2006). *Filemaker Pro 8: Visual QuickStart guide*. Berkeley, CA: Peachpit Press.

Reliability

The issue of reliability (i.e., repeatability or reproducibility) is crucial in selecting and developing the most appropriate item, scale, or instrument. Reliability refers to the extent to which an instrument or the measurements of a test will consistently produce the same result, measure, or score if applied two or more times under identical conditions. A technique is reliable, or has achieved a high level of agreement, if it yields consistent results on repetition. If repeated measurements produce different results, and the entity being measured is assumed to not have changed, the instrument would be considered unreliable.

Methods of Assessing or Estimating Reliability

There are a variety of methods for estimating instrument reliability. DeVellis classifies these methods into two categories: (1) the type of instrument (observer or external source vs. self-report) and (2) time instrument applied or method (single administration or multiple administration). Reliability is estimated in one of four ways:

1. *Internal Consistency.* This estimation is based on the correlation among the variables comprising the set or the homogeneity of the items comprising a scale (usually estimated with Cronbach's alpha).
2. *Split-Half Reliability.* This estimation is based on the correlation of two equivalent forms of the scale (usually estimated with the Spearman-Brown coefficient).
3. *Test-Retest Reliability.* This estimation is based on the correlation between scores from two (or more) administrations of the same item, scale, or instrument for different times, locations, or populations, when the two administrations do not differ in other relevant variables (usually estimated with the Spearman-Brown coefficient).
4. *Interrater Reliability.* This estimation is based on the correlation of scores between/among two or more raters who rate the same item, scale, or instrument (usually estimated with intraclass correlation, of which there are six types discussed below).

These four reliability estimation methods are sensitive to different sources of error and are not necessarily mutually exclusive. Therefore, the reliability scores measured using these methods should not be expected to be equal nor need they lead to the same results. All reliability coefficients are forms of correlation coefficients and are thus sample dependent. In other words, another sample may well result in a different estimate.

Internal Consistency Reliability

Cronbach's coefficient alpha is the classic form of internal consistency reliability and is widely used as a measure of reliability. Cronbach's alpha can be interpreted as a measure of mean intercorrelation among item responses obtained at the same time. Cronbach's alpha is influenced by the number of items in a scale, so alpha will increase as the number of items in the scale increases, if new items have the same average intercorrelation of items. There are no absolute standards for knowing when reliability is adequate, but an often-used rule of thumb is that alpha should be at least .70 for a scale to be considered adequate, and many researchers require a cutoff of .80 for a "good scale."

Cronbach's α is defined as

$$\frac{N}{N-1}\left(\frac{\sigma_X^2 - \sum_{i=1}^{N}\sigma_{Y_i}^2}{\sigma_X^2}\right),$$

where N is the number of items, σ_X^2 is the variance of the observed measure, and $\sigma_{Y_i}^2$ is the variance of sum of the items.

Cronbach's α is closely related to the correlation among items, and when evaluating whether an individual item should be retained in a scale, it is good to look at the squared multiple correlation, R^2 for an item when it is predicted from all other items in the scale. The larger this R^2, the more the item is contributing to internal consistency. The lower the R^2, the more the researcher should consider dropping it. Note

that a scale with an acceptable overall Cronbach's α may have some items with a low R^2. *The Kuder-Richardson (KR20) coefficient* is a special version of Cronbach's α for items that are dichotomous.

Split-Half Reliability

Split-half reliability measures equivalence among two measurement instruments or between two halves of the same instrument. It is related to the concept of parallel-forms reliability, in which two different measurement instruments, which are assumed to be equivalent, are administered twice to the same people. Spearman-Brown split-half reliability coefficient, also called the *Spearman-Brown prophecy coefficient*, is used to estimate full test reliability based on split-half reliability measures. The Spearman-Brown "prophecy formula" predicts what the full-test reliability would be, based on half-test correlations. This coefficient will be higher than the half-test reliability coefficient. This coefficient is usually equal to and easily calculated by hand as twice the half-test correlation divided by the quantity 1 plus the half-test reliability.

$$r_{SB1} = (k \times r_{ij})/[1 + (k - 1) \times r_{ij})],$$

where

r_{SB1} = the Spearman-Brown split-half reliability,

r_{ij} = the Pearson correlation between forms i and j, and

k = total sample size divided by sample size per form (k is usually 2).

As with other split-halves measures, the Spearman-Brown reliability coefficient is highly influenced by alternative methods of sorting items into the two forms, which is preferably done randomly. Random assignment of items to the two forms should ensure equality of variances between the forms, but this is not guaranteed and should be checked by the researcher.

Test-Retest Reliability

Test-retest reliability, which measures the temporal stability or stability over time, is administering the same test to the same subjects at two points in time. In other words, test-retest reliability is replicating the measurement and computing the correlation coefficient to see how constant the scores remain from one occasion to another. Statistically, test-retest reliability is treated as a variant of split-half reliability and also uses the Spearman-Brown coefficient.

Test-retest methods, as an appropriate way of gauging reliability, are subject to several restrictions. Among these restrictions are the following: (1) it must be assumed that the underlying phenomenon or the true score has not changed; (2) if the scale itself is unreliable for a single administration, test-retest reliability cannot be evaluated; and (3) there are no carry-over effects, that is, the scores from the second episode of testing are not influenced by the subject's memory (or physical traces, such as drugs remaining in the bloodstream) from the first episode. Researchers using test-retest reliability must weigh and understand the special validity concerns to make informed judgments when designing a measurement or evaluative study.

Interrater Reliability

Interrater reliability is a method of determining how well raters agree in their judgment of some event; it is often reviewed to evaluate agreement among several individuals assigned to review medical charts, for instance. Interrater reliability is evaluated by having two or more raters or interviewers administer the same form to the same people or evaluate the same objects (such as medical charts) to establish the extent of consensus on use of the instrument by those who administer it. Raters should be as blind as possible to expected outcomes of the study and should be randomly assigned. There are different ways to calculate interrater agreement; for dichotomous items, the most common choices are simple percent agreement and kappa, which is percent agreement corrected for the amount of agreement expected by chance alone. For continuous data, consensus is measured by *intraclass correlation* (ICC); note that the term *ICC* is sometimes applied to percent agreement and kappa also.

The guidelines for choosing the appropriate form of the ICC varies depending on whether a one-way or two-way analysis of variance (ANOVA) is suitable, whether the differences between the judges' mean ratings are relevant, and whether reliability is to be measured based on an individual rating or the mean of several ratings. ICC may be conceptualized as the ratio of between-groups variance to total variance and is interpreted similarly to Kappa. Shrout and Fleiss (1979) have defined six kinds of ICC:

1. *ICC(1,1).* Used when each subject is rated by multiple raters, raters assumed to be randomly assigned to subjects, all subjects have the same number of raters; one-way (random targets are the grouping variable) single measure reliability.
2. *ICC(2,1).* Used when all subjects are rated by the same raters, who are assumed to be a random subset of all possible raters; a two-way random effects model single measure reliability.
3. *ICC(3,1).* Used when all subjects are rated by the same raters, who are assumed to be the entire population of raters; two-way mixed effects model single measure reliability.
4. *ICC(1,k).* Same assumptions as for ICC(1,1) but reliability is for the mean of *k* ratings; one-way model single and average measure reliability.
5. *ICC(2,k).* Same assumptions as for ICC(2,1) but reliability is for the mean of *k* ratings; a two-way random effects model average measure reliability.
6. *ICC(3,k).* Same assumptions as for ICC(3,1) but reliability is for the mean of *k* ratings. This additionally assumes no subject by judges interaction; the two-way mixed effects model average measure reliability.

—Kevin Robinson

See also Bias; Item Response Theory; Kappa; Response Rate; Validity

Further Readings

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* London: Sage.
Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.
DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
Nunnaly, J. (1978). *Psychometric theory*. New York: McGraw-Hill.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *2*, 420–428.

Reproductive Epidemiology

Reproductive epidemiology is the study of reproduction-related morbidity, mortality, and other health issues in males and females. The topics covered in reproductive epidemiology include development and physiology of reproductive systems and functions, conception, pregnancy, birth outcomes, and maternal morbidity and mortality.

Measures of Reproductive Health

The number of measures of reproductive health is substantial. Several select indicators commonly used in reproductive epidemiologic studies are described below.

Maternal Mortality

Maternal mortality is defined by the World Health Organization (WHO) as the death of a woman during pregnancy or within 6 weeks of termination of pregnancy from any cause related to or aggravated by the pregnancy or its management. The causes of the death can be categorized into direct and indirect obstetric deaths. Direct obstetric death is caused by complications of pregnancy, delivery, or the puerperium (the period immediately after childbirth and lasting about 6 weeks, during which the mother's body returns to its prepregnant condition). The five major causes of direct obstetric deaths worldwide are hemorrhage, complications of unsafe abortion, eclampsia, infection, and obstructed labor. Indirect obstetric death results from previously existing conditions or conditions physiologically aggravated by the pregnancy. Common examples of such conditions are malaria, anemia, HIV/AIDS, and cardiovascular disease. Nonobstetric deaths include other deaths during but not caused by the pregnancy, such as those caused by accidents or by intentional acts not caused directly by the pregnancy (e.g., murder). According to a report jointly prepared by the WHO, UNICEF (United Nations Children's Fund), and UNFPA (United Nations Population Fund), there were 529,000 maternal deaths in 2000, of which more than 99.5% occurred in developing regions. Three commonly used measures related to maternal mortality are the maternal mortality rate, maternal mortality ratio, and lifetime risk of maternal death.

Maternal Mortality Rate

The maternal mortality rate is calculated as the number of maternal deaths in a given period per 1,000 women of reproductive age (usually 15–49

years of age) during the same time period and reflects the frequency with which women are exposed to mortality risk through fertility.

Maternal Mortality Ratio

This is a measure of the risk of death associated with pregnancy. It is calculated as the number of maternal deaths during a given time period per 100,000 live births during the same time period. In other words, the numerator is the number of maternal deaths multiplied by 100,000, and the denominator is the number of live births. This measure is often referred to as a rate, although it is really a ratio.

Lifetime Risk of Maternal Death

The lifetime risk of maternal death is the probability that a woman will die from complications of pregnancy or childbirth at some point during her reproductive years. It is a cumulative risk across a woman's reproductive years and is often used as an index of risk faced by women in developed and developing countries.

Infant Mortality

The infant mortality rate (IMR) is defined as the rate per 1,000 live births at which babies less than 1 year of age die. It is calculated by dividing the number of infant deaths in a given year by the number of live births in the same given year. The IMR is often used to compare the general health and well-being of populations within and between countries and is sometimes considered a proxy or indicator of the quality of health care available to the relevant population. Comparing different countries' IMRs can sometimes be difficult when different definitions of "live birth" are employed. For example, the WHO defines a live birth as any born human being who demonstrates independent signs of life, including breathing, voluntary muscle movement, or heartbeat. Some European states and Japan, however, only count as live births those in which an infant breathes at birth, thereby causing their IMRs to be somewhat lower and their perinatal mortality rates to be somewhat higher than in settings using other definitions. Excluding high-risk infants from the denominator or numerator in reported IMRs also makes comparing rates problematic.

The IMR has declined steadily over the past several decades in the United States, from a national average of 26.0 per 1,000 live births in 1960 to 6.9 per 1,000 live births in 2000, but large racial and ethnic disparities persist. In 2000, the IMR among whites in the United States was 5.7 per 1,000 live births, compared with 14.1 per 1,000 live births for African Americans. Reducing the IMR overall and closing gaps between white and minority IMRs are national objectives put forth in *Healthy People 2010*. In the United States and other western nations, common causes of infant death include congenital malformations, preterm birth and low birthweight, sudden infant death syndrome, problems related to pregnancy complications, and respiratory distress syndrome.

Differences in IMRs are also evident among developed versus developing countries. The United Nations estimated the 2000 IMR among developed countries to be 8 per 1,000 live births, compared with 62 per 1,000 live births for less developed countries. The causes of IMR in developing countries tend to be different as well and include infectious disease, communicable disease, and dehydration.

Pregnancy Outcomes

Low Birthweight

Low birthweight is typically divided into three categories: low birthweight, very low birthweight, and extremely low birthweight. Low birthweight is defined as weight at birth of $\leq$2,500 g (5.5 lb), very low birthweight refers to babies born weighing $\leq$1,500 g, and extremely low birthweight is defined as weight at birth of $\leq$1,000 g. It is estimated that in 2000, more than 20 million infants (approximately, 15.5% of all live births worldwide) were low birthweight. Prevalence of low birthweight varies substantially by countries' development status. The prevalence of low birthweight is approximately 7.0% in developed countries, 16.5% in developing countries, and 18.6% in the least developed countries. Low birthweight is associated with child physical growth and psychosocial development and with chronic medical conditions later in life. Measurement error is not a major concern for birthweight in developed countries because birthweight can be measured accurately; however, it can be of substantial concern in developing countries since most babies are not born in a medical setting and therefore not often weighed at birth.

Pregnancy Loss

Pregnancy loss refers to the loss of a pregnancy before birth and may occur through miscarriage (sometimes referred to as "spontaneous abortion"), termination, or stillbirth. Early pregnancy loss occurs prior to 20 weeks of completed gestation, before a fetus can survive outside the womb. Miscarriage occurs in approximately 15% to 20% of all pregnancies, most commonly within the first 13 weeks of pregnancy. Some miscarriages occur before a woman even realizes she is pregnant, even before missing a menstrual period, so the true number of miscarriages is probably significantly underestimated. The cause of miscarriage is frequently unknown, but chromosomal abnormalities in the fetus, maternal health problems such as infections (e.g., bacterial vaginosis) or chronic disease (e.g., diabetes or lupus), maternal lifestyle behaviors (e.g., smoking, substance use), or uterine impairments can be contributing factors. Chromosomal abnormalities are more common among women above 35 years of age, which places these women at higher risk of miscarriage compared with younger women.

Infertility

Infertility refers to the phenomenon of couples who try to conceive but fail to have a pregnancy for more than a year. Rather than the number of live births, infertility denotes reproductive capacity. Studies of infertility need to be cautious of case ascertainment as infertility diagnosis is prone to selection bias. For instance, women or couples who seek infertility care may differ from those who choose not to try to conceive (and may not ever realize they are infertile) or who cannot afford to obtain infertility diagnosis or treatment. Couples with infertility problems may have no apparent clinical symptoms and may appear to be healthy otherwise. Psychosocial factors and personal choices also can be significant influences and should be considered when measuring infertility.

Offspring Morbidity

The survival rate of newborns, including infants with low birthweight and those born as a result of assisted reproductive technology, has increased dramatically over the past two decades. This is particularly true in developed countries. The increased survival of newborns with some medical conditions (e.g., birth defects, extremely low birthweight) may lead to a higher prevalence of infant morbidity because those infants are at high risk for morbidity and mortality, and previously they would not have survived the birth process.

Some researchers have studied whether later health outcomes are influenced in utero. David Barker was one of the first to propose these notions in what is known as "Barker theory" or the "fetal origins hypothesis." Barker hypothesized that biophysiologic programming occurs at certain critical periods of fetal development, thereby strongly influencing health in later life. Since then, many studies have tested the programming hypothesis, exploring causal relationships between fetal and childhood exposures and adult chronic disease. Barker's hypothesis could help explain some of the socioeconomic and racial disparities in chronic illnesses such as cardiovascular disease and hypertension, although lifestyle and other environmental influences may also play a role.

Some researchers have critiqued studies based on fetal programming theory, citing methodological concerns such as selection bias, failure to assess and control for potential confounding, ecologic fallacy, and so on. Recent studies, however, have continued to find evidence supportive of the association between exposures and experiences in utero with adult blood pressure and certain cause-specific mortalities. Research based on fetal programming theory can be difficult to conduct due to challenges such as recall bias related to exposure, gene and environment interaction, expense of long-term follow-up, loss to follow-up, and determination of causality.

Health Disparities

Although some of the disparities in reproductive health are related to biological differences between populations, many are due to inequality in health care availability, social and cultural issues, and differences in lifestyle. The following are some of the areas in which health disparities are often addressed through reproductive epidemiologic studies.

Preventive Care

Preventive services in reproductive health include education, screening, treatment for sexually transmitted diseases (STDs) and sexually transmitted infections (STIs), preconceptional and prenatal care, and

counseling concerning lifestyle factors such as smoking and drinking alcohol. Preconceptional and early comprehensive prenatal care can reduce the risk of pregnancy- and birth-related complications. Contraceptive practice can lengthen intervals between pregnancies to protect the health of the mother and her children. Despite its importance, the relation between the number of prenatal care visits and pregnancy outcomes is not linear. Research evidence suggests that receiving less (or no) and more than the recommended number of prenatal care visits all are associated with poor pregnancy outcomes. Receiving more than the recommended amount of prenatal care could reflect a problematic pregnancy, whereas receiving less or no prenatal care is likely due to barriers or lack of knowledge of its importance. Some common barriers a woman and her family (especially her partner) face in obtaining appropriate preventive care services include health care unavailability, lack of health insurance and underinsurance, lack of transportation, and lengthy waiting time for care. As with many other chronic conditions, unhealthy lifestyle (e.g., substance abuse, obesity) is associated with many adverse reproductive health outcomes in women and men. Education and policy can play important roles in promoting healthy behaviors, which in turn improves reproductive health.

Infectious Diseases

STIs or STDs are those that are typically transmitted between people by sexual contact, such as vaginal or anal intercourse, or oral sex. Other possible routes of transmission include birth, breastfeeding, blood transfusions, or sharing intravenous needles. Many people with STIs, especially women, are asymptomatic and unaware of their condition but are still able to spread infection. For women, regular Papanicolaou (Pap) screening can help identify asymptomatic and symptomatic STIs, allowing treatment that could help prevent serious complications, such as pelvic inflammatory disease and infertility.

The incidence of STIs is high in most of the world, despite the existence of effective protection (e.g., condoms), diagnosis, and treatment. In the United States, more than 2.8 million new cases of chlamydia were diagnosed in 2005, the rate of gonorrhea was 115.6 per 100,000, and the syphilis rate increased 11.1% between 2004 and 2005. As is the case with many reproductive health issues, socioeconomic and racial/ethnic disparities are evident in STI infection rates. For various reasons, funding for STI prevention and treatment is insufficient, and in many parts of the world, frank discussion of issues related to sexual behavior is not common.

Some STIs, such as chlamydia and gonorrhea, are easily treated with antibiotics, while others, such as herpes and HIV/AIDS, either cannot be cured currently or are difficult to treat. Recently, advances have been made in developing a prophylactic vaccine for females and males to protect against human papillomavirus, the virus that is responsible for the majority of cervical cancer cases.

Sociocultural Factors

Although women and men have the right to determine the course of their reproductive lives, numerous sociocultural factors may prevent them from being able to do so. These factors vary across different countries or regions and may differ within the same country. Politics and religion may influence how an individual experiences or manages his or her reproductive health. Other factors, such as insurance coverage for access to birth control and emergency contraception, pregnancy termination or abortion, and general reproductive health care, also play influential roles.

Nutrition

Evidence indicates that maternal excess body mass and body fat are associated with menstrual disorders, infertility, pregnancy complications, and pregnancy outcomes. Currently, more than 60% of U.S. women of childbearing age are either obese or overweight. In 2003, there were about 4 million live births in the United States. This suggests that more than 2 million children were born to mothers who were obese or overweight in the United States in 2003 alone. Researchers have begun to investigate pregnancy as a catalyst for maternal obesity after childbirth; however, the findings are inconclusive. Given the current obesity epidemic observed in developed countries, further investigation of obesity-related reproductive health problems is greatly needed. Findings from this research will inform the public and health providers and guide the development of intervention strategies and effective prevention programs. In contrast, many poor maternal health conditions and adverse birth

outcomes in developing countries are more frequently due to poor or undernutrition. Examples of these health conditions and birth outcomes include hypertensive disorders, anemia and infection during pregnancy, low birthweight, maternal depletion, and neural tube defects. When body mass index (BMI: weight in kilograms divided by height in meters squared) is used as an indicator or proxy of nutrition, both underweight (< 18.5 BMI) and obese (≥ 30 BMI) status are risk factors for poor maternal health conditions and birth outcomes. A great disparity exists between the health and nutritional status of populations living in the developing and developed world. For example, low BMI (< 18.5), a known predictor of poor pregnancy outcome, is prevalent in 34% of women of childbearing age in south Asia and in 18% of women of childbearing age in sub-Saharan Africa in contrast to a prevalence of 4% in women of childbearing age from the developed world.

Large disparities in many reproductive health outcomes continue to exist between disadvantaged and higher socioeconomic groups within the same country, between different racial and ethnic populations, and between lower- and higher-resource countries. Most of the disparities can be attributed to poverty, lack of education, and ineffective or nonexistent policies. To eliminate the inequities, collaborative efforts from communities, governmental agencies, and the global society are needed.

Key Study Design and Measurement Issues

As with research on most rare diseases, researchers exploring uncommon reproductive conditions may consider adopting a case-control study design to obtain a satisfactory sample size that would permit detection of an effect in a more efficient way with given resources and time. However, information bias (e.g., recall bias) and measurement errors can be a concern. Case-case study design can be carried out in the situation in which data were collected for cases only. For example, the Autism Genetic Resource Exchange is a large collaborative gene bank for which biosamples are collected only from families of children with autism. One can analyze the data by comparing cases with and without a perinatal exposure of interest and their autism subtype diagnosis or genetic traits. The prospective cohort study design is ideal for studying the effect of an early exposure that does not manifest until later in life. This approach, however, can be very costly and impractical. One way to handle the problems of prospective cohort studies, such as long follow-up time and low occurrence of cases or events, is to select a cohort that is at high risk of disease. For example, autism is one of the most heritable neuropsychiatric disorders. To investigate whether perinatal suboptimality is related to autism, one can recruit a cohort that consists of mothers who have a child with autism and are planning or intending to have another pregnancy in the near future.

Some methodologic considerations that occur more often in reproductive epidemiologic research than in chronic or infectious disease research are worthy of mention. First, some reproductive health studies may need to consider the couple as a unit rather than as two individuals. For example, the decision to conceive, ideally, involves decisions made by both partners. Second, the heterogeneity in risk of the individuals comprising the couple needs to be considered; focus should not be given only to the female, for instance. Third, researchers must remember that pregnancy outcomes can be competing. For example, a research interest is to investigate a specific exposure and infant mortality. Epidemiologists need to recognize that pregnancy outcomes, such as early pregnancy loss, spontaneous abortion, miscarriage, stillbirth, neonatal, postneonatal, and infant mortality compete with one another, and to bear in mind that a harmful exposure artificially observed as "protective" for infant mortality is not addressed if early pregnancy loss or spontaneous abortion occurred beforehand.

Public Health Implications

Reproductive epidemiologic research has made great contributions in terms of informing treatment approaches and policy making aimed at improving human reproductive health. A noteworthy example is periconceptional folic acid supplementation to prevent neural tube defects. While progress has been made in many areas, including advances in tools for measuring environmental toxicants and genotyping in reproductive health research, disparity issues remain and must be addressed. Some of the reproductive health disparities that persist are those between men and women,

between populations in developing versus developed countries, between wealthy and impoverished persons, and across racial and ethnic groups.

The breadth of reproductive epidemiologic research needs to expand to understand etiologic risks more fully. Instead of solely measuring perinatal factors and immediate birth outcomes, the expansion should address the time prior to the periconception period and assess long-term maternal and child health impacts that occur later in life. Because reproduction involves both males and females, both male and female reproductive function warrants continuous investigation.

Some reproductive health issues are influenced heavily by politics (e.g., contraception and abortion), some are more influenced by familial and cultural factors (e.g., intimate partner violence), while others depend on the individual's financial status, access to health care, and reproductive history (e.g., assisted reproductive technology). As a consequence, the social and biologic context of reproduction will sustain its epidemiologic study.

—Li-Ching Lee, Deborah L. Dee, and Amy Tsui

See also Birth Defects; Fertility, Measures of; Fetal Death, Measures of; Gestational Age; Maternal and Child Health Epidemiology; Newborn screening programs; Oral Contraceptives; Preterm Birth; Sexually Transmitted Diseases

Further Readings

Centers for Disease Control and Prevention. (2007). *Trends in reportable sexually transmitted diseases in the United States, 2005: National surveillance data for chlamydia, gonorrhea, and syphilis*. Retrieved February 12, 2007, from http://www.cdc.gov/std/stats/trends2005.htm.

Maternal Mortality in 2000: Estimates developed by WHO, UNICEF and UNFPA. (2004). *Department of reproductive health and research World Health Organization.* Geneva, Switzerland: WHO.

National Center for Health Statistics. (2002). Infant, neonatal, and post neonatal mortality rates by race and sex: United States, 1940, 1950, 1960; 1970, and 1975-2000 (Table 34). *National Vital Statistics Report*, *50*(15), 100–101. Retrieved February 12, 2007, from http://www.cdc.gov/nchs/fastats/pdf/nvsr50_15tb34.pdf.

Walboomers, J. M., Jacobs, M. V., Manos, M. M., Bosch, F. X., Kummer, J. A., Shah, K. V. et al. (1999). Papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology*, *189*, 12–19.

Weinberg, C. R., & Wilcox, A. J. (1998). Reproductive epidemiology. In J. R. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 585–608). Philadelphia: Lippincott-Raven.

Winer, R. L., Hughes, J. P., Feng, Q., O'Reilly, S., Kiviat, N. B., Holmes, K. K., et al. (2006). Condom use and the risk of genital human papillomavirus infection in young women. *New England Journal of Medicine*, *354*, 2645–2654.

Web Sites

World Health Organization, Sexual and Reproductive Health: http://www.who.int/reproductive-health/index.htm.

Response Rate

A study's response rate is an important gauge for the quality of data collection. The response rate, in its most basic form, refers to the proportion of people eligible for a study who actually enroll and participate. In fact, despite the name, the response rate is a proportion rather than a rate.

Although the concept is simple, the computation of response rates can be complex, and the use of multiple formulas diminishes the ability to compare studies by degree of nonresponse. Any comparison of response rates between studies requires knowledge of the study designs, sampling frames, modes of study recruitment, and formulas for computing response rates.

The American Association for Public Opinion Research (AAPOR) has attempted to standardize response rates for surveys conducted by mail or random-digit dialing by offering guidance on different computation methods. For example, calculating response rates for cases in a case-control study are reasonably straightforward because a list of cases is likely available (e.g., incident cases of a specific cancer received by a cancer registry). Thus, a simple proportion of the individuals with incident disease who agree to participate in the study can be computed. For controls selected from the general population, response rates need to combine information about who could be contacted, and among who could be contacted, who agrees to participate. The response rate for controls can be computed using one of the AAPOR standard formulas. Cohort studies and randomized trials tend to sample from

defined subpopulations with a complete list of eligible participants or clinical settings with methods that allow for straightforward response rate computation.

It is widely recognized that response rates for all study types have decreased. This decrease may correlate with the increase in (and dislike of) telemarketing, overscheduled lifestyles, and lack of trust in government, academia, and medicine to use time efficiently and effectively. The Behavioral Risk Factor Surveillance System is a national random-digit-dial (RDD) telephone survey that collects information about health behaviors and health care access and is administered by the Centers for Disease Control and Prevention in conjunction with the states. This survey provides an example of the decrease in response rates over the past two decades. The BRFSS response rate declined from 71% in 1993 to 51% in 2005.

Response Rates for Telephone Surveys With Random Samples

Telephone surveys have their own set of issues regarding response rates. Not all telephone numbers belong to households; some belong to businesses. Because many people use technology to screen phone calls, it may not be possible to separate those who refuse to participate from those who are simply unavailable (for instance, not at home). Those who answer may not provide information to determine if an eligible person for the study resides at home, and some eligible people refuse to participate. The computation of a response rate for RDD telephone surveys requires multiple levels of information. The RDD response rate typically comprises two elements: the contact rate and the cooperation rate, both of which are really proportions (not rates). The *contact rate* is the proportion of nonbusiness numbers dialed resulting in households reached. The *cooperation rate* (sometimes called the *participation rate*) is the proportion of contacted eligible units resulting in completed interviews. While it may seem simple to construct numerators and denominators for these proportions, the myriad formulas take into account the almost 20 ways a phone call may or may not result in an eligible household or person being contacted. When comparing response rates in an assessment of data collection quality, it is clearly important to take into account the formulas used.

Response Rates in Clinical Settings

Studies conducted in clinical settings have the major advantage of having convenient sampling options. Consecutive sampling, for example, considers every patient to be eligible from the beginning of the study period until the end if they meet specific criteria (e.g., age, diagnosis). Thus, a list of all patients in the order of their appointments would constitute the sampling frame. The response rate is simply the proportion of eligible patients who agreed to participate in the study. This type of simple computation is possible for all studies where a list of eligible individuals can be developed, either before the study starts or throughout the study (e.g., a daily list of patients who have medical appointments).

The Target Response Rate

What response rate would convey good coverage of the target population? The answer is ''it depends.'' The higher the response rate, the better it is. Historically, a response rate of 80% or more was required to establish scientific validity. However, the decline in response rates over the past two decades has eroded this standard. Some researchers have resorted to the inherently flawed ''same as other studies'' standard as justification for their response rates.

Because response rates are declining for both telephone and mailed surveys, researchers are studying the implications of low response rates. This research is in its infancy and, not surprisingly, has generated mixed results. Studies that compared low-response telephone surveys with higher-response interview surveys found similar response patterns for the major components, such as health behaviors and access to health care. Such a comparison conducted for the BRFSS found that it provided very similar estimates as the National Health Interview Survey, which has a response rate of about 90%. Other studies identified the potential for bias due to nonresponse. For example, in a study of response to telephone surveys on domestic violence, individuals who had experienced domestic violence were more likely to participate than those who had no such experience. In general, individuals who have a particular interest in the study topic are more likely to participate than others. The implication is that disease or events may be

overestimated by studies with a particular focus; such a possibility must be carefully considered when evaluating the findings.

Reasons for Nonresponse

Many reasons for nonresponse exist, only some of which can be controlled by the researcher. Today's lifestyle is busier than ever, and people are more careful in assessing survey requests, selecting only those with the highest value for themselves, their families, and the community. Trust is an issue, especially for telephone surveys; potential respondents may evaluate whether they believe the request comes from a scientific study or a telemarketer, for instance. Another issue is conflict between a common RDD protocol that prohibits leaving a message on an answering machine, and the custom of some families to not answer the phone unless they know who is calling.

Once contact is made with an eligible individual, several factors can affect willingness to participate. Longer surveys will garner fewer participants than those projected to take little time (e.g., 5 min or less). Interviewers themselves have a good deal of impact on response, as rapport is or is not established within seconds. Thus, untrained interviewers or those with an unfamiliar accent may be less likely to elicit cooperation. For mailed surveys, visual appeal and ease of completion are important factors in determining likelihood of response.

Methods to Improve Response Rates

Several methods are known to improve response rates. In-person recruitment and interviews tend to be more successful than telephone recruitment, which tends to be more effective than mail surveys. It is not yet clear where Web-based recruitment and survey may fall in terms of response. Recent information suggests that response rates may be reasonably high for some Web-based surveys, but if surveys flood the Web, response rates are likely to drop just as they did with RDD and mailed surveys.

Because in-person recruitment and interview tends to be expensive and logistically difficult, it is important to maximize the quality of alternative methods. Several relatively simple and cost-effective methods can improve response rates:

- First, establish credibility. Clarifying the purpose of the research and the organization conducting the research can greatly encourage participation.
- Send a postcard, letter, or e-mail to inform individuals of the upcoming study and evoke their interest before surprising them with a telephone call or mailed survey. Provide a phone number as part of the information to allow potential participants to ask questions in advance.
- Use reminder messages to encourage participation and improve response.
- The use of incentives is now normative. Incentives range from trinkets to cash. The amount needs to be carefully weighed for usefulness and the risk of coercion.
- Keep the survey simple and pleasant.

Sampling in special populations (e.g., medical clinics) requires constant attention to improve and maintain high response rates while protecting privacy. It is critical that good surveys and sampling plans be designed to maximize response rates. Studying a sample of nonrespondents compared with participants will help interpret the results of studies and inform the reader about the data quality.

—Shazia Hussain and Louise-Anne McNutt

See also Bias; Interview Techniques; Questionnaire Design; Random-Digit Dialing; Sampling Techniques

Further Readings

Langer, G. (2003). About response rates: Some unresolved questions. *Public Perspective*, pp. 16–18. Retrieved August 8, 2007, from http://www.ropercenter.uconn.edu/pubper/pdf/pp14_3c.pdf.

Nelson, D. E, Powell-Griner, E., Town, M., & Kovar, M, G. (2003). A comparison of national estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System. *American Journal of Public Health*, *93*(8), 1335–1341.

Schmidt, L., Morton, S. C., Damberg, C., & McGlynn, E. A. (1998). An overview of methods for conducting surveys. In E. A. McGlynn, C. Damberg, E. A. Kerr, & R. H. Brook (Eds.), *Health information systems: Design issues and analytic applications* (pp. 189–212). Santa Monica, CA: RAND. Retrieved August 8, 2007, from http://www.rand.org/pubs/monograph_reports/MR967/mr967.chap8.pdf.

Smeeth, L., & Fletcher, A. E. (2002). Improving the response rates to questionnaires. *British Medical Journal*, *324*, 1168–1169.

Ricketts, Howard
(1871–1910)

Howard Taylor Ricketts was an American pathologist and an ambitious pioneer of infectious disease who became renowned as the first to establish the identity of the infectious organism that causes Rocky Mountain spotted fever. The groundbreaking efforts of Ricketts and his research team was one of the earliest collaborations between physicians and entomologists, and the results have had an enormous impact on the often interdisciplinary field of epidemiology. His findings opened new pathways of knowledge in understanding the etiology of diseases.

Ricketts was born in Findley, Ohio. He completed his undergraduate degree in zoology at the University of Nebraska and went on to Northwestern University where he attained his medical degree. While working as a professor of pathology at the University of Chicago, he became interested in the mysterious and widely feared disease that was causing a very high fever and spots on the skin and was killing people who were spending a great deal of time outdoors. In 1906, Dr. Ricketts devoted his research on the discovery of the etiology of Rocky Mountain spotted fever. He characterized the basic epidemiologic features of the disease, including the role of tick vectors.

His definitive studies in the endemic area of Montana's Bitterroot Valley (where the disease was especially virulent) found that Rocky Mountain spotted fever was caused by a microorganism now called *Rickettsia rickettsii*. These unique microorganisms have both bacterial and viral characteristics and are pathogenic in humans. Ricketts demonstrated that Rocky Mountain spotted fever is not only transmitted by wood ticks but also caused by a bloodborne bipolar bacillus. Although he observed a small bacillus, Rickets was unable to culture a causal agent. His work suggested that bacterial diseases could be biologically passed from pests to people in his published findings in 1909, "A Micro-Organism Which Apparently Has a Specific Relationship to Rocky Mountain Spotted Fever: A Preliminary Report."

Through a series of groundbreaking investigations, now considered landmark epidemiological achievements, Ricketts used noninfected guinea pigs as hosts for ticks carrying the disease, and afterward the guinea pigs developed the infection. He proved that a nonfilterable virus, not protozoa, was the etiologic agent for Rocky Mountain spotted fever. Ricketts was quite devoted to his research and was known to inject himself with pathogens on several occasions to measure their effect.

Four years later, he showed that typhus is caused by a similar organism carried by lice. Tragically, Dr. Ricketts died of typhus (another rickettsial disease) in Mexico in 1910 at the age of 39, shortly after completing his remarkable studies on Rocky Mountain spotted fever. His death came only a few days after he isolated the organisms he believed caused typhus. The two organisms Ricketts discovered were the first of what were later shown to be an unusual genus of virus-like bacteria—the *Rickettsiae*. He is now remembered as one of the great martyrs of epidemiologic research.

—*Sean Nagle*

See also Etiology of Disease; Insect-Borne Disease; Parasitic Diseases

Further Readings

Heyneman, D. (2001). The blight of the Bitterroot, the Mysterious Rocky Mountain Spotted Fever, and the significant role of Wilson and Chowning: A commentary. *Wilderness and Environmental Medicine*, *12*(2), 118–120. Retrieved August 8, 2007, from http://www.wemjournal.org/pdfserv/i1080-6032-012-02-0118.pdf.

Ricketts, H. T. (1909). A microorganism which apparently has a specific relationship to Rocky Mountain spotted fever. *Journal of the American Medical Association*, *52*, 379–380.

Robust Statistics

Maximum likelihood (ML) is the most widely used approach for statistical inference. Although it has the advantage of employing straightforward calculations, the ML approach lacks robustness, giving rise to spurious results and misleading conclusions. Researchers in epidemiology and a variety of other experimental and health sciences are becoming increasingly aware of this issue and are informed about the available alternatives for more reliable inference.

Concept of Robustness

What is robustness? Although it is intuitively clear what robustness should be, there is no unique statistical definition, in part because of the diverse aspects of robustness. The generally accepted notion is that a robust statistical procedure should be insensitive to changes not involving the parameters, but sensitive to changes in model parameters. For example, the ML approach is the most powerful for detecting changes in the parameters under the model. However, it is generally sensitive to model assumptions, yielding biased estimates and incorrect inference when the study data depart from the model. A robust procedure aims to provide good power under the model, while still yielding reliable estimates when data drift away from the model.

To elucidate the basic idea, consider a relatively simple problem of comparing two independent groups. The most common procedure is the t test developed based on ML under normal distribution assumption. This procedure compares the two sample means for evidence of group differences. If the data are normally distributed for both groups, the difference statistic between the two group means has a t distribution, providing the basis for inference (i.e., p values and confidence intervals). In many applications, however, data often deviate from the normal model. Such departures from normality can affect both the estimate and inference. For example, the difference statistic may severely over- or underestimate the true group difference in the presence of outliers, giving rise to biased estimates. In many applications, the difference statistic may be unbiased, but the skewness and sparseness in the data distribution may seriously affect the sampling distribution of the statistic, making inference based on the t distribution incorrect. Thus, a robust procedure must address either one or both issues.

Robustness Approaches

A common cause of bias in the estimate is outliers (observations that are exceptionally large or small). Although the sample mean is easy to interpret and work with, it is sensitive to such outlying observations. The common approach to address the effect of outlier is the use of order statistics. By ordering the observations from the smallest to the largest, we can define estimates that are not influenced, or are less influenced, by outliers. For example, the trimmed mean is the sample mean calculated based on the data after removing a certain percentage of observations in the smallest and largest range of the order statistic. Alternatively, one may downweight such outliers to lessen their effect. For example, the winsorized mean is the sample mean after replacing a fraction of the lowest and highest values by the next values counting inward from the extremes, respectively. The sample median is yet another common robust estimate based on the order statistic. Thus, for comparing two groups, we can also form a difference statistic by using any of these robust estimates.

Although these order statistic-based estimates are all more robust than the sample mean, they are not widely used in real study applications. First, these estimates contain a subjective element regarding what constitute outliers and how they are treated. For example, the trimmed and winsorized means involve subjective decisions for trimming and downweighting observations. Second, order statistic-based estimates often give rise to very complex sampling distributions even for univariate outcomes. As a result, existing methods do not apply to cohort and longitudinal studies, which are becoming increasingly popular in epidemiologic and other health-related research. Thus, for most real study applications containing outliers, two sets of analyses are usually performed to examine their effect; one that includes all data and the other that leaves out the outliers. In cases where the outliers have substantial effect on inference, investigation is conducted to determine the nature of their occurrence and whether they should be included in the analysis. Such sensitivity analysis is easy to perform and bypasses the technical difficulties for inference using the order statistic-based estimates.

Even in the absence of outliers, inference based on ML may still be wrong if the distribution assumptions are violated. For example, in the two-group comparison case, if the data from one or both groups are skewed or heavily tailed, the difference statistic generally does not follow the t distribution. Thus, although the statistic may be unbiased, the t distribution is no longer appropriate for inference. The two most popular alternatives are the asymptotic theory-based large sample and permutation-based exact inference. While large sample procedures require large sample size for valid inference, exact methods apply without this restriction.

The estimating equations (EE) approach is the most widely used large sample procedure. Rather than

relying on the likelihood for inference as with the ML approach, the EE constructs estimates and sampling distributions based on a set of estimating equations. Since the equations can be set up without any distribution assumption, this approach provides valid inference regardless of the data distribution. The EE approach applies to many commonly used models such as multiple regression, logistic, and log-linear models. Its extension to longitudinal (or panel) data and other types of clustered data, known as the generalized estimating equations (GEE), has been widely used in biomedical, epidemiologic, behavioral, and social science research.

For data with relatively small sample sizes, large sample procedures may not be applicable; in such cases, exact methods provide an alternative for inference. Exact methods are developed based on the permutation distribution of a test statistic under the null hypothesis. Fisher's exact test for analysis of a 2×2 contingency table is the most familiar example of this approach for discrete outcomes. We can also readily apply such methods to the two-group comparison problem in our example. Under the null hypothesis of no between-group difference, the two groups have the same distribution. Thus, group membership does not matter and can be arbitrarily mixed or permuted between the two groups. By calculating the difference statistic for all possible permutations, we obtain the permutation distribution of the statistic under the null hypothesis and use it like a sampling distribution for inference. Even for small sample size and a simple problem such as two-group comparison, it is difficult to find the exact permutation distribution because of the astronomically large number of permutations and formidable computing problems. In practice, we often approximate the permutation distribution by considering 1,000 to 5,000 different permutations. Such a Monte Carlo implementation generally provides reasonably good results.

Discussion

Robust estimate and robust inference are related, but differ in both concept and application. Estimate robustness is concerned with locally contaminated data such as outliers, while inference robustness addresses global distribution assumptions such as normality. For example, when comparing two groups using the median or trimmed mean, we may still assume normality, in which case inference may still be wrong if the normal model is violated. On the other hand, although inference based on EE is robust against violations of distribution assumptions, it can still be wrong if EE-based estimates are seriously biased, as in the presence of outliers.

We may also apply robust inference to robust estimates. For example, instead of large sample inference, we can compare two-group medians using exact inference. This combination of robust estimate and inference is robust against not only outliers, but data distributions as well.

Rank statistics offer another approach to address robustness. However, as rankings of observations have no direct relationship with the scale of the data, such methods are generally used to provide inference rather than modeling data as the EE and GEE procedures do. For example, the Mann-Whitney-Wilcoxon rank sum test is only used to compare whether there is a shift in location between two otherwise identical distributions. It does not provide any information regarding the distributions or features of the distributions such as mean, median, and standard deviation.

Although EE and GEE are widely used for robust inference for cross-sectional and longitudinal (or panel) data, applications of robust estimates are largely limited to relatively simple, cross-sectional data analyses. In addition, specialized software such as StatXact by Cytel Inc. are often required, as most major statistical packages, including SAS, Splus, SPSS, and STATA, do not provide support for inference based on robust estimates. Alternatively, user-written and supported functions may also be used and are often available free of charge. For example, a set of Splus programs for performing robust analyses, including the median, trimmed, and winsorized mean, can be downloaded from www.apnet.com/updates/ireht.htm.

—Wan Tang, Qin Yu, and Xin Tu

See also Fisher's Exact Test; Nonparametric Statistics; Panel Data

Further Readings

Diggle, P. J., Heagerty, P. J., Liang, K. Y., & Zeger, S. L. (2002). *Analysis of longitudinal data.* New York: Oxford University Press.

Efron, B., & Tihshirani, R. J. (1998). *An introduction to the Bootstrap.* Boca Raton, FL: Chapman & Hall.

Wilcox, R. R. (1977). *Introduction to robust estimation and hypothesis testing.* San Diego, CA: Academic Press.

Rochester Epidemiology Project

The Rochester Epidemiology Project (REP) is a population-based medical records linkage system that was established by Leonard T. Kurland in 1966. It exploits the geographic isolation of Olmsted County in southeastern Minnesota from other urban centers so that almost all medical care for residents of the county (or the central city of Rochester) is delivered within the community. Care is mainly provided by Mayo Clinic and Olmsted Medical Center. Mayo Clinic is a major referral center but has always provided primary, secondary, as well as tertiary care to local residents. It employs a dossier (or unit) medical record for each individual containing the details of every admission to its two affiliated hospitals (St. Marys and Rochester Methodist), every outpatient office or clinic visit, all emergency room and nursing home care, all laboratory results, all pathology reports including autopsies, and all correspondence concerning each patient.

Mayo Clinic now holds medical histories on more than 6.3 million unique individuals (including referral patients); less than a 1,000 of these dossiers have been lost over the past century. The records of the other providers have also been maintained and are available for use in approved research studies. These original and complete (inpatient and outpatient) records that span each person's entire period of residency in the community are easily retrievable for study because Mayo has maintained, since 1910, extensive indices based on clinical and histologic diagnoses and surgical procedures. With continuous support from the NIH for the past 40 years, the REP created similar indices for the records of the other providers of medical care to county residents, most notably Olmsted Medical Center with its affiliated Olmsted Community Hospital.

The result is a unique medical record system capable of addressing research questions in a community population with more than 5.2 million person-years of experience from 1950 through 2005. These detailed data, accumulated over a long period of time, have provided the basis for almost 1,700 studies by investigators inside and outside Mayo. Most of these studies could not have been carried out as efficiently anywhere else. Complete ascertainment of diagnosed cases supports descriptive studies of the incidence and outcomes of diverse diseases and diagnostic/therapeutic procedures. Inception cohorts with verified exposures can be identified for long-term retrospective (historical) cohort studies, and the local community can be enumerated and sampled for population-based cross-sectional and case-control studies. Due to the nature of the database, the main focus has been on clinical risk factors and clinical outcomes; because of the contemporary documentation available, these are much more accurately ascertained compared with self-report. The population of Olmsted County was 124,000 in 2000 and is largely white (99% in 1950, 90% in 2000). Except for a higher proportion of the working population in the health care industry, its population resembles U.S. whites generally. Judged by previous studies of a variety of chronic diseases, results can probably be extrapolated to that population.

—Lee Joseph Melton III

See also Administrative Data; Biomedical Informatics; Clinical Epidemiology

Further Readings

Melton, L. J., III. (1996). History of the Rochester Epidemiology Project. *Mayo Clinic Proceedings*, *71*, 266–274.

Rural Health Issues

The definition of a rural area is complex. According to the U.S. Census Bureau, rural areas are all territories, populations, and housing units not classified as urban. An urban area is defined as one with a total population of at least 2,500 for urban clusters or at least 50,000 for urbanized areas. Rural areas can be located in both metropolitan and nonmetropolitan areas. According to the 2000 census, 21% of the U.S. population (60 million people) live in rural areas.

Many of the health challenges faced by rural America, including chronic illnesses such as hypertension and diabetes, are similar to those faced by all Americans; however, populations in rural areas also have unique health concerns. This entry discusses some of these concerns, including occupational

health, environmental health, and access to health care and also addresses the health of minority groups and migrant workers in rural areas.

Occupational Hazards

Agriculture, fisheries, logging, mining, hunting, and trapping are among the common industries found in rural areas. These industries include the most hazardous occupations for occupational morbidity and mortality. All the above industries require heavy physical labor under demanding weather and environmental conditions. The National Institute of Occupational Safety and Health estimates that 4.5 million people work in agriculture, of whom 23% are minorities. Although agricultural workers make up less than 3% of the U.S. total labor force, they suffer 12% of fatal workplace injuries. In addition to the higher mortality, about 500 agricultural workers experience disabling injuries daily, and 5% of these injuries result in permanent impairment.

The types of occupational hazards that result in increased morbidity and mortality include machinery-related deaths and injuries, noise exposure leading to hearing loss, vibratory exposure leading to neurovascular degeneration of the hands, and the risk of blindness from flying objects. Respiratory irritants can result in asthma, farmer's lung, silo-filler's disease, black lung disease, asbestosis, and asphyxiation. Working with animals can result in zoonotic diseases. In the fisheries industry, there is the risk of drowning or capsizing vessels; in the mining industry, there is the risk of mine collapse. In all these industries, there is physical isolation that results in increased stress and depression.

Environmental Hazards

In addition to the above occupational exposures, there are environmental hazards as well. Rural populations face exposure to chemicals and pesticides, lack of clean drinking water, lack of hand-washing facilities and toilets, temperature extremes, and exposure to zoonotic (animal) diseases. Outside labor with more exposure to sunlight increases the risk of skin cancers.

Pesticides are known to cause a multitude of acute and chronic problems. The EPA estimates that 1.2 billion pounds of pesticides are used annually in the United States, 76% in agriculture. If pesticides seeped into groundwater, they could easily contaminate 90% of rural America's supply of drinking water. Some minor effects of pesticides include eye and nose irritation, fatigue, nausea, and diarrhea. Pesticides can cause human reproductive and developmental toxicity, infertility, neural tube defects, and limb reduction defects. They have also linked pesticides to neurobehavioral problems, Parkinson's disease, depression, and many types of skin problems. Chemical and pesticide exposure can increase risks of non-Hodgkins lymphoma, leukemia, prostate, and stomach and brain cancers.

Health-Related Behaviors

Lifestyle choices such as alcohol and tobacco use are seen in higher proportions in rural areas. In both rural adults and youth, alcohol abuse, alcohol dependence, tobacco use, and illegal drug use are more common than in their urban counterparts. Forty percent of rural 12th graders reported using alcohol while driving compared with 25% in urban areas.

In rural areas, there is an increased use of smokeless tobacco and smoking of homemade or unfiltered cigarettes. Approximately, 28% of rural high school students state that they regularly smoke some type of tobacco. Eleven percent of urban teens report daily tobacco use. Rural 8th graders are twice as likely to smoke cigarettes than their urban counterparts.

Although illegal drug use is common in rural areas, production of amphetamine has increased in isolated rural areas where it is more likely to go unnoticed. In 1999, there were 300 times more seizures of methamphetamine labs in Iowa than in New York and New Jersey combined.

Access to Health Care

Providing health care in rural areas of the United States poses different challenges than in urban areas, because rural populations tend to be poorer, are more likely to travel greater distances to obtain health care, and are less likely to have health insurance. Other issues relating to health care access include inflexible work schedules, loss of pay, language and cultural barriers, poverty, and lack of both public and private transportation.

Access to health care and funding for health care in rural areas are major concerns. Although certain groups of populations have insurance coverage, many do not. The government provides Medicare for the

elderly, Medicaid for those in extreme poverty, and government-sponsored children's health insurance plans for some children. However, during any given year, approximately 15% to 20% of the rural population below age 64 is without health insurance. If you live in a rural area, odds are 80% higher that you will be uninsured.

For those patients seeking health care (with or without insurance), physicians are still difficult to access. Approximately, 75% of all rural counties are considered health care professional shortage areas. There are limited numbers of physicians and specialists per capita. Only about 10% of all physicians practice in rural areas, despite the fact that nearly 25% of the population lives in rural areas. Medicare payments to rural hospitals and physicians are dramatically less than those to their urban counterparts for equivalent services. This correlates closely with the fact that more than 470 rural hospitals have closed in the past 25 years. Since rural hospitals rely on government funding, budget cuts in Medicaid and Medicare make it difficult for many hospitals to continue operations.

Individuals below the age of 64 who work typically do not qualify for government health insurance. Many of the employment opportunities in rural areas are either self-employed or blue-collar jobs, neither of which provides health insurance. These populations are responsible for purchasing private health insurance that can dramatically decrease their monthly income. Instead, many chose to forego health insurance and hope that they will never need to use medical facilities. Many patients without health insurance do not seek regular preventive medical care and only seek care in emergency situations. This lack of preventive care hinders their overall health and quality of care, as well as raising final costs to both the patient and to society. The Institute of Medicine estimates that the United States loses $65 to $130 billion annually because of poorer health and earlier death of those that lack health insurance.

Migrant Farmworkers

The definition of a migrant farmworker is someone who changes residences during the year to accommodate crop harvests. Although the total number of migrant farmworkers is unknown, it is estimated that there are between 2.5 million and 4 million working in the United States. An increasing number of these farmworkers are born outside the United States. In 1998, 89% of migrant farmworkers were Hispanic and 50% admitted that they lacked legal documentation. The health of migrant farmworkers is an important issue in rural health. Not only do these workers harvest and package fresh produce for the United States but their health affects their families, their local communities, and the population at large.

Migrant workers are also exposed to a variety of environmental and occupational hazards, including exposure to extreme temperatures and pesticides. Some of the health issues that affect this population include living in remote isolated areas that do not have health care facilities, increased distance from health care providers, limited choices of providers, poverty, transportation difficulties, inflexible work schedules, discrimination, and language and cultural barriers. They are often housed in crowded living conditions that allow the spread of infectious disease, particularly in populations that have inadequate immunization. Rubella, varicella, and tuberculosis are easily spread in these conditions. Intestinal parasites are common among agricultural workers and may be either imported from their country of origin or may be secondary to polluted water sources within the United States. Their work is often physically demanding and repetitive, and they often suffer from sleep deprivation and separation from family, which may lead to alcohol and drug abuse, violence, and boredom. Prostitution, an additional source of income for some, increases the spread of disease. Lack of experience with machinery and tools can increase injuries. The management of chronic conditions such as hypertension or diabetes is especially challenging in a mobile population, and migrant farmworkers may also be unavailable to obtain the results of preventive screening tests such as pap smears or mammograms. Needs-based programs such as Medicaid, WIC (Women, Infants, Children), and food stamps are rarely used by migrant farmworkers. Approximately 50% of this population has never visited a dentist.

Minority Groups

American Indians and Alaska Natives (AI/AN)

There are approximately 3.3 million AI/AN in the United States, and a higher percentage live in rural areas (including reservations) as compared with the general population. This population has a higher rate

of chronic disease than the average population. Cardiovascular disease, hypertension, and diabetes mellitus are leading causes of morbidity and mortality in this population. Cancer is also higher in this population, with the top cancers including lung, breast, prostate, and colon cancer with lower overall cancer survival rates. AI/AN are three times more likely to die from diabetes as compared with all U.S. races.

The great epidemic among this population is injuries, both intentional and unintentional. Unintentional injuries are the leading cause of death for AI/AN up to 45 years of age. Motor vehicle accidents make up half of all unintentional injuries. The associated factors are alcohol abuse, speeding, and substandard road conditions. There is also a twofold increase in other unintentional injuries such as drowning, falls, poisoning, and burns as compared with the normal population. There are high risks of suicide believed to be associated with loss of traditional culture and exposure to the Western culture. Only 50% of males will reach their 45th birthday.

With improvements in both health care and access to health care, life expectancy has improved dramatically but still lags behind that of the general population. This improvement was primarily seen through sharp reductions in infant mortality. The Indian Health Service (IHS) provides a nationally integrated system of lifetime health care for this population. In addition to providing health care to approximately 1.8 million AI/AN, the IHS provides education, outreach activities and opportunity for tribal involvement in developing and managing programs to meet their health needs.

African Americans

Nearly, 90% of rural African Americans live in the southeastern United States. This population faces higher rates of chronic illness and injury-related mortality and has lower levels of participation in health-promoting behaviors than their urban counterparts. Recent studies have shown that those living in the southeast suffer from higher rates of diabetes, stroke, and obesity. African American populations in rural areas also experience high rates of teenage pregnancy, alcoholism, and uncontrolled blood pressure. Many of these health problems are attributed to lower socioeconomic status, poorer diets, greater risk of occupational injury, higher levels of stress, and less access to health care. Many African Americans also face limited opportunity for quality education and employment.

The Road to Improvement

There are multiple areas for improvement in the health of rural populations. Main areas include increasing safety in the workplace, improving standards regarding environmental exposure, increasing access to health care, and increasing education on the prevention of disease.

Programs such as those of the U.S. Department of Labor's Mine Safety and Health Administration (MSHA) have decreased injuries in the workplace for miners through stricter guidelines and workplace standards. This agency enforces compliance with mandatory safety and health programs required by the 1977 Mine Safety and Health Act. MSHA also aims to prevent or decrease illness and injury on the job through training programs for workers and mine inspectors, technical assistance to miners and mine operators, and efforts to encourage employers to provide safer equipment and more education for the mining community.

Community health programs must stress preventive measures, such as screening and monitoring for chronic illness (especially hypertension and diabetes) and providing public health education and information on issues such as prenatal care, symptom management for chronic illness, diet, exercise, and tobacco use. An often-undervalued resource for youth in rural communities is the cooperative extension service. This service provides education aimed toward pregnant teens and youth development and discusses a multitude of topics, including sex education, ATOD (alcohol, tobacco, other drugs) use, smoking prevention, and nutrition. The 4-H program encourages citizenship, leadership, and teaches life skills to teenagers in rural areas.

Access to health care and increasing the number of insured patients needs to be at the top of the list to improve rural health. Improved health status among rural Americans would reduce health care expenditures, and it would improve the quality of life and productivity of the population. Although the Health Service Corp and the IHS have increased the number of physicians in rural areas, more needs to be done to recruit and retain physicians in these areas to provide needed medical assistance.

—Amanda Bush Flynn

See also African American Health Issues; Environmental and Occupational Epidemiology; Geographical and Social Influences on Health; Health Disparities; Immigrant and Refugee Health Issues; Native American Health Issues; U.S. Public Health Services

Further Readings

Coward, R., Davis, L., Gold, C., Smiciklas-Wright, H., Thorndyke, L., & Vondracek, F. (2006). *Rural women's health: Mental, behavioral, and physical issues*. New York: Springer.

Glasgow, N., Morton, L., & Johnson, N. (Eds). (2004). *Critical issues in rural health*. Ames, IA: Blackwell.

Loue, S., & Quill, B. (Eds). (2001). *Handbook of rural health*. New York: Kluwer Academic.

Straub, L., & Walzer, N. (Eds.). (1992). *Rural health care: Innovation in a changing environment*. Westport, CT: Praeger.

Web Sites

Indian Health Service, U.S. Dept of Health and Human Services: http://www.ihs.gov.

Mine Safety and Health Administration, U.S. Department of Labor: http://www.msha.gov.

National Health Service Corps, U.S. Dept of Health and Human Services, Health Resources and Services Administration: http://nhsc.bhpr.hrsa.gov.

National Industry for Occupational Safety and Health: http://www.cdc.gov/niosh/homepage.html.

National Rural Health Association: http://www.nrharural.org.

Rush, Benjamin
(1745–1813)

Acknowledged as a father of modern psychiatry as well a signatory of the Declaration of Independence, Benjamin Rush modeled a social activist's tireless approach to public health issues in early American society, while becoming the most famous physician and medical educator of his generation. As a boy in Byberry Township outside Philadelphia, Rush was a student at West Nottingham Academy (now the oldest boarding school in the country) under Reverend Samuel Finley, who subsequently became the president of the College of New Jersey (now Princeton University). Although initially intent on becoming a lawyer, Rush switched to medicine. He attended medical school in England, graduating from the medical department of the University of Edinburgh in 1768. During his time abroad, Rush attended medical lectures in Paris, where he befriended Benjamin Franklin, who became a benefactor. Later, he also developed close relationships with John Adams and Thomas Jefferson.

Medicine and Politics

On his return to the United States, Rush was appointed Professor of Chemistry at the College of Philadelphia and launched a public career as a political activist and advocate of colonial rights. As a physician, he practiced extensively among the poor. In 1771, Rush published essays on slavery, temperance, and health. In 1774, he delivered the annual guest lecture at the Philosophical Society, titled "Natural History of Medicine Among the Indians of North America." Two years later, after the adoption of the Declaration of Independence, he was elected to Congress. The same year, he married Julia Stockton, daughter of another signatory of the Declaration, and together they had 13 children. Indiscreet criticism of George Washington, however, somewhat sullied his reputation and led him to resign a medical post in the Continental Army. Nonetheless, he attended the wounded at various Revolutionary War battles while retaining his medical posts in Philadelphia.

In 1793, Rush was credited with overcoming the yellow fever epidemic in his home city, during which he treated over a 100 patients a day. A founder of Dickinson College, Rush was also a strong advocate of public education, including education of women. Rush served as Professor of Medical Theory and Clinical Practice at the University of Pennsylvania for 22 years, and from 1799 until his death, he was also Treasurer of the U.S. Mint. In 1803, Rush succeeded Benjamin Franklin as President of Pennsylvania Society for the Abolition of Slavery.

Health Beliefs and Publications

Rush's most important publications include *An Inquiry Into the Effects of Ardent Spirits on the Human Mind and Body* (1784), *Medical Inquiries and Observations* (a five-volume set, last published in 1806), *Essays, Literary, Moral, and Philosophica* (1798), *Sixteen Introductory Lectures* (1811), and *Diseases of the Mind* (1812, 5th ed., 1835). He also edited several other medical books. Some of Rush's

medical ideas, such as his beliefs about the value of bloodletting and purging, have proved to be ill founded. On the other hand, he sensed the dangers of tobacco use and is recognized as the first physician to define alcohol dependence, or what he termed *inebriety*, as a disease.

—Merrill Singer

See also Alcohol Use; Tobacco; Yellow Fever

Further Readings

D'Elia, D. (1966). Dr. Benjamin Rush and the American medical revolution. *Proceedings of the American Philosophical Society*, *110*(4), 227–234.

Powell, J. (1965). *Bring out your dead: The great plague of yellow fever in Philadelphia in 1793.* New York: Time.

Shryock, R. (1972). *Medicine and society in America: 1660–1860.* Ithaca, NY: Cornell University Press.

S

Sample Size Calculations and Statistical Power

Sample size determination and prospective power analysis are important factors in planning a statistical study, because a study executed with an inappropriate sample size may result in wasted resources. If the sample is too small, the inference goal may not be achieved and true effects may not be detected, and if it is too large, money and resources have been unnecessarily expended, and subjects may have been exposed unnecessarily to a drug or the treatment. Conversely, with an optimal sample, the investigator has increased chances of detecting true effects without wasting resources or exposing subjects to unnecessary risk. For this reason, many funding agencies required sample size determinations to be supported by statistical power analysis. To illustrate, the National Institutes of Health (NIH) Policy Manual (NIH, 1988) states that requests for approval from the Office of Management and Budget to conduct epidemiological studies should include a discussion of sample size and statistical power analysis, among other things. In general, a power of at least 80% is considered satisfactory.

The process of determining the optimum sample size that will give adequate power to detect effects is a task that involves the entire research team. Several factors influence the statistical power of the study and consequently the required sample size. The factors include the hypothesis to be tested, the probability model to test the hypothesis, the significance level α, and a guess on the variance and effect size; that is, in the simplest terms, the expected difference between groups based on scientific considerations. Because some of the factors that influence the statistical power of the study have to be guessed, the statistical power should be tested under different scenarios of sample size, assumptions, and study conditions. For example, if it is important to detect a relative risk of 3, then power analysis should also include a range of neighboring values such as 2 and 4.

There are many formulas available for sample size calculations, although some of the commonly used formulas are based on approximations that assume that large sample sizes will be used in the study. In 1989, in an article titled "How Appropriate Are Popular Sample Size Formulas?," Kupper and Hafner (1989) discuss the use of some of the formulas that have large-sample approximation and recommend the use of formulas that consider statistical power. In epidemiological studies that plan to test hypotheses, it is very important to select a formula that will consider power. Note that not all the formulas consider power estimates. Generally speaking, power analysis is important in obtaining a balance between Type I and Type II errors.

Statistical Power

The concept of power is analogous to the concept of error type. The significance level α is the probability of Type I error—that is, the probability of finding a statistically significant difference by chance—when it does not truly exist. The power is $1 - \beta$, where β is the probability of Type II error. Recall that the Type II error is the probability of not finding a statistically significant difference when it exists. Therefore, the power is the

probability that the null hypothesis will be rejected given that the alternative hypothesis is true. This means that a study with low power will not be able to detect significant effects. Thus, power analysis has the objective of balancing Type I and Type II errors.

The statistical power will depend on the level at which the significance was fixed. For example, imagine that the significance level is fixed at 0.01 instead of 0.05. That will require a larger confidence level $(1-\alpha)$ of 99% instead of 95%. It will be harder to find significance with 99% than it would be with 95%, and therefore, the statistical power of the study will be lower.

The approach taken in a power analysis to define the optimum sample size depends on the nature of the data (e.g., Are the variables continuous, binary, or ordinal?) and the statistical test or model that will be used. For example, for comparisons of means with *t* tests and analysis of variance (ANOVA), power can be approximated using the noncentral *t* distribution or noncentral *F* distribution. With categorical data, for comparisons of proportions in contingency tables with Pearson chi-square tests, power can be approximated using the noncentral chi-square distribution. Agresti (2002) and O'Brien and Muller (1993) discuss methods of statistical power analysis for many other tests such as the *t* test of correlated means, generalized linear models, and logistic regression. Statistical power analysis is an active area of statistical research and better approaches to power calculations are being developed, in particular for more complex models such as specific mixed models. Another approach to power calculation are methods based on simulation: This approach is a particularly effective tool for verifying if the approximations chosen are reasonable and for performing calculations in cases where good approximations are not yet available.

Examples of Calculations

Two-Sample t Test

Suppose that one plans to test the difference between the means of two groups, $\mu_1 - \mu_2$. For example, the null hypothesis can be that there is no difference between the means, $\mu_1 - \mu_2 = 0$, and that it can be assumed that the parameters μ_1 and μ_2 have a common standard deviation, σ. In this case, the power calculation will involve the noncentral *t* distribution, where the power will be the probability

$$1 - p(z \leq t(df)^* - \delta),$$

where z is the standard normal random variable, t^* is the *t* distribution critical value, at the selected significance level α, with degrees of freedom *df* (which is equal to the total sample -2), and the noncentrality parameter δ defined as

$$\delta = \frac{|\mu_1 - \mu_2|}{\sigma}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2}.$$

Note that if $\mu_1 = \mu_2$, then the noncentrality parameter δ will be 0. This is equivalent to the null hypothesis. The larger the noncentrality parameter δ is, the higher the departure from the null hypothesis and the power.

Imagine the following numerical example. An investigator is planning a study with two groups consisting of 45 subjects in each group. He or she wants to be able to detect a difference of at least 9 points $(\mu_1 - \mu_2 = 9)$ in the group means, at 5% confidence level. He or she has a guess that the standard deviation will be equal to 15. From a table, he or she obtains the value of t^* (upper tail probability $= \alpha = 0.05$ and $df = 45 + 45 - 2 = 88$) of 1.99. The statistical power of the study will be

$$1 - p(z \leq t^* - \delta) = 1 - p\left(z \leq 1.99 - \frac{9}{15}\left(\frac{1}{45} + \frac{1}{45}\right)^{-1/2}\right) \approx 0.8.$$

The study has a statistical power of 80%. Therefore, if all the assumptions hold, the study will correctly reject the null hypothesis at 5% significance level in 80% of all possible samples.

Analysis of Variance

The power test for the one-way ANOVA is fairly similar to the power test for the two-sample *t* test, except that instead of the noncentral *t* distribution, the noncentral *F* distribution will be used. Consider that one plans to test the null hypothesis that the mean of *I* groups is the same, $\mu_1 = \mu_2 = \cdots = \mu_i$, and that we can assume a common standard deviation σ. Let *I* be the total number of groups, *N* be the total sample size, and n_i and μ_i be the sample size and the mean of each group, respectively. In this case, the power will be probability

$$p(F(df_{\text{hypothesis}}, df_{\text{error}}, \lambda) \geq F(df_{\text{hypothesis}}, df_{\text{error}})^*),$$

where $F(df_{\text{hypothesis}}, df_{\text{error}})^*$ is the F distribution critical value at the selected significance level α and $F(df_{\text{hypothesis}}, df_{\text{error}}, \lambda)$ is the noncentral F distribution value with degrees of freedom $df_{\text{hypothesis}}(=I-1)$ and $df_{\text{error}}(=N-1)$, and the noncentrality parameter λ defined as

$$\lambda = \frac{\sum n_i(\mu_i - (N^{-1}\sum n_i\mu_i))^2}{\sigma^2}.$$

Similar to δ, if all the means are equal, then the noncentrality parameter λ will be zero; and the larger the noncentrality parameter λ is, the higher the departure from the null hypothesis and the power.

To illustrate, imagine that the investigator is now planning a study with three groups consisting of 30 subjects in each group. His or her null hypothesis is that there is no difference in the group means (H_0: $\mu_1 = \mu_2 = \mu_3$), at a 5% significance level. Based on his or her experience, he or she guesses that the means are $\mu_1 = 55$, $\mu_2 = 60$, and $\mu_3 = 65$, with a standard deviation of 15.

From a table, he or she obtains the value of F^* (upper tail probability $= \alpha = 0.05$, $df_{hypothesis} = 3-1 = 2$ and $df_{error} = 120-1 = 119$) of 3.07, $\lambda = 6.67$, and the power of the study will be

$$p(F(2, 119, 6.67) \geq 3.07) \approx 0.62.$$

The study's statistical power of 62% would probably be considered low, and the investigator may opt to increase the sample size.

Pearson Chi-Square Tests of Independence

Imagine that one plans to have a contingency table with two levels of exposure versus levels of severity of a certain disease. One plans to test the null hypothesis (H_0) that there is independence between levels of exposure to a virus and levels of disease severity. Let π represent the true proportion in each of the cells in the contingency table, and let $\pi(H_0)$ represent the proportions under the null hypothesis. If the sample is large, the power calculation can use the noncentral chi-square distribution as discussed by Alan Agresti, where the power will be the probability

$$p(\chi^2(df, \lambda) > \chi^2(df)^*),$$

where $\chi^2(df)^*$ is the chi-square distribution critical value at the selected significance level α, $\chi^2(df, \lambda)$ is the noncentral chi-square distribution, and the noncentrality parameter λ is defined as

$$\lambda = \frac{\sum n_i[\pi_i - \pi_i(H_0)]^2}{\pi_i(H_0)}.$$

To illustrate, imagine that the investigator is now planning a study with four groups consisting of 50 subjects in each group. The proportion of subjects with disease is expected to be 0.05, 0.12, 0.14, and 0.20 in each of the groups. For each group, the proportion of subjects without disease will be 1 − the proportion of subjects with disease, 0.95, 0.88, 0.86, and 0.8, respectively. Because the sample size is equal for all the four groups, the expected proportion can be calculated by summing the proportions of subjects with disease and dividing by 4. In this case, $\pi(H_0)$ for subjects with disease is equal to 0.1275, and $\pi(H_0)$ for subjects without disease is equal to 0.8725. Using the formula, he or she estimates λ to be approximately 5.158. From the table, he or she obtains the value of $\chi^2(df)^*$ (upper tail probability $= \alpha = 0.05$, $df = 4-1 = 3$) of 7.81, and the power of the study will be

$$p(\chi^2(3, 5.158) > 7.81) \approx 0.45.$$

The study's statistical power of 45% would probably be considered low, and the investigator may opt to increase the sample size.

Packages and Programs

It is often more convenient to use commercial software packages designed for power calculations and sample size determination than to write customized code in statistical packages such as SAS or R to do the calculations. There are many packages available that perform sample size analyses. Some packages are specific for power analysis and sample sizes calculation, such as PASS®, nQuery Advisor®, and SamplePower from SPSS. SAS includes two procedures (POWER and GLMPOWER) that estimate power for many different statistical tests. Additionally, UnifyPow is a freeware SAS module/macro (written by Ralph O'Brien) that performs power analysis (but requires that SAS be available on the user's computer), and it is available at www.bio.ri.ccf.org/UnifyPow.

—Ana W. Capuano

See also Hypothesis Testing; Sampling Techniques; Study Design; Type I and Type II Errors

Further Readings

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Castelloe, J. (2000). *Sample size computations and power analysis with the SAS system.* Paper presented at the proceedings of the 25th annual SAS User's Group International Conference [Paper 265-25], Cary, NC: SAS Institute.

Kupper, L. L., & Hafner, K. B. (1989). How appropriate are popular sample size formulas? *American Statistician*, *43*(2), 101–105.

National Institutes of Health. (1988, December). *NIH policy manual. Chapter 1825: Information collection from the public*. Retrieved April 16, 2007, from http://www1.od.nih.gov/oma/manualchapters/management/1825.

O'Brien, R. G. (1998). *A tour of UnifyPow: A SAS module/macro for sample-size analysis*. Paper presented at the proceedings of the 23rd annual SAS Users Group International Conference (pp. 1346–1355). Cary, NC: SAS Institute.

O'Brien, R. G., & Muller, K. E. (1993), Unified power analysis for *t*-tests through multivariate hypotheses. In L. K. Edwards (Ed.), *Applied analysis of variance in behavioral science* (chap. 8, pp. 297–344). New York: Marcel Dekker.

Whitehead, J. (1993). Sample size calculations for ordered categorical data. *Statistics in Medicine*, *12*, 2257–2271.

Sampling Distribution

The sampling distribution of a statistic, S, gives the values of S and how often those values occur. The sampling distribution is a theoretical device that is the basis of statistical inference. One use is to determine if an observed S is rare or common.

Creating a Sampling Distribution

Recall that a statistic describes the sample and a parameter describes the population. First, one has the population. Then, one takes a random sample of size n from the population and finds the value of S. This value is the first value of the sampling distribution of the statistic. Take another random sample of size n from the population; find the value of S. This value is the second value of the sampling distribution. This is repeated. The resulting collection of Ss is the sampling distribution of S. Figure 1 gives a visual representation of the process of creating a sampling distribution. The S value varies from sample to sample.

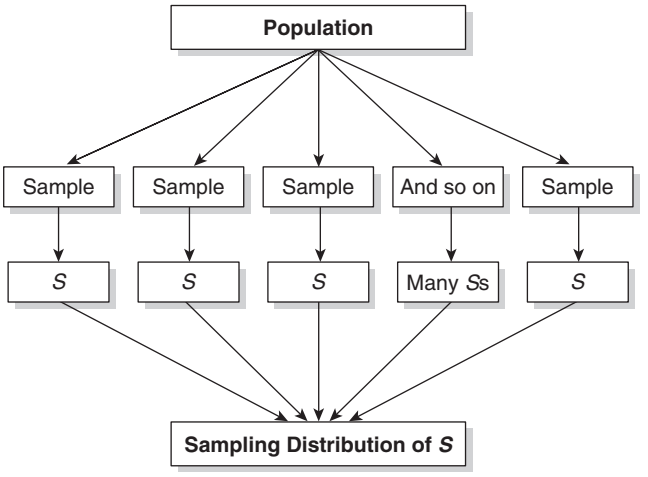


Figure 1 Model of Creating a Sampling Distribution

The sampling distribution of S enables one to see that variability.

The population size is either finite or infinite or large enough to be considered "infinite." When the population size is finite, that is, small, then a proper random sample would be taken from the population *without replacement.* So the number of possible random samples from the population of size N is

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}.$$

Example 1: Finite Population

Example 1 (created by author for this article) contains $N = 6$ items (presented in Table 1), and suppose we are interested in a sample size of 3, $n = 3$, so there are

$$\binom{6}{3} = \frac{6!}{3!(6-3)!} = 20$$

unique random samples from the data set. The population median is 6, the population mean is 5.8, and the population standard deviation is 2.4.

Table 2 contains the 20 unique samples as well as the values for the following statistics: sample mean, sample median, and sample standard deviation. Figure 2 shows

Table 1 Example 1: Finite Population

2	4	5	7	8	9

Table 2 The 20 Possible Random Samples of Size 3 and the Values of Three Statistics

Sample	*Median*	*Mean*	*St. Dev.*	*Sample*	*Median*	*Mean*	*St. Dev.*
2 4 5	4	3.7	1.5	4 5 7	5	5.3	1.5
2 4 7	4	4.3	2.5	4 5 8	5	5.7	2.1
2 4 8	4	4.7	3.1	4 5 9	5	6.0	2.6
2 4 9	4	5.0	3.6	4 7 8	7	6.3	2.1
2 5 7	5	4.7	2.5	4 7 9	7	6.7	2.5
2 5 8	5	5.0	3.0	4 8 9	8	7.0	2.6
2 5 9	5	5.3	3.5	5 7 8	7	6.7	1.5
2 7 8	7	5.7	3.2	5 7 9	7	7.0	2.0
2 7 9	7	6.0	3.6	5 8 9	8	7.3	2.1
2 8 9	8	6.3	3.8	7 8 9	8	8.0	1.0

the sampling distributions of the sample mean, sample median, and sample standard deviation. Notice that the sample distributions are centered on the associated parameter; that is, the sampling distribution of the median is centered on 6, and the same is true for the mean (centered at 5.8) and standard deviation (centered at 2.4). Also, notice that there is variability in the sampling distributions. This variability is known as *sampling variability*—the variability that is induced by the act of taking a random sample. The value of a statistic does not equal the value of the parameter (typically) but varies around the parameter.

By looking at the sampling distribution, one can determine if an observed statistic is rare or common. For example, suppose one takes a random sample of three items and finds that the sample mean is 4 or below. If the random sample is from the original population, then the chance of observing an $\bar{x}$ equal to 4 or less is $1/20 = 0.05$. So there is only a 5% chance that the random sample is from the original population.

Example 2: Infinite Population, Quantitative Values

In most situations, the size of the population is infinite or large enough; therefore, when one takes random samples from the population, one can take the sample with replacement and then take a large number of samples.

In Example 2, the population has an exponential distribution with mean and standard deviation equal to 3, and the median is $2.0794 = -3 \ln(0.5)$. Figure 3 gives

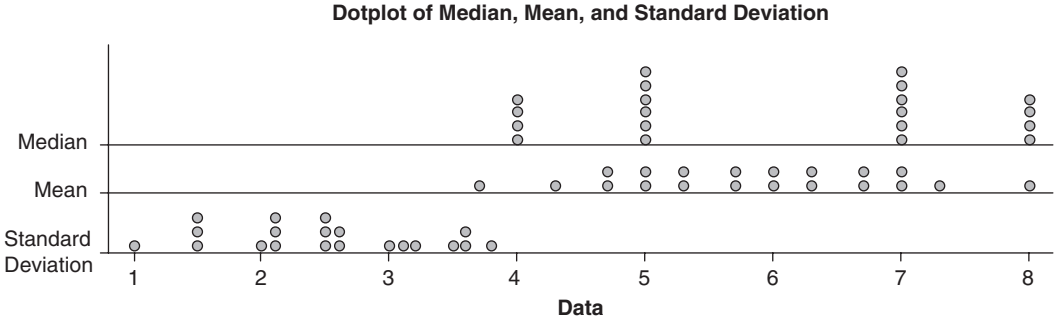


Figure 2 Sampling Distributions for the Median, Mean, and Standard Deviation, Finite Population

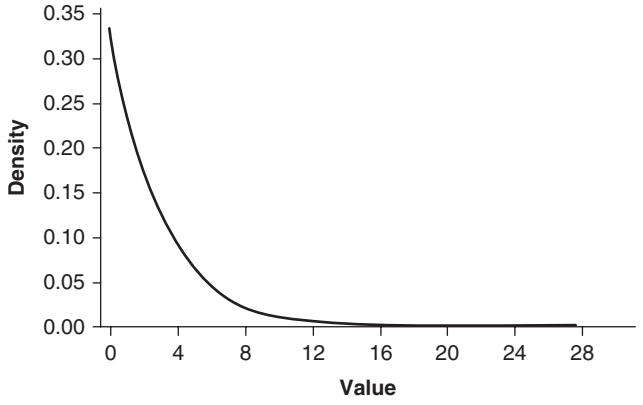


Figure 3 Exponential Population With Mean = 3

the density function of the population. The sampling distributions given in Figures 4, 5, and 6 were created by taking 10,000 samples of size n. The associated statistics were found, and then those statistics were used in the sampling distribution. This process is repeated for five different sample sizes, 5, 10, 20, 30, and 40.

For all three statistics, one observes that the variability of the distributions decreases as the sample size increases. Intuitively, this makes sense because as one has more items in a sample one would expect that the statistic should vary less, that is, the statistic becomes more consistent.

Sampling Distribution of Common Statistics

Sampling Distribution of the Sample Mean

The *central limit theorem* explains that if the population has a finite population mean, μ, and a finite population standard deviation, σ, then

- The population mean of the sampling distribution of the sample mean is μ.
- The population standard deviation of the sampling distribution of the sample mean is $\sigma/\sqrt{n}$.
- The sampling distribution of the sample mean converges to a normal distribution as n is increased.

Sampling Distribution of the Sample Proportion

The sample proportion is the main statistic of interest with categorical data. The population has a population

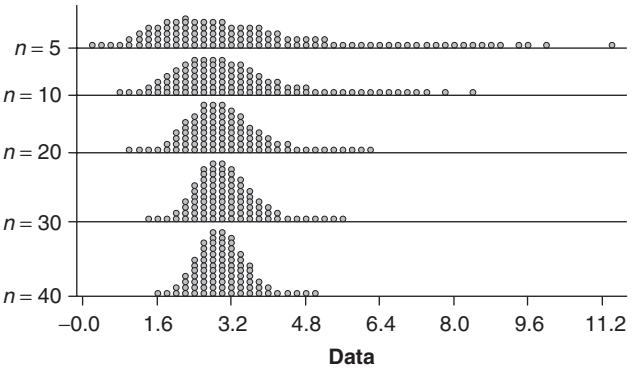


Figure 4 Sampling Distribution of the Mean

Note: Each symbol represents up to 136 observations.

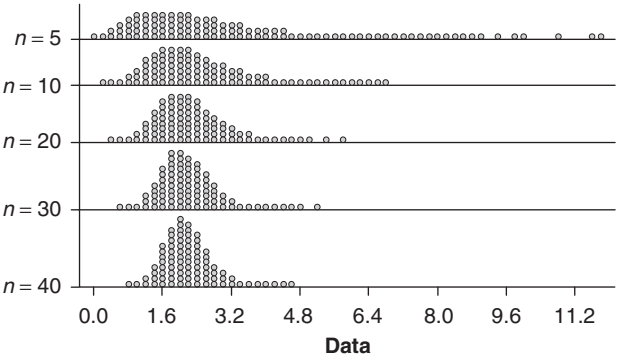


Figure 5 Sampling Distribution of the Median

Note: Each symbol represents up to 141 observations.

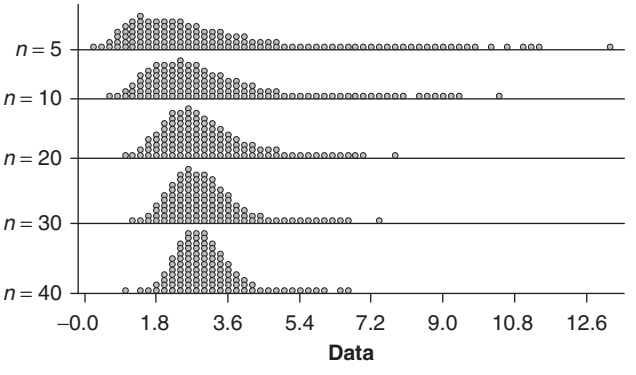


Figure 6 Sampling Distribution of the Standard Deviation

Note: Each symbol represents up to 111 observations.

proportion, p, which measures the proportion or probability of a "success." The sampling distribution of the sample proportion has the following properties:

- The population mean of the sampling distribution is p.
- The population standard deviation of the sampling distribution is $\sqrt{p(1-p)/n}$.
- As the sample size increases, the sampling distribution converges to a normal distribution.

Although one has seen visually the creation of sampling distribution in this entry, statisticians do not solely rely on visual images. The proofs of the sampling distribution of the sample mean and sample proportion are not based on graphs but are based on statistical theory. Statisticians use a variety of tools, such as the moment generating function or characteristic functions and expected values, to determine the distribution of the statistic. However, the key to understanding sampling distributions is not in the statistical theory. The key is understanding that the sampling distribution of S gives the possible values of S and how often S takes those values. In the real world, one does not have multiple samples from a population. Instead, one has a single sample and a single statistic. But by making assumptions about the population, one is able to test those assumptions based on the single sample:

- If the population assumption is true, then one can determine the sampling distribution of S.
- Using this sampling distribution, one determines whether the observed S from the random sample is common or rare.
- If S is rare, then one's assumptions about the population may not be true.
- If S is common, then one does not have sufficient evidence against the assumptions.

—Marjorie E. Bond

Author's Note: All data for this article were created by the author for this entry or simulated using Minitab.

See also Central Limit Theorem; Confidence Interval; Hypothesis Testing

Further Readings

Agresti, A., & Franklin, C. (2007). *Statistics: The art and science of learning from data.* Upper Saddle River, NJ: Pearson Prentice Hall.

Moore, D. S. (2007). *The basic practice of statistics* (4th ed.). New York: W. H. Freeman.

Sampling Techniques

Proper scientific sampling is an important element in the study of populations. The population of interest may be a general population, such as all people in the United States 18 years of age or older, or a targeted subpopulation, such as people in the United States 18 years of age or older living in poverty or living in urban areas. Many epidemiologic studies involve gathering information from such populations. The population of interest in such studies may be specific—such as physicians, people in homeless shelters, people with a specific chronic disease—or may involve any population that can be accurately defined. Drawing an appropriate sample of the target population is the foundation on which such studies must be built. An improper sample can negate everything that the study wishes to discover. This entry describes different techniques that can be used for drawing a proper sample and when they are most appropriate. The Further Readings section provides a more in-depth look at the specific statistical properties of various sampling techniques, such as variance estimation and the approximation of required sample sizes.

Simple Random Samples

Any discussion of sampling techniques must begin with the concept of a simple random sample. Virtually all statistics texts define a simple random sample as a way of picking a sample of size n from a population of size N in a manner that guarantees that all possible samples of size n have an equal probability of being selected. From an operational standpoint, consider a list of N population members. If one used a random number generator to assign a random number to each of the N population members on the list and then selected the n smallest numbers to make up one's sample of size n, this would constitute a valid simple random sample. If one replicated this process an infinite number of times, each possible sample of size n would be expected to be selected an equal number of times. This then meets the definition of a true simple random sample. Many statistical software packages contain

a sampling module that usually applies this type of simple random sampling.

It is important to point out that virtually all statistical procedures described in textbooks or produced as output of statistical software programs assume simple random sampling was performed. This includes the estimation of means, variances, standard errors, and confidence intervals. It is also assumed in the estimation of standard errors around regression coefficients, correlation coefficients, odds ratios, and other statistical measures. In other words, the development of statistical estimation is built around the concept of simple random sampling. This does not imply that proper statistical estimates cannot be derived if simple random sampling was not performed, but that estimates derived by assuming simple random sampling may be incorrect if the sampling technique was something different.

In practice, it is usually the case that simple random sampling will lead to the best statistical properties. It usually will be found to produce the smallest variance estimates and smallest confidence intervals around estimates of interest. It also leads to the least analytic complexity as again all statistical software programs can handle simple random sampling without any problem. It must be noted that a simple random sample will not always produce the best statistical properties for estimation, but that in practice this usually will be the case.

The question then arises as to why other sampling techniques are ever used. The answer is that while simple random sampling may be shown to generally have the best statistical properties, it also frequently may be the most expensive to conduct, may at times be impossible to do, and frequently will not address all the goals of a particular study. In these instances, more complex sampling techniques must be investigated.

Systematic Samples

A systematic sample is the sampling technique that is closest to simple random sampling in nature. A systematic sample is generally defined as an interval sampling approach. From a list or other device that defines a target population, first select a random number between 1 and m. Then starting with this random number, take every mth element into the sample. As an example, one may want to draw a systematic sample of 100 elements from a target population of 1,000 elements. In this case, m could be set to 10. Drawing a random number between 1 and 10 might lead to the number 7. The resulting systematic sample would then be the elements from the original list that were numbered 7, 17, 27, 37, . . . , 997. There are many instances in practice that could be seen as practical applications of such an interval sampling technique. Examples are drawing a sample of files from file cabinets, telephone numbers from the page of a telephone book, housing units on a specific city block, or every mth patient who visits a clinic on a specific day. In practice, it may be more practical to implement a systematic sampling procedure than to draw a simple random sample in these types of instances.

In general, systematic samples usually can be assumed to behave similarly to simple random samples for estimation purposes. Still, great care must be taken to make sure a systematic bias does not result. As an example, consider a proposed study of people in homeless shelters. On a given night, it is decided to take a systematic sample of every third bed in every shelter in a city. From the people placed into these beds, assessments would be made concerning the medical condition and needs of these people. Suppose it was discovered that in numerous shelters beds are arranged as triple-decked bunk beds. Furthermore, the practice in these shelters is to place the most physically impaired people in the bottom bunks, the less impaired in the middle bunks, and the healthiest in the top bunks. In this instance, taking every third bed would result in a completely biased sample. If systematic samples are to be used, it is critical to examine such potentially unforeseen biases that may result from the interval selection.

Complex Random Samples

The term *complex sample* is generally used to mean any sampling technique that adds a dimension of complexity beyond simple random sampling. Three of the most common forms of complex samples will be discussed. They are stratified samples, cluster samples, and samples selected with probability proportionate to size.

Stratified Samples

Suppose the population of interest can be divided into nonoverlapping groups. Examples of such a situation are states into counties, physicians into primary specialties, hospitals into size (measured by number of beds), or patients into primary clinical diagnosis. The

groups in these instances are referred to as strata. The identifying characteristic is that every member of the target population is found in one and only one of the strata. If one is not only interested in studying the whole population but also interested in comparing strata, then stratified sampling is the best sampling technique. For example, suppose one wants to study physicians and make estimates about the characteristics of how they spend their time and the degree of difficulty of their practices. One might also want to be able to compare physicians across specialties. If one were to draw a simple random sample of physicians from some list (such as AMA membership), one would find that there are many more internists than cardiologists, as one example. The resulting simple random sample would either not contain enough cardiologists to be able to accurately compare them to internists or the size of the original simple random sample would have to be so large to ensure that there are enough cardiologists that the study is not feasible. In instances such as these, stratified sampling is the better approach. The original population can be broken into strata by primary specialty. Within each stratum, a simple random sample of physicians can be drawn. The size of the samples within each stratum can be controlled to guarantee that comparisons across strata can be accurately performed. This type of approach can be used in many similar settings. In theory, the resulting stratified sample may even have better sample characteristics, such as smaller variance estimates. In practice, it should be expected that this is not the case.

Cluster Samples

At times, the population of interest can be more easily located by knowing that they reside in naturally nonoverlapping groups. The groups are of no particular interest to the study, but they serve as a mechanism to more easily identify or find the population members that are desired. In these instances, the groups are known as clusters. Many examples exist in practice. If a study wants to sample households within a city to investigate the possible presence of environmental risk factors that can adversely affect the health of household residents, one possibility is to get a list (if it exists) of all households in the city and then to draw a simple random sample of households. If this is done, the cost of sending interviewers across all parts of the city, usually to examine a single house here and another there, can be very expensive. If it is recognized that household residents are in one and only one city block, then the city can be initially divided into nonoverlapping city blocks. A simple random sample of city blocks can be drawn and then a simple random sample of households within each selected block. This makes it far more efficient and less expensive to examine the sample households as they are in groups or clusters of close proximity. These blocks (or clusters) have no intrinsic interest to the study but simply serve as a mechanism for more efficiently getting to the households of interest.

This type of sampling technique is called cluster sampling. It is frequently used in practice as it usually reduces costs. Sampling college students within colleges, residents or interns within hospitals, and patients within physician practices are just a few examples. Although this sampling may be necessary, it frequently increases estimated sample variances and standard errors because variables that are to be measured can be correlated within the clusters. The higher the amount of correlation, the larger the increase in estimated standard errors. This must be factored in and compared with the estimated cost savings to determine if cluster sampling is advisable.

Probability Proportionate to Size Samples

Sampling done proportionate to some measure of size may sometimes be preferable to simple random sampling where each element of a population has an equal probability of being selected into the sample. An example can serve best. Suppose a sample of college students is the goal. Since there is no list that contains all college students in the country, it is necessary to sample college students by initially sampling colleges. Each college does have a list of its own students. Drawing a simple random sample of colleges can be done using a list of all accredited colleges. However, because a random sample of students is the ultimate goal and colleges can have radically different sizes in terms of number of students, a random sample of these colleges would not yield the desired result. In fact, a closer look shows that approximately 50% of the college students in the country can be found in approximately 15% of the colleges. There are a great many small colleges, and it takes many of them to equal the student body of a large state university. In this instance, drawing a simple random sample of colleges would lead to a disproportionately large sample

of students from small colleges. This is not reflective of how students are distributed. A better sampling technique in such instances is sampling proportionate to size. In this instance, the number of students at a college can serve as a measure of size, and colleges are selected using this measure. As a result, the larger the college, the higher the probability its students will be selected. This more closely resembles the actual distribution of students across colleges. If simple random samples of students of equal size are then selected from within each selected college, it can be shown that each student in the original population had an equal probability of being included in the sample. Again, many examples exist, including sampling hospitals using number of beds or surgical procedures as a measure of size, sampling patients in a hospital by length of stay, and sampling census blocks within a city using the estimated number of people with a specified characteristic (e.g., above 65 years of age) as a measure of size. Care must be taken with this approach as measures of size that are in error can lead to biases in the sample.

Summary

The complex sampling techniques just specified can be used alone or in conjunction. For example, the country can be divided into strata by using states; then, within each state, a sample of clusters (e.g., colleges or hospitals) can be drawn using probability proportionate to size sampling and then simple random sampling can be used within each cluster. Whenever complex sampling techniques are used, they *must* be reflected in the resulting statistical analysis. Failure to do so undermines the validity of any stated statistical result. Fortunately, many statistical software programs have modules for handling complex sample designs. The design must be entered correctly into such software programs to get the appropriate results. In practically all cases, failure to do so will result in estimated variances and standard errors that are too small. This in turn leads to potentially stating that results are statistically significant when they are not.

The techniques specified are all probability sampling techniques. Nonprobability samples, such as volunteer samples and convenience samples, cannot be used to make population estimates as they lack the statistical foundation. For clinical trials, volunteers can be randomly assigned to treatment and control groups. Comparing these groups can then result in a valid analysis of the treatment but only in the context of the initial volunteer sample. How volunteers differ from the population in general will be a potential unknown. The techniques mentioned here do not handle all sampling situations but are the most common that are used. In any truly complicated study, the advice of a sampling statistician should always be sought.

—Anthony Roman

See also Questionnaire Design; Randomization; Study Design; Target Population

Further Readings

Cochran, W. G. (1977). *Sampling techniques.* New York: Wiley.

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology.* New York: Wiley.

Kish, L. (1995). *Survey sampling.* New York: Wiley.

Scatterplot

The scatterplot is a graphical technique that is often used to explore data to get an initial impression of the relationship between two variables. It is an important component of descriptive epidemiology and can be helpful in providing clues as to where further exploration of the data may be valuable. It gives a sense of the variability in the data and points to unusual observations. The scatterplot is most frequently used when both the variables of interest are continuous.

To create a scatterplot, the value of the y variable, the dependent variable, is plotted on the y axis against the corresponding value of the x variable, the independent variable, which is plotted on the x axis. This is done for each data point. The resulting plot is a graphical description of the relationship between the two variables. In other words, it is a visual description of how y varies with x. See Figure 1 for an illustration of a scatterplot. The circled diamond corresponds to a data point with a value of 53 in. for height and 80 lb for weight.

Besides being an important technique that allows epidemiologists to get a visual description of the data, aiding in the detection of trends that may exist between two variables, the scatterplot is a useful tool as an initial step in the determination of the type of

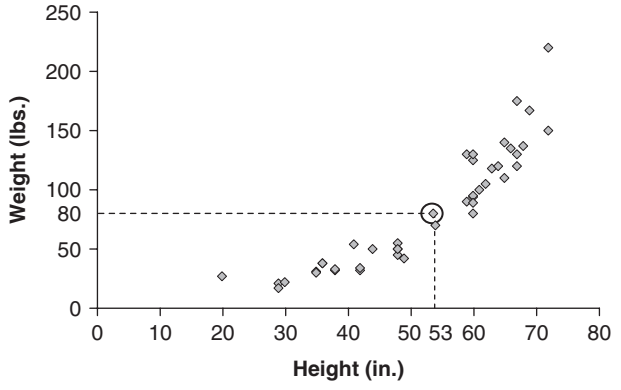


Figure 1 Weight and Height in U.S. Children Aged 0 to 17 Years

Source: Data from random selection of 43 points from Centers for Disease Control and Prevention, National Center for Health Statistics, State and Local Area Integrated Telephone Survey, National Survey of Children's Health (2003).

model that would best explain the data. For example, if the relationship appears linear, linear regression may be appropriate. If, however, the relationship does not appear linear, a different type of regression or the inclusion of a term such as the quadratic term or a spline term may be more appropriate.

Another use of the scatterplot is to explore the relationship between two predictor, or independent, variables. If two predictor variables appear to be highly correlated, this may be an indication that they are collinear. In other words, they may be providing the same information to the model. Adding both variables to the model, therefore, is giving redundant information and may result in inflated standard errors, reducing the precision of the model. For this reason, it is of value to understand how predictor variables are related. The scatterplot is one way to begin to explore this.

The scatterplot can be used when both the variables of interest are not continuous; however, there are other types of graphs that may be preferable and give a better depiction of the data. Box-and-whisker plots are of value when comparing binary or categorical data. Bar charts are useful not only for ordinal data but also for categorical data.

In short, the scatterplot is an important tool to use, especially in the beginning stages of data exploration, to describe the data. It provides insight into how variables are related to each other, what steps should be taken next in the data analysis process, and allows for checks as to whether or not model assumptions are met.

—Rebecca Harrington

See also Bar Chart; Box-and-Whisker Plot; Collinearity; Dependent and Independent Variables; Inferential and Descriptive Statistics

Further Readings

Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

SCHIZOPHRENIA

Schizophrenia is a disorder of psychotic intensity characterized by profound disruption of cognition and emotion, including hallucinations, delusions, disorganized speech or behavior, and/or negative symptoms. This entry reviews the epidemiology of schizophrenia, along with its natural history and risk factors. It focuses on the period since the review by Yolles and Kramer in 1969, concentrating on results that are most credible methodologically and consistent across studies, and on the most recent developments.

Descriptive Epidemiology

The most credible data on the epidemiology of schizophrenia come from registers, including inpatient and outpatient facilities for an entire nation, in which the diagnosis is typically made carefully according to the standards of the World Health Organization's *International Classification of Diseases* and in which treatment for schizophrenia in particular and health conditions in general is free (e.g., Denmark). The global point prevalence of schizophrenia is about 5 per 1,000 population. Prevalence ranges from 2.7 per 1,000 to 8.3 per 1,000 in various countries. The incidence of schizophrenia is about 0.2 per 1,000 per year and ranges from 0.11 to 0.70 per 1,000 per year. The incidence of schizophrenia peaks in young adulthood (15 to 24 years, with females having a second peak at 55 to 64 years). Males have about 30% to 40% higher lifetime risk of developing schizophrenia than females.

Natural History

The onset of schizophrenia is varied. In 1980, Ciompi found that about 50% of cases had an acute onset and about 50% had a long prodrome. About half of the individuals had an undulating course, with partial or full remissions followed by recurrences, in an unpredictable pattern. About one third had a relatively unremitting course with poor outcome; and a small minority had a steady pattern of recovery with good outcome. Several studies have shown that negative symptoms and gradual onset predict poor outcome. There is variation in the course of schizophrenia around the world, with better prognosis in the so-called developing countries. Although there is a long literature on the relation of low socioeconomic position to risk for schizophrenia, it seems likely that the association is a result of its effects on the ability of the individual to compete in the job market. Recent studies suggest that the parents of schizophrenic patients are likely to come from a higher, not lower, social position.

Some individuals with schizophrenia differ from their peers even in early childhood in a variety of developmental markers, such as the age of attaining developmental milestones, levels of cognitive functioning, neurological and motor development, and psychological disturbances, but no common causal paths appear to link these markers to schizophrenia. Minor physical anomalies, defined by small structural deviations observed in various parts of the body (e.g., hands, eyes, and ears), are more prevalent in individuals with schizophrenia and their siblings as compared with the rest of the population. This evidence on developmental abnormalities is consistent with the hypothesis that schizophrenia is a neurodevelopmental disorder, with causes that may be traced to early brain development.

Risk Factors

A family history of schizophrenia is the strongest known risk factor, with first-degree relatives having about 5- to 10-fold relative risk and monozygotic twins having about 40- to 50-fold relative risk. In addition to family history, there are several risk factors that have been identified in the past 50 years, including complications of pregnancy and birth, parental age, infections and disturbances of the immune system, and urban residence.

It has been long known that individuals with schizophrenia are more likely to be born in the winter. This risk factor is interesting in part because it is indisputably not genetic in origin. The relative risk is approximately a 10% increase for those born in the winter compared with those born in the summer. The effect exists in both hemispheres, with more births during the winter in the Southern Hemisphere, which does not coincide with the beginning of the calendar year. One possible explanation is that winter months coincide with seasonal peaks in infectious agents (e.g., influenza) that may affect prenatal development. Many studies have reported a doubling of odds for developing schizophrenia among those with a birth complication (e.g., preeclampsia). Recently, several population-based studies have provided strong evidence about the role of paternal age, rather than maternal age, in schizophrenia.

A series of ecological studies suggested that persons whose mothers were in their second trimester of pregnancy during a flu epidemic had a higher risk for schizophrenia. Prenatal infection as a risk factor is consistent with the neurodevelopmental theory of schizophrenia. Consistent evidence shows that individuals with antibodies to *Toxoplasma gondii* (the parasite that causes toxoplasmosis) have higher prevalence of schizophrenia. A relatively small but consistent literature indicates that persons with schizophrenia have unusual resistance (e.g., rheumatoid arthritis) or susceptibility (e.g., celiac disease) to autoimmune diseases. A single weakness in the immune system in schizophrenic patients may explain both the data on infections and the results on autoimmune disorders, with ongoing studies examining this hypothesis.

In the 1930s, Faris and Dunham showed that admissions for schizophrenia in and near Chicago, Illinois, tended to come from the city center, with decreasing rates among individuals living in zones of transition (the less central part of the city; zones of transition are characterized by mixed-use [commercial and housing] and is a transition to the primarily residential areas that would be even further removed from the city center), a pattern not seen in manic depression. The relative risk of schizophrenia is about two to four times higher for those born in urban areas. Additionally, evidence suggests that schizophrenia is a disease of relatively recent origin. While identifiable descriptions of manic depression have been found during the time of Galenic medicine in the 2nd century AD, descriptions of schizophrenia are vague and rare. There appears to be an

upward trend in schizophrenia over four centuries, with a doubling or quadrupling of the prevalence. Many of the possible explanations for the rise in prevalence of schizophrenia with modernization parallel the explanations for the higher risk in urban areas—animals in the household, crowding in cities, and difficulty formulating a life plan when the future is uncertain.

As late as a quarter century ago, the epidemiology of schizophrenia was nearly a blank page. The only risk factors that seemed strong and consistent were the conditions of lower social class in life and the family history of schizophrenia. Since that time, there has been considerable progress delineating a more or less consistent picture of the descriptive epidemiology and the natural history of schizophrenia. In the future, concerted efforts will be made to study risk factors in combination. Such efforts will make the prospects for prevention more targeted and effective.

—Briana Mezuk and William W. Eaton

See also Child and Adolescent Health; Neuroepidemiology; Psychiatric Epidemiology

Further Readings

Brown, A., & Susser, E. (2002). In utero infection and adult schizophrenia. *Mental Retardation and Developmental Disabilities Research Reviews*, *8*, 51–57.

Cannon, M., Jones, P., & Murray, R. (2002). Obstetric complications and schizophrenia: Historical and meta-analytic review. *American Journal of Psychiatry*, *159*, 1080–1092.

Ciompi, L. (1980). The natural history of schizophrenia in the long term. *British Journal of Psychiatry*, *136*, 413–420.

Eaton, W. W., & Chen, C.-Y. (2006). Epidemiology. In J. A. Lieberman, T. Scott Stroup, & D. O. Perkins (Eds.), *The American Psychiatric Publishing textbook of schizophrenia* (pp. 17–37). Washington, DC: American Psychiatric Publishing.

Faris, R., & Dunham, W. (1939). *Mental disorders in urban areas*. Chicago: University of Chicago Press.

Murray, R. W., & Lewis, S. W. (1987). Is schizophrenia a neurodevelopmental disorder? *British Medical Journal (Clinical Research Edition)*, *295*, 681–682.

Torrey, E. F., Miller, J., Rawlings, R., & Yolken, R. H. (1997). Seasonality of births in schizophrenia and bipolar disorder: A review of the literature. *Schizophrenia Research*, *7*, 1–38.

Yolles, S., & Kramer, M. (1969). Vital statistics. In L. Bellak (Ed.), *The schizophrenic syndrome* (pp. 66–113). New York: Grune & Stratton.

Screening

Screening is the process of systematically searching for preclinical disease and classifying people as likely or unlikely to have the disease. It involves using a reasonably rapid test procedure, with more definitive testing still required to make a diagnosis. The goal of screening is to reduce eventual morbidity or mortality by facilitating early treatment. Mass or population screening targets a whole population or population group. Clinical use of screening tests for diseases unrelated to patient symptoms is referred to as opportunistic screening. This entry discusses the general principles related to mass screening and study designs for the evaluation of screening programs.

Screening is primarily a phenomenon of the 20th century (and beyond). Early screening tended to deal with communicable diseases, such as tuberculosis and syphilis, and had as a goal reducing transmission as much as treatment for the individual's sake. As the public health burden associated with these diseases was reduced in developed countries, the practice of screening evolved to focus on chronic diseases, such as cancer and diabetes, and risk factors for coronary heart disease. Screening programs related to infant and child development, such as newborn screening for genetic and metabolic disorders or vision and lead screening in children, have also become a routine part of public health practice. Newer or emerging concepts include prenatal screening and testing for genetic susceptibility to diseases such as cancer.

Criteria for Screening

The criteria for an ethical and effective screening program were first outlined by Wilson and Jungner in 1968. Their list has been reiterated and expanded over the years and remains pertinent even today. Meeting a list of criteria does not guarantee the success of a screening program, and on the other hand, there may be utility in programs that do not meet every guideline. However, the following are important considerations.

Disease

The disease should represent a significant public health problem because of its consequences, numbers affected, or both. Additionally, the natural history of the disease must include a *preclinical detectable*

period. This is a period of time when it is feasible to identify early disease using some existing test, but before the onset of symptoms that would otherwise lead to diagnosis. This preclinical detectable period should be of long enough duration to be picked up by tests at reasonably spaced time intervals. This is why screening is generally applied for chronic conditions that develop slowly. In an acute disease with rapid onset, even if a short preclinical detectable period exists, testing would need to be extremely frequent to pick it up before symptomatic disease developed. Finally, there should be a treatment for the disease available that improves outcome and is more effective when given earlier.

Screening Test

Practically, the screening test should be relatively simple and quick to carry out, acceptable to the population receiving it, and not prohibitively expensive. It should also be sensitive, specific, and reliable. Having high sensitivity means that there will be few cases of the disease that get "missed"—and miss the opportunity for early treatment. High specificity means that there are few false positives—that is, people without the disease who may undergo further testing and worry needlessly. Although both are ideally high, a trade-off between sensitivity and specificity may occur, especially for tests, such as a blood glucose test for diabetes, where a positive or negative result is determined by dichotomizing a continuous measure. The sensitivity and specificity can only be determined if there is an accepted "gold standard" diagnostic test to define true disease status. Although considered properties of the test, sensitivity and specificity can differ according to population characteristics such as age.

Population

When a population subgroup is targeted for screening, it should be one where the disease is of relatively high prevalence. Higher prevalence of disease results in a higher positive predictive value, which means that there will be fewer false positives. Also, the number of screening tests administered per case identified is lower, improving cost and resource efficiency. For example, the U.S. Preventive Services Task Force recommends screening for syphilis in high-risk groups, such as those engaging in risky sexual behavior or incarcerated populations, but recommends against routinely screening those who are not at increased risk. Screening in a high-risk population is sometimes referred to as *selective screening.*

System

Screening should occur within a system where follow-up testing to confirm disease presence is available for all persons who test positive, and treatment is available to those in whom disease is confirmed. Screening persons who will not have access to these services may not be acceptable ethically, as they will be subjected to the psychological consequences of a positive result without the presumed benefit of early treatment. As a corollary, the screening program should be economically feasible as a part of total health care expenditures within the system. Finally, screening for many disorders should be a continuing process with tests repeated at regular intervals.

Evaluation of Screening Programs

Despite the intuitive appeal of finding and treating disease early, screening programs cannot simply be assumed effective. Because a positive screening result may produce worry for the individual or family, a risk of side effects from diagnostic testing and treatment, and the expenses involved, screening may not be warranted if it does not result in decreased morbidity or mortality. Data may not exist prior to the inception of a screening program to fully evaluate all criteria, and the effectiveness of a program may vary over time or across populations. Therefore, evaluating programs or methods is an important function of epidemiology related to screening. Several different types of study design may be used.

Ecologic studies compare disease-specific mortality rates or indicators of morbidity between screened and unscreened populations, defined, for example, by geographic area or time period before versus after the screening program began. Outcomes between areas with different per capita screening frequencies could also be compared. Such studies may be relatively simple to conduct, especially if outcome data have already been systematically collected (i.e., vital statistics). However, because individual-level data are not collected, it is not shown whether those with better disease outcomes are actually the ones who participated in screening.

Case-control studies can be used to compare cases dying from the disease or having an otherwise "poor"

outcome (e.g., cancer that has metastasized) with controls drawn from all members of the source population without the outcome (death or advanced disease). This control group can include persons with the disease who have not developed the outcome. Exposure status is then defined by screening history. Another strategy for study design is to classify newly diagnosed cases as exposed or not according to screening history and then followed up to death or other defined outcome.

Several types of bias may arise in observational studies of screening. Selection bias may occur because those who choose to participate in screening are different from those who refuse and may tend to have other behaviors in common that affect disease risk. *Lead time bias* happens when a survival time advantage appearing in screen-detected cases is actually because they have by definition been diagnosed earlier in the course of illness than the cases detected due to symptomatic disease. *Length bias* can occur if cases with a long preclinical detectable period tend to have a slower-progressing or less fatal disease variant. Since these cases are also more likely to be detected through screening, the screen-detected group of cases may have better outcomes even if treatment has no effect. *Overdiagnosis* involves disease diagnosis, in the screening group only, of persons truly without disease or with disease that would never become symptomatic within the natural lifespan. This has been advanced as an explanation for cancer screening trials that have found more disease in screened groups but then failed to find any difference in mortality.

Due to the biases inherent in observational studies, randomized trials, where subjects are randomly assigned to be screened (or not), are typically considered the optimal study design. However, a large sample size and long follow-up period may be required to observe a sufficient number of events. Additionally, feasibility is limited to situations where participants and their physicians are willing to go along with randomization. This may not be the case if a test has become generally accepted, even if its effectiveness is not proven.

—Keely Cheslack-Postava

See also Negative Predictive Value; Newborn Screening Programs; Positive Predictive Value; Preclinical Phase of Disease; Prevention: Primary, Secondary, and Tertiary; Sensitivity and Specificity

Further Readings

Gordis, L. (2004). The epidemiologic approach to the evaluation of screening programs. In *Epidemiology* (3rd ed., pp. 281–300). Philadelphia: Elsevier Saunders.

Morabia, A., & Zhang, F. F. (2004). History of medical screening: From concepts to action. *Postgraduate Medical Journal*, *80*, 463–469.

Morrison, A. S. (1998). Screening. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 499–518). Philadelphia: Lippincott Williams & Wilkins.

Parkin, D. M., & Moss, S. M. (2000). Lung cancer screening. Improved survival but no reduction in deaths: The role of "overdiagnosis." *Cancer*, *89*, 2369–2376.

U.S. Preventive Services Task Force. (2004, July). *Screening for syphilis infection: Recommendation statement*. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved December 8, 2006, from http://www.ahrq.gov/clinic/3rduspstf/syphilis/syphilrs.htm.

Wilson, J. M. G., & Jungner, G. (1968). *Principles and practice of screening for disease.* Geneva, Switzerland: World Health Organization.

Secondary Data

Researchers in epidemiology and public health commonly make a distinction between *primary data*, data collected by the researcher for the specific analysis in question, and *secondary data*, data collected by someone else for some other purpose. Of course, many cases fall between these two examples, but it may be useful to conceptualize primary and secondary data by considering two extreme cases. In the first case, which is an example of *primary data*, a research team collects new data and performs its own analyses of the data so that the people involved in analyzing the data have some involvement in, or at least familiarity with, the research design and data collection process. In the second case, which is an example of *secondary data*, a researcher obtains and analyzes data from the Behavioral Risk Factor Surveillance System (BRFSS), a large, publicly available data set collected annually in the United States. In the second case, the analyst did not participate in either the research design or the data collection process, and his or her knowledge of those processes come only from the information available on the BRFSS Web site and from queries to BRFSS staff.

Secondary data are used frequently in epidemiology and public health, because those fields focus on monitoring health at the level of the community or

nation rather than at the level of the individual as is typical in medical research. In many cases, using secondary data is the only practical way to address a question. For instance, few if any individual researchers have the means to collect the data on the scale required to estimate the prevalence of multiple health risks in each of the 50 states of the United States. However, data addressing those questions have been collected annually since the 1980s by the Centers for Disease Control, in conjunction with state health departments, and it is available for download from the Internet. Federal and state agencies commonly use secondary data to evaluate public health needs and plan campaigns and interventions, and it is also widely used in classroom instruction and scholarly research.

There are both advantages and disadvantages to using secondary data. The advantages relate primarily to the fact that an individual analyst does not have to collect the data himself or herself and can obtain access through a secondary data set information much more wide-ranging than he or she could collect alone. Specific advantages of using secondary data include the following:

- *Economy*, because the analyst does not have to pay the cost of data collection
- *Speed and convenience*, because the data are already available before the analyst begins to work
- *Availability of data from large geographic regions*, for instance, data collected on the national or international level
- *Availability of historical data and comparable data collected over multiple years*, for instance, the BRFSS data are available dating back to the 1980s, and certain topics have been included every year
- *Potentially higher quality of data*, for instance, the large surveys conducted by federal agencies such as the Centers for Disease Control commonly use standardized sampling procedures and professional interviewers in contrast to many locally collected data sets that represent a convenience sample collected by research assistants.

The disadvantages of using secondary data relate primarily to the potential disconnect between the analyst's interests and the purposes for which the data were originally collected and the analyst's lack of familiarity with the original research design and data collection and cleaning processes. These disadvantages include the following:

- *Specific research questions may be impossible to address through a secondary data set*, either because the relevant data were not collected (because it was not germane to the original research project) or because it was suppressed due to confidentiality concerns (e.g., home addresses of respondents).
- *Data may not be available for the desired geographic area or time period.* This is of particular concern to researchers studying a small geographic area, such as a neighborhood within a city, or to those who are interested in specific time periods, such as immediately before and after a particular historic event.
- *Different definitions or categories* than what the analyst desires may have been used for common constructs such as race and ethnicity, and certain constructs may not have been recognized at the time of the survey—for instance, same-sex marriage.
- *Potential lack of information about the data collection and cleaning process* may leave the analyst uninformed about basic concerns such as the response rate or the quality of the interview staff. In some cases, some of this information is available, in others it is not, but in any case every research project has its idiosyncrasies and irregularities that affect data quality and the lack of information about these details may lead the analyst to reach inappropriate conclusions using the data.

Primary and secondary data analyses are not in competition with each other. Each has its place within the fields of epidemiology and public health, and the most useful approach is to choose the data to be analyzed for a particular project based on what is most appropriate for the primary research questions and based on the resources available.

There are many sources of secondary data in epidemiology and public health that are easily accessible. The best-known sources are the data from large-scale governmental-sponsored surveys, such as the BRFSS, which are available on the Internet and may be accessed through the CDC Wonder Web site or through the Web site of the relevant agency. Claims records and other data relating to the Medicare and Medicaid systems are also available, with certain restrictions. Secondary data may also be accessed through clearinghouses that collect data from other sources, including private researchers, and make it available for use, such as that of the Interuniversity Consortium for Social and Political Research located at the University of Michigan. Vital statistics data

(records of births, deaths, marriages, divorces, and fetal deaths) are often available at the local or state level, and some of this information is available at the national level as well. Access to private administrative data, for instance, claims records of insurance companies, must be negotiated with the company in question.

—Sarah Boslaugh

See also Centers for Disease Control and Prevention; National Center for Health Statistics

Articles on individual secondary data sets and sources of data can be located through the Reader's Guide.

Further Readings

Boslaugh, S. (2007). *Secondary data sources for public health: A practical guide*. New York: Cambridge University Press.

Web Sites

Centers for Disease Control Wonder: http://wonder.cdc.gov.
Interuniversity Consortium for Social and Political Research: http://www.icpsr.umich.edu.

Self-Efficacy

In 1977, Albert Bandura introduced the concept of self-efficacy, which is defined as the conviction that one can successfully execute the behavior required to produce a specific outcome. Unlike efficacy, which is the power to produce an effect (i.e., competence), self-efficacy is the belief that one has the power to produce that effect. Self-efficacy plays a central role in the cognitive regulation of motivation, because people regulate the level and the distribution of effort they will expend in accordance with the effects they are expecting from their actions. Self-efficacy is the focal point of Bandura's social-cognitive theory as well as an important component of the health belief model.

Theories and models of human behavior change are used to guide health promotion and disease prevention efforts. Self-efficacy is an important psychosocial concept for epidemiologists to understand because it influences study participants' behavior, intervention uptake, and potential for long-term program maintenance. Ignoring the role that psychosocial concepts, such as self-efficacy, play in intervention efforts may cause study effects to be misinterpreted and project results to be misattributed. For instance, individuals participating in a smoking cessation program may be given solid smoking cessation strategies, social support, and alternative stress reduction activities, but if they do not believe that they can stop smoking—that is, if they lack self-efficacy—the program will be less successful. On the other hand, if a smoking cessation expert recognizes that overcoming an individual's lack of self-efficacy to stop smoking is critical to his or her quitting, the program could readily be designed with this factor in mind, and the program evaluation could determine the success of raising self-efficacy. Understanding self-efficacy can shed light on the determinants of health and disease distributions.

Factors Influencing Self-Efficacy

Bandura points to four sources affecting self-efficacy—experience, modeling, social persuasions, and physiological factors. Experiencing mastery is the most important factor for deciding a person's self-efficacy; success raises self-efficacy, failure lowers it. During modeling, an individual observes another engage in a behavior; when the other succeeds at the behavior, the observing individual's self-efficacy will increase—particularly if the observed person is similar in meaningful ways to the person doing the observation. In situations where others are observed failing, the observer's self-efficacy to accomplish a similar task will decrease. Social persuasions relate to encouragement and discouragement. These can be influential—most people remember times where something said to them severely altered their confidence. Positive persuasions generally increase self-efficacy, and negative persuasions decrease it. Unfortunately, it is usually easier to decrease someone's self-efficacy than it is to increase it. Physiologic factors play an important role in self-efficacy as well. Often, during stressful situations, people may exhibit physical signs of discomfort, such as shaking, upset stomach, or sweating. A person's perceptions of these responses can markedly alter his or her self-efficacy. If a person gets "butterflies in the stomach" before public speaking, a person with low self-efficacy may take this as a sign of his or her inability, thus decreasing efficacy further. Thus, it is how the person interprets the

physiologic response that affects self-efficacy rather than the physiologic response per se.

Self-Efficacy: Influences on Beliefs and Behavior

Self-efficacy can enhance human accomplishment and well-being in numerous ways. It may influence the *choices* people make because individuals tend to select tasks in which they feel competent and avoid those in which they are not. Self-efficacy may also determine how much effort to expend, how long to persevere, and how resilient to be when faced with adverse situations. The higher the sense of efficacy, the greater the effort, persistence, and resilience a person will generally demonstrate. Self-efficacy can also influence an individual's thought patterns and emotional reactions. High self-efficacy helps create feelings of calm or competence in approaching difficult tasks and activities. Conversely, people with low self-efficacy may believe that things are tougher than they really are, a belief that fosters anxiety and stress and may serve to narrow problem-solving capacity. Self-efficacy can also create a type of self-fulfilling prophecy in which one accomplishes only what one believes one can accomplish. It is not unusual for individuals to overestimate or underestimate their abilities; the consequences of misjudgment play a part in the continual process of efficacy self-appraisals. When consequences are slight, individuals may not need to reappraise their abilities and may continue to engage in tasks beyond their competence. Bandura argued that strong self-efficacy beliefs are the product of time and multiple experiences; therefore, they are highly resistant and predictable. Weak self-efficacy beliefs, however, require constant reappraisal. Both, of course, are susceptible to a powerful experience or consequence.

—Lynne C. Messer

See also Health Behavior; Health Belief Model; Intervention Studies

Further Readings

Glanz, K., Lewis, F. M., & Rimer, B. K. (Eds.). (1997). *Health behavior and health education.* San Francisco: Jossey-Bass.

Sensitivity and Specificity

In medicine, diagnostic tests are administered to patients to detect diseases so that appropriate treatments can be provided. A test can be relatively simple, such as a bacterial culture for infection, or a radiographic image to detect the presence of a tumor. Alternatively, tests can be quite complex, such as using mass spectrometry to quantify many different protein levels in serum and using this "protein profile" to detect disease. Screening tests are special cases of diagnostic tests, where apparently healthy individuals are tested with the goal of diagnosing certain conditions early, when they can be treated most effectively and with higher success. For example, in the United States, annual screening mammograms are recommended for healthy women aged 40 years and older to facilitate early detection of breast cancer.

Whether for diagnosis or screening, the objective is to use some kind of a test to correctly classify individuals according to their true disease state. Sensitivity and specificity are two related measures used to quantify the performance of a screening or diagnostic test compared with the true condition or disease state. The test yields a binary result for each individual, positive for the presence of a condition or negative for its absence, which is compared with the true disease state for the same individuals. In practice, since the true disease state cannot always be identified with absolute certainty, a gold standard test that also yields a binary result (presence or absence of a condition) is considered the true status for an individual. For example, screening tests conducted among pregnant women to measure the chance of a child having birth defects usually take results from an amniocentesis as the gold standard, since the true status of the infant may not be determined until birth. In most cases, the amniocentesis results reflect the truth with extremely high probability.

A comparison of diagnostic or screening test results with the true disease state (or gold standard test results) can be displayed in a 2×2 table, as in Table 1.

Sensitivity, calculated as $a/(a + c)$, gives the proportion of individuals who truly have the disease for whom the test gives a positive result. Sensitivity is also sometimes referred to as the "true positive fraction" or "true positive rate." That is, among those individuals in whom the condition is truly present, sensitivity reflects how often the test detects it.

Table 1 True Disease State and Screening Test Result

	True Disease State or Gold Standard Result		
Screening Test Result	*Positive (D +)*	*Negative (D–)*	*Total*
Positive (T +)	a	b	$a+b$
Negative (T–)	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

Notes: Cell a = number of true positives; Cell b = number of false positives; Cell c = number of false negatives; Cell d = number of true negatives.

Specificity, calculated as $d/(b+d)$, gives the proportion of those individuals who truly do not have the disease for whom the test gives a negative result—the true negatives. That is, for those individuals in whom the disease or condition is truly absent, specificity reflects how often the test gives a negative result. Closely related to specificity is the proportion of false-positive test results, which is calculated as $b/(b+d)$ or 1 − specificity. In the health sciences literature, sensitivity, specificity, and false-positive rate are usually reported as a percentage rather than as proportions (i.e., they are multiplied by 100).

Sensitivity and specificity are also interpretable as conditional probabilities. Sensitivity is the probability of a case having a positive test given that the case truly has the disease. Sensitivity is the probability of a case having a negative test given that the disease is truly absent. Both can be represented mathematically as

$$\text{Sensitivity} = \Pr(T+|D+),$$
$$\text{Specificity} = \Pr(T-|D-),$$

where T represents the test result (either positive or negative) and D represents the true disease state (either with or without disease). Because these are each probabilities for a given disease state, they are not dependent on the prevalence (probability) of the disease in a population. In the context of the cells displayed in Table 1, this means that sensitivity and specificity do not require that $(a+c)/(a+b+c+d)$ and $(b+d)/(a+b+c+d)$ represent the proportions of those with and without disease in the population of interest. For example, one could still estimate sensitivity and specificity for a test that detects a rare condition even if equal numbers of individuals with and without the condition were assessed. In fact, the strategy of enrolling equal numbers of $D+$ and $D-$ individuals is quite common in some of the early stages of diagnostic test development.

Ideally, a screening or diagnostic test would maximize both sensitivity and specificity. However, there is usually a trade-off between the two measures for any given test. For many tests, the result is a numeric value rather than simply positive or negative for a disease. A culture may provide bacterial counts or a laboratory assay may provide serum protein concentration levels, for example. Still other tests may be reported as ordinal categories, such as breast cancer tumors, which are classified as Stage 0, 1, 2, 3, or 4 based on their size, cell types, and lymph node involvement. These continuous or ordinal results may be translated to a binary classification of "positive" or "negative" for a disease by establishing a cutpoint such that values on one side of the cutpoint (usually higher) are interpreted as positive and values on the other side (usually lower) are interpreted as negative. In the case of tumor staging, a binary classification may be used to designate "severe" versus "less severe" forms of cancer. Body mass index (BMI), which is computed by dividing an individual's weight (in kg) by the individual's height (in m) squared, is a widely used method to diagnose whether a person is overweight or obese. Typically, individuals with BMI values between 25 and 29.9 are considered overweight, while those with BMI values of 30 or greater are considered obese. It is important to note that BMI is an indirect measure of true body composition—a complex relationship between lean mass (i.e., muscle) and fat mass. However, lean and fat mass measures are difficult to obtain compared with weight and height. Therefore, BMI, when used in the context of overweight and

obesity, is used as an easily obtainable measure of body fat, as is frequently used to diagnose overweight and obesity.

Altering a cutpoint that designates a positive versus a negative result on a diagnostic test will alter the sensitivity and specificity of that test. If a cutpoint is lowered for a test in which a score over the cutpoint is interpreted as having the disease, the test may identify more people with the disease with a positive test (increased sensitivity), but included in those positive tests will be a greater number of people without the disease (higher proportion of false positives, which implies decreased specificity). It is estimated that approximately 65% of people in the United States are overweight or obese when the BMI > 25 definition is used. If we lower the cutpoint from 25 to, say, 24, we would classify more individuals as overweight. Using this classification, we may correctly classify some individuals as overweight and who were missed using the standard classification, increasing the specificity. However, we may also increase the number of false positives; that is, we may mistakenly classify some individuals who have high lean mass, rather than fat mass, relative to their height as overweight. On the other hand, if we increased the cutpoint to diagnose overweight to 27 or 28, we would classify fewer people as overweight and have fewer false-positive results. However, we would potentially miss diagnosing some overweight individuals who could benefit from treatments or lifestyle changes that may help lower body fat and reduce risk for future medical conditions related to overweight and/or obesity.

When considering what the ideal sensitivity and specificity of a particular diagnostic or screening test should be, one must consider the "costs" (not necessarily strictly monetary) of failing to identify someone who truly has the disease compared with the "cost" of falsely identifying a healthy person as having the disease. When the consequence of missing a person with the disease is high, as is the case with rapidly fatal or extremely debilitating conditions, then sensitivity should be maximized, even at the risk of decreasing specificity. Falsely identifying someone as diseased in this instance has less dire consequences than failing to identify a diseased person. However, when the consequence of wrongly classifying someone as diseased is greater than the risk of failing to identify a diseased person, as in the case of a slowly progressing but highly stigmatizing condition, then one would prefer to maximize specificity even in the face of decreasing sensitivity.

Predicted Values

Once test results are known, a patient or medical professional may wish to know the probability that the person truly has the condition if the test is positive or is disease free if the test is negative. These concepts are called positive and negative predicted values of a test and can be expressed probabilistically as $\Pr(D+|T+)$ and $\Pr(D-|T-)$. Unlike sensitivity and specificity, positive and negative predicted values cannot always be computed from a fourfold table such as Table 1, as they are dependent on the prevalence of disease in the population of interest. Only if $(a+c)/(a+b+c+d)$ and $(b+d)/(a+b+c+d)$ accurately reflect the proportions of those with and without disease in the population can the fourfold table can be used to estimate the predicted values. In such cases where this holds true, the positive predicted value (PPV) can be computed as $a/(a+b)$, and the negative predicted value (NPV) can be computed as $c/(c+d)$. However, the PPV and the NPV can always be obtained by using Bayes's theorem. If one can obtain estimates of sensitivity and specificity for a particular test, and, from another source, estimate the prevalence of disease in a population of interest, Bayes's theorem yields the following:

$$\text{PPV} = \frac{\text{Sensitivity} \times \text{Prevalence}}{(\text{Sensitivity} \times \text{Prevalence}) + (1 - \text{Sensitivity}) \times (1 - \text{Prevalence})}$$

and

$$\text{NPV} = \frac{\text{Sensitivity} \times (1 - \text{Prevalence})}{\text{Sensitivity} \times (1 - \text{Prevalence}) + (1 - \text{Sensitivity}) \times (1 - \text{Prevalence})}.$$

All else being equal, the lower the prevalence of disease (i.e., rare conditions), the lower the proportion of true positives among those who test positive; that is, the PPV will be lower if a condition is rare than if it is common, even if the test characteristics, sensitivity and specificity, are high. Screening pregnant women for a rare condition such as Down syndrome (trisomy 21), whose prevalence in the United States is approximately 9.2 per 10,000 live births (0.0092), results in more false positive than true positive tests despite the fact that the screening tests have adequate sensitivity (75% or higher) and excellent specificity (95% or higher).

Receiver Operating Characteristic Curves

As previously noted, binary classifications are frequently derived from numerical or ordinal test results via a cutoff value. When this is the case, the performance of different cutoff values can be assessed using a receiver operating characteristic (ROC) curve. ROC curves are plots of the true-positive fraction of a test (sensitivity) versus the false-positive fraction (1 − specificity) across the entire spectrum of observed test results. An example of an ROC curve for a simulated diagnostic test that is measured on a continuum is presented in Figure 1. Test 1 (bold line in Figure 1) has higher sensitivity for certain false-positive fractions than Test 2 (dotted line). However, both tests have low, but comparable, sensitivity when false-positive fractions are 0.10 and lower.

The area under the ROC curve can range from 0.5, which represents results from an uninformative test, whose results are the same as flipping a coin to make a diagnosis, to 1.0, which represents a perfectly discriminative test with 100% sensitivity and 100% specificity. The area under the ROC curve for BMI as a test for obesity (obtained via direct measures of body fat) is in the range of 0.8 for adult men and above 0.9 for adult women.

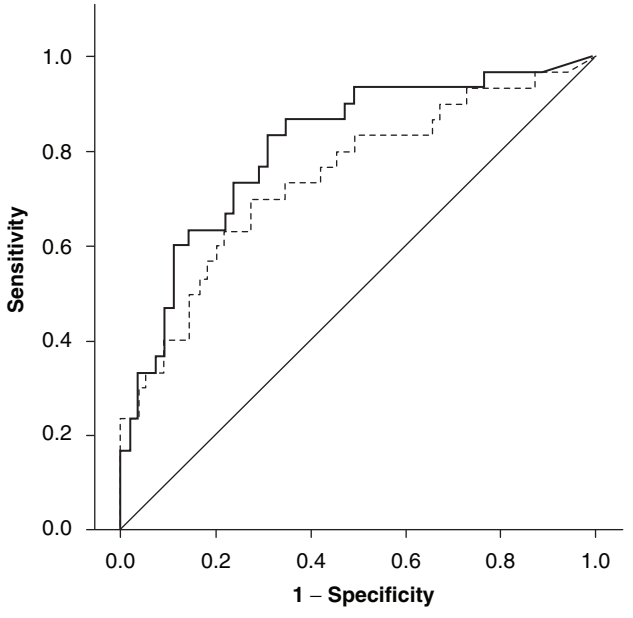


Figure 1 Receiver Operating Characteristic Curve for Simulated Diagnostic Tests That Are Reported as a Numerical Value

Note: Diagonal reference line represents an uninformative test.

The areas under ROC curves are an appropriate summary index to compare performance of the multiple diagnostic tests for the same condition. The area under the ROC for Test 1 in Figure 1 is 0.81, but only 0.74 for Test 2, leading us to prefer Test 1 over Test 2. Like sensitivity, specificity, and predicted values, the area under the ROC curve also has a probabilistic interpretation—the probability that a randomly selected pair of individuals with and without disease or condition is correctly classified. Thus, ROC curves, which are simply graphical representations of sensitivity and specificity, are extremely helpful in illustrating the trade-off between increasing sensitivity and increasing the proportion of false positives (or decreasing specificity).

—Jodi Lapidus

See also Bayes's Theorem; Clinical Epidemiology; Evidence-Based Medicine; Receiver Operating Characteristic (ROC) Curve

Further Readings

Loong, T. W. (2003). Understanding sensitivity and specificity with the right side of the brain. *British Medical Journal*, *327*, 716–719.

Rosner, B. (1999). *Fundamentals of biostatistics* (5th ed., chap. 3). Pacific Groove, CA: Duxbury Press.

Sentinel Health Event

A sentinel health event is the occurrence of a specific health condition, or the occurrence of an adverse outcome to a medical intervention, which signals that (a) the incidence of a condition or disease has exceeded the threshold of expected cases, (b) changes are occurring in the health levels of a population, or (c) the quality of medical care may need to be improved. Each of these circumstances may indicate the presence of a broader public health problem, and each requires epidemiologic investigation and intervention.

For example, the occurrence of a single case of smallpox anywhere in the world would signal the recurrence of a disease that has been eradicated in the wild, indicating that a highly unusual and unexpected event has occurred that requires immediate

investigation. Likewise, an increase in the incidence of birth defects in a community that has had a stable incidence over time also indicates that the health level of that population has changed, possibly due to a change in the environment or due to a common exposure, which also warrants investigation. And when physicians began seeing cases of Kaposi's sarcoma, a very rare cancer seen almost exclusively in older men of Mediterranean or eastern European extraction, occurring in young American men, this was the harbinger of the presence of an entirely new condition, human immunodeficiency virus infection.

Sentinel health events have also been defined for use in monitoring health care settings for potential problems in health care quality. In this setting, the sentinel health event is an illness, disability, or death that was otherwise preventable or avoidable, except for a problem in health care delivery or quality. An increase in postoperative infections in a particular hospital or specific ward may indicate a problem involving individual operating room staff, ventilation systems, sterilization equipment, or other systems involved in the treatment and care of surgical patients.

To be able to identify when conditions or occurrences are unusual, good baseline information is required. Public health surveillance systems are used to gather such baseline data for a given population and facilitate the monitoring of sentinel health events.

—Annette L. Adams

See also Applied Epidemiology; Bioterrorism; Epidemic; Field Epidemiology; Notifiable Disease; Outbreak Investigation; Public Health Surveillance

Further Readings

Gregg, M. B. (Ed). (2002). *Field epidemiology* (2nd ed.). New York: Oxford University Press.

Teutsch, S. M., & Churchill, R. E. (2000). *Principles and practice of public health surveillance* (2nd ed.). New York: Oxford University Press.

Sequential Analysis

Sequential analysis refers to a statistical method in which data are evaluated as they are collected, and further sampling is stopped in accordance with a predefined stopping rule as soon as significant results are observed. This contrasts to classical hypothesis testing where the sample size is fixed in advance. On average, sequential analysis will lead to a smaller average sample size compared with an equivalently powered study with a fixed sample size design and, consequently, lower financial and/or human cost.

Sequential analysis methods were first used in the context of industrial quality control in the late 1920s. The intensive development and application of sequential methods in statistics was due to the work of Abraham Wald around 1945. Essentially, the same approach was independently developed by George Alfred Barnard around the same time.

Interestingly, sequential analysis has not been frequently used in epidemiology despite the attractive feature of allowing the researcher to obtain equal statistical power at a lower cost. In 1962, Lila Elveback predicted a great increase in the application of sequential methods to problems in epidemiology in the coming few years; however, the prediction has not come true. The reluctance to apply sequential methods might be attributed to the fact that making a decision following every observation is complicated. Furthermore, in epidemiological studies, complex associations among outcome variable and predictor variables (rather than simply the primary outcome) are often a major interest and need to be determined in a relatively flexible multivariable modeling framework in light of all available data, while sequential analysis usually requires a well-defined and strictly executed design. Nevertheless, sequential analysis could be appropriately applied to some of the epidemiology research problems where the data are monitored continuously, such as in the delivery of social and health services and disease surveillance.

Wald's Sequential Probability Ratio Test

Sequential analysis is a general method of statistical inference. Wald's sequential probability ratio test (SPRT) is one of the most important procedures for hypothesis testing in sequential analysis. Its application is suitable for continuous, categorical, or time-to-event data, and the test includes sequential t tests, F tests, or χ^2 tests, among others. Generally, the test is performed each time a new observation is taken. At each step, the null hypothesis is either rejected or accepted, or based on predefined criteria, the study continues by taking

one more observation without drawing any conclusions. In practice, Wald's SPRT need not be started with the first observation, but after a certain number of observations have been taken, since small sample size often does not provide enough evidence to reject or accept the null hypothesis. Sequential estimation procedures have also been developed to allow the estimation of confidence intervals in sequential sampling. Since binomial data are frequently encountered in public health and epidemiology applications, Wald's SPRT for binomial proportions is shown here.

Suppose that null hypothesis $H_0: p = p_0$ versus alternative hypothesis $H_1: p = p_1\ (> p_0)$ is tested here. The criterion for accepting or rejecting the null hypothesis is given by two parallel straight lines. The lines are functions of p_0, p_1, Type I error (α), Type II error (β), and the number of total observations to date:

$$d_l = \beta_{01} + \beta_1 N \text{ (lower line)},$$

$$d_u = \beta_{02} + \beta_1 N \text{ (upper line)},$$

where N is the number of total observations, and

$$\beta_1 = \frac{\log\left(\frac{1-p_0}{1-p_1}\right)}{\log\left[\left(\frac{p_1}{p_0}\right)\left(\frac{1-p_0}{1-p_1}\right)\right]},$$

$$\beta_{01} = -\frac{\log\left(\frac{1-\alpha}{\beta}\right)}{\log\left[\left(\frac{p_1}{p_0}\right)\left(\frac{1-p_0}{1-p_1}\right)\right]},$$

and

$$\beta_{02} = \frac{\log\left(\frac{1-\beta}{\alpha}\right)}{\log\left[\left(\frac{p_1}{p_0}\right)\left(\frac{1-p_0}{1-p_1}\right)\right]}.$$

The power of a test is defined as $1 - \beta$. The repeated testing of hypothesis was incorporated in the construction of Wald's SPRT so the Type I error is preserved at the level of α. The commonly accepted value for α is 0.05; for β, 0.10 or 0.20 is typically used, which translates to 90% or 80% power. At each new observation, d_l and d_u are calculated based on prespecified values of p_0, p_1, α, and β. Denote the number of cases as d, then if $d \le d_l$, the null hypothesis is accepted; if $d \ge d_u$, the null hypothesis is rejected; otherwise, sampling is continued, until the null hypothesis is rejected or accepted. The regions are illustrated in Figure 1. The smaller the difference between p_0 and p_1, the greater the number of observations needed.

One example of using Wald's SPRT is monitoring of a breast screening program for minority women to determine whether the program was reaching its target population. The information was to be collected over time, and it was the researcher's goal to have 95% of the screened women to be African American since 95% of the target area residents were African American. If not and if at least 90% of the screened women were African American, it would indicate that the women in the target population were not being adequately reached. For this problem, the hypotheses could be set up as follows:

> H_0: the program adequately reached its target population ($p_0 = 5\%$, proportion of non–African American women screened).
>
> H_1: the program did not adequately reach its target population ($p_1 = 10\%$).

Note that $p_1 > p_0$. Given $\alpha = 0.05$ and $\beta = 0.10$, Wald's SPRT could be performed based on the methods described above, and the study continued until enough women were observed to draw a conclusion.

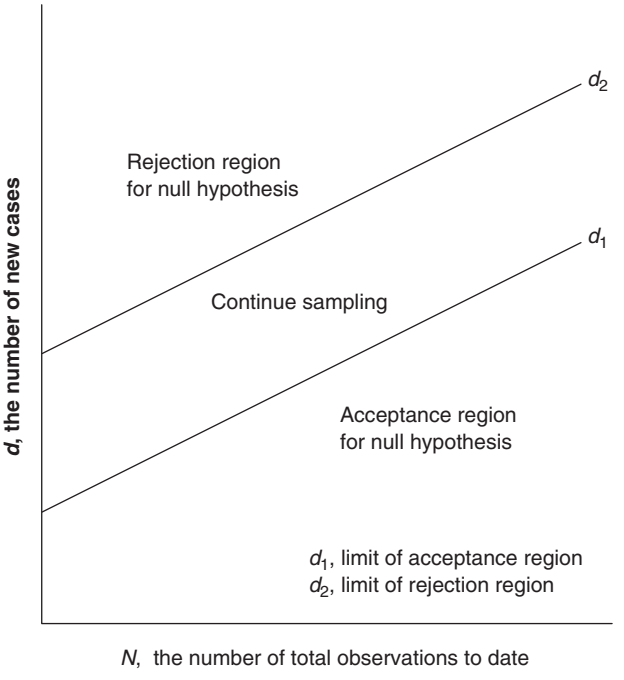


Figure 1 Graphical Illustration of Wald's Sequential Probability Ratio Test

Group Sequential Design

Sequential methods seemed initially to be an attractive solution to achieve the scientific and ethical requirements of clinical trials. However, sequential analysis was designed to perform a hypothesis testing after data on each new case were collected, which do not work well for clinical trials. A modified sequential method, called *group sequential design*, was developed for clinical trials, in which the accruing efficacy data are monitored at administratively convenient intervals and treatments are compared in a series of interim analyses. Important decisions concerning the future course of the study are made along the way, such as early stopping for futility or benefit, sample size readjustment, or dropping ineffective arms in multiarm dose-finding studies. Investigators may also stop a study that no longer has much chance of demonstrating a treatment difference. This method has now been routinely used in clinical trials. A well-known example is the hormone therapy trials of Women's Health Initiatives, where two treatment arms were stopped earlier because it was decided that the risks exceeded the benefits.

—Rongwei (Rochelle) Fu

See also Clinical Trials; Hypothesis Testing; Type I and Type II Errors

Further Readings

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. New York: Cambridge University Press.

Howe, H. L. (1982). Increasing efficiency in evaluation research: The use of sequential analysis. *American Journal of Public Health*, *72*, 690–697.

Wald, A. (1946). Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, *16*, 117–186.

Whitehead, J. (1997). *The design and analysis of sequential clinical trials* (2nd Rev. ed.). Chichester, UK: Wiley.

Severe Acute Respiratory Syndrome (SARS)

Severe acute respiratory syndrome (SARS) is a severe atypical pneumonia resulting from infection with a novel coronavirus, the SARS coronavirus (SARS-CoV). SARS is thought to have first occurred in humans in Guangdong province, China, in November 2002. International recognition of SARS occurred in March 2003 after many persons infected at a Hong Kong hotel seeded outbreaks of SARS in several countries and areas within days of each other. On March 12, 2003, the World Health Organization (WHO) issued a global alert regarding an acute respiratory syndrome of unknown aetiology and 3 days later issued an emergency travel advisory. By July 2003, cases of SARS had been identified in 29 countries and areas on most continents resulting in at least 774 deaths and 8,096 probable cases of SARS reported to the WHO. Global collaboration, coordinated by the WHO, was vital for the rapid identification of the causative agent, and prompt detection and isolation of cases, strict adherence to infection control procedures, vigorous contact tracing, and implementation of quarantine measures proved effective in containing the global outbreak.

The emergence of SARS provided a major scientific and public health challenge and resulted in rapid scientific achievements, a speedy epidemiological response, and international collaboration. Although SARS appears to be contained, the threat of another global outbreak remains as SARS may reemerge from unidentified animal reservoirs, laboratories that store the virus, or via undetected transmission within the human population. The epidemiology of SARS serves as a reminder of the global nature of infectious diseases and their continuing threat. The lessons learned provide a useful template for future international public health preparedness and response strategies.

Global Spread

The rapid global spread of SARS from the southern Chinese province of Guangdong was amplified by international air travel and "superspreading" events. A superspreading event may be defined as a transmission event whereby one individual with SARS infects a large number of persons. In February 2003, an infected physician from Guangdong who had treated SARS patients flew to Hong Kong and stayed in the Metropole Hotel, where he infected at least 14 hotel guests and visitors. Several of these infected individuals subsequently seeded outbreaks in Hanoi, Hong Kong, Singapore, Toronto, and elsewhere. Such superspreading events were reported from all sites with sustained local transmission before the implementation of strict hospital infection control measures. Although several theories have been proposed to explain why superspreading events occurred, such as

patient immune status, underlying disease, higher level of viral shedding at the peak of infection or other environmental factors, a definitive reason has not been identified, and further research is required to explain this phenomenon.

On July 5, 2003, the WHO declared that the last known chain of human-to-human transmission had been broken and that the global SARS outbreak was contained. By this time, the global cumulative total of probable cases was 8,096, with 774 deaths. Mainland China, Hong Kong, and Taiwan accounted for 92% of all cases and 89% of all deaths. Globally, most cases (21%) occurred in health care workers and their close contacts; however, secondary community transmission did occur in particular—the Amoy Gardens outbreak in Hong Kong in March 2003 where more than 300 residents were infected.

In the current postoutbreak period, 17 SARS cases have been reported. Six cases of laboratory-acquired infection have been reported from China, Singapore, and Taiwan. One of the laboratory-acquired cases in China resulted in seven additional cases. In addition to the laboratory-acquired cases, a further four community-acquired cases were reported from Guangdong province, China.

Origin of SARS

The actual reservoir of SARS-CoV in nature is unknown; however, it is suggested that SARS-CoV originated from a wild animal reservoir in mainland China, supported by a number of factors. Masked palm civets and raccoon dogs in animal markets in China had a SARS-CoV almost identical to that seen in SARS patients. Additionally, more than one third of the early SARS patients in Guangdong were involved in either the trade or preparation of food from wild animals in markets. Furthermore, there was a much higher seroprevalence of SARS-CoV among wild animal handlers than among controls in Guangdong. Further surveillance on animals is needed to help understand the reservoir in nature that led to the SARS outbreak.

Transmission of SARS

The incubation period for SARS is between 3 and 10 days, with a median of 4 to 5 days. However, it may be as long as 14 days. The SARS-CoV is transmitted predominately through droplets from the respiratory tract of the infected person, particularly when coughing, sneezing, and even speaking. The risk of transmission is highest when there is close face-to-face contact. SARS-CoV transmission is believed to be amplified by aerosol-generating procedures such as intubation or the use of nebulizers. The detection of SARS-CoV in fecal as well as respiratory specimens indicates that the virus may be spread by both fecal contamination and via respiratory droplets. SARS-CoV also has the ability to survive on contaminated objects in the environment for up to several days and therefore transmission may occur via fomites. Although SARS spreads rapidly around the world as a result of international travel, relatively few cases were acquired by this route.

Efficient environments for transmission of the SARS-CoV were health care facilities and households, leading to a preponderance of SARS cases who were either health care workers or household contacts of cases. Several risk factors may account for this. In the health care setting, close contact is required to care for severely ill patients. Additionally, efficiency of transmission appears to be directly related to the severity of the illness, and those more severely ill are more likely to be hospitalized. Furthermore, patients with SARS-CoV appear to be most infectious during the second week of their illness, and although transmission can also occur during the first week, they are more likely to present to hospital as their clinical condition worsens, usually during the second week. In Singapore, the secondary attack rate among household members of SARS cases was approximately 6%. The risk factors associated with household transmission included older age of the index case and non–health care occupation of the index case. Interestingly, there were no reports of transmission from infected children to other children or to adults.

Diagnostic Criteria and Treatment

The WHO issued updated case definitions for SARS during the outbreak using a combination of clinical signs and symptoms together with epidemiologic factors to assist in the identification of hospital cases. These have been revised in the postoutbreak period and include radiographic and laboratory findings. To date no consistently reliable rapid test is available for SARS-CoV.

Sudden onset of fever is the most common initial symptom of SARS and may also be associated with

headache, myalgia, malaise, chills, rigor, and gastrointestinal symptoms. However, documented fever did not occur in some cases, particularly the elderly.

Although a variety of treatment protocols have been tested, at present sufficient evidence is not available to recommend any specific therapy for the treatment of SARS.

Prognostic Factors

Factors associated with a poor prognosis or outcome (i.e., admission to an intensive care unit or death) included advanced age, coexisting illness such as diabetes or heart disease, and the presence of elevated levels of lactate dehydrogenase or a high neutrophil count on admission. Clinically, compared with adults and teenagers, SARS-CoV infection was less severe in children (< 12 years) and had a more favorable outcome. Additionally, preterm and term infants born to women infected with SARS-CoV were not found to be clinically affected or shedding the virus after birth.

Prevention and Control

In the absence of a vaccine, the most effective way to control a viral disease is to break the chain of human-to-human transmission. This was achieved for SARS using the traditional public health methods of early case detection and isolation, strict infection control measures, contact tracing, and quarantine measures. The SARS outbreak also clearly demonstrated the effectiveness of international collaboration.

Early identification of SARS cases is crucial in making it possible to initiate appropriate precautions, and studies of transmission show this to be a critical component in controlling any future outbreaks of SARS. However, clinical symptoms alone are not enough to diagnose SARS, especially as these are nonspecific in the early stages and cases may present during seasonal outbreaks of other respiratory illnesses. Furthermore, SARS-CoV laboratory tests have limited sensitivity when used in the early stages of illness. However, most SARS cases could be linked to contact with SARS patients or a place where SARS transmission was known or suspected. Transmission was also interrupted by the use of stringently applied infection control precautions, including the appropriate use of personal protective equipment. The relatively long incubation period of SARS facilitated contact tracing, the implementation of quarantine measures, and the institution of control measures for contacts who developed illness. Other measures were also instituted, including closing hospitals and schools, wearing masks in public, banning public gatherings, screening inbound and outbound international travelers, and issuing travel advisories. There is a need to fully evaluate the effectiveness of these measures.

—Karen Shaw and Babatunde Olowokure

See also Epidemic; Outbreak Investigation; Zoonotic Disease

Further Readings

Chowell, G., Fenimore, P. W., Castillo-Garsow, M. A., & Castillo-Chavez, C. (2003). SARS outbreaks in Ontario, Hong Kong and Singapore: The role of diagnosis and isolation as a control mechanism. *Journal of Theoretical Biology*, *224*, 1–8.

Goh, D. L.-M., Lee, B. W., Chia, K. S., Heng, B. H., Chen, M., Ma, S., et al. (2004). Secondary household transmission of SARS, Singapore [Electronic version]. *Emerging Infectious Diseases.* Retrieved March 7, 2006, from http://www.cdc.gov/ncidod/EID/vol10no2/03-0676.htm.

Hui, D. S. C., Chan, M. C. H., Wu, A. K., & Ng, P. C. (2004). Severe acute respiratory syndrome (SARS): Epidemiology and clinical features. *Postgraduate Medical Journal*, *80*, 373–381.

Mackenzie, J., & Olowokure, B. (2006). Biocontainment and biosafety issues. In *SARS: How a global epidemic was stopped* (pp. 234–240). Manila, Philippines: World Health Organization, Western Pacific Region.

Olowokure, B., Merianos, A., Leitmeyer, K., & Mackenzie, J. (2004). SARS. *Nature Reviews. Microbiology*, *2*, 92–93.

Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., et al. (2003). Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, *300*, 1961–1966.

Shaw, K. (2006). The 2003 SARS outbreak and its impact on infection control practice. *Public Health*, *120*, 8–14.

World Health Organization. (2003). *SARS breaking the chains of transmission*, *2003*. Retrieved March 7, 2006, from http://www.who.int/features/2003/07/en.

World Health Organization. (2003). *WHO consensus on the epidemiology of SARS*. Retrieved March 7, 2006, from http://www.who.int/csr/sars/archive/epiconsensus/en.

World Health Organization. (2003). *World Health Report. Chapter 5: SARS: Lessons from a new disease.* Retrieved March 7, 2006, from http://www.who.int/whr/2003/chapter5/en/print.html.

Sexually Transmitted Diseases

Sexually transmitted diseases (STDs), also commonly referred to as sexually transmitted infections (STIs), are primarily spread through the exchange of bodily fluids during sexual contact. STDs may also be transmitted through blood-to-blood contact and from a woman to her baby during pregnancy or delivery (congenital transmission). Exposure to STDs can occur through any close exposure to the genitals, rectum, or mouth. Unprotected sexual contact increases the likelihood of contracting an STD. Abstaining from sexual contact can prevent STDs, and correct and consistent use of latex condoms reduces the risk of transmission. STDs include gonorrhea, chlamydia, genital herpes, syphilis, human papillomavirus or genital warts, lymphogranuloma venereum, trichomoniasis, bacterial vaginosis, human immunodeficiency virus (HIV), which causes AIDS, and hepatitis B. Other infections that may be sexually transmitted include hepatitis C, cytomegalovirus, scabies, and pubic lice. Having an STD increases the risk of becoming infected with HIV if one is exposed to HIV. Most STDs can be treated and cured, though important exceptions include HIV and genital herpes. Successful treatment of an STD cures the infection, resolves the clinical symptoms, and prevents transmission to others. HIV is not currently curable and can cause death. Genital herpes symptoms may be managed, but genital herpes is a recurrent, lifelong infection.

STDs remain a major public health concern due to their physical and psychological effects, as well as their economic toll. The Centers for Disease Control and Prevention (CDC) estimates that 19 million new infections occur each year, almost half of them among young people aged 15 to 24 years. Women, especially young women, ethnic minorities, and men who have sex with men (MSM) are often most affected by STDs (except chlamydia). The following are descriptions of key features associated with STDs.

Gonorrhea

Gonorrhea is caused by the bacterium *Neissera gonorrheae* and can be spread during vaginal, anal, or oral sex, as well as from a woman to her newborn during delivery regardless of whether the mother is symptomatic at the time. Gonorrhea can affect the urethra, rectum, throat, pelvic organs, and, in rare cases, conjunctiva, as well as the cervix in women. Colloquially referred to as the "clap" or the "drip," the incubation period is usually 2 to 5 days but may take up to 30 days to become visible. Gonorrhea is symptomatic in approximately 50% of infected individuals and possible symptoms include painful urination, abnormal discharge from the penis or vagina, genital itching or bleeding, and, in rare cases, sore throat or conjunctivitis. In women, symptoms may also include lower abdominal pain, fever, general tiredness, swollen and painful Bartholin glands (opening of the vaginal area), and painful sexual intercourse. In men, symptoms can include discharge from the penis that is at first clear or milky and then yellow, creamy, excessive, and sometimes blood tinged. If the infection disseminates to sites other than the genitals, possible symptoms include joint pain, arthritis, and inflamed tendons.

Left untreated, gonorrhea may cause complications to the female reproductive system, including pelvic inflammatory disease (PID), which can result in an increased risk of infertility (a danger that increases with subsequent episodes), tubo-ovarian abscess, inflammation of the Bartholin glands, ectopic (tubal) pregnancy, chronic pelvic pain, and, in rare occurrences, Fitz-Hugh-Curtis syndrome (inflammation of the liver). Left untreated in men, complications can include urethritis (infection of the urethra), epididymitis (inflammation and infection of epididymis), prostatitis, and infertility. If gonorrhea is not treated, complications may arise from disseminated gonococcal infection and include fever, cellulites, sepsis, arthritis, endocarditis (inflammation of the heart valves and the chambers of the heart), and meningitis.

Diagnosis of gonorrhea involves a medical history; physical exam, including a pelvic or genital exam; and collection of a sample of body fluid or urine. Discovered early and treated before complications arise, gonorrhea causes no long-term problems. Antibiotic treatment is recommended for individuals who have a positive gonorrhea test or who have had sex partners within the past 60 days who tested positive.

Chlamydia

Chlamydia is caused by the bacterium *Chlamydia trachomatis* and is the most common STD/STI in the United States. Chlamydia infects the urethra of men and the urethra, cervix, and upper reproductive organs

of women. It may also infect the rectum, throat, pelvic organs, and conjunctiva. Chlamydia can be passed from mother to newborn during vaginal delivery and, in rare instances, during Caesarean delivery. Symptoms develop in only about 10% of those with chlamydia. In women, symptoms include painful urination, cloudy urine, abnormal vaginal discharge, abnormal vaginal bleeding during intercourse or between periods, irregular menstrual bleeding, genital itching, lower abdominal pain, fever and general tiredness, swollen and painful Bartholin glands, and conjunctivitis. In men, symptoms include painful urination or itching during urination, cloudy urine, watery or slimy discharge from the penis, crusting on the tip of the penis, a tender anus or scrotum, and conjunctivitis. Complications in women include cervicitis (inflammation of the cervix), urethritis, endometritis, inflammation of the Bartholin glands, PID, pelvic abscess, infertility, and Fitz-Hugh-Curtis syndrome. Complications in men include urethritis, epididymitis, prostatitis, and infertility. Complications in both sexes include conjunctivitis, inflammation of the mucous membrane of the rectum (proctitis), and Reiter's syndrome, which is caused by a bacterial infection and results in varied symptoms, including joint and eye inflammation.

Chlamydia is diagnosed with a medical history; physical exam, including a pelvic or genital exam; and collection of body fluid or urine. Discovered early and treated before complications arise, chlamydia causes no long-term problems. Antibiotic treatment is recommended for individuals who have a positive chlamydia test or who have had sex partners within the past 60 days who tested positive. Because of the high incidence of chlamydia in the United States, the CDC recommends annual screening for sexually active women up to the age of 25 years and for women above the age of 25 years who engage in high-risk sexual behaviors. Health professionals are obligated to report a positive diagnosis to the State Health Department for the purposes of notification of sexual partner(s).

Genital Herpes

Genital herpes is a lifelong, recurrent viral infection caused by herpes simplex virus type 2 (HSV-2) and less frequently by herpes simplex virus type 1 (HSV-1). At least 50 million persons in the United States have genital HSV infection, and most exhibit minimal or no symptoms. HSV-1, which more commonly causes infections of the mouth and lips ("fever blisters"), can also be transmitted to the genitals through oral-genital (during oral outbreaks) or genital-genital contact. The painful multiple ulcerative lesions that typically characterize those infected with genital herpes are often absent. Most persons with genital herpes have mild or unrecognized infections. Although they are not aware of their condition, those with undiagnosed HSV are still contagious, as the virus periodically "sheds" in the genital tract. The majority of genital herpes infections are transmitted by persons who are unaware of infection or who are aware but asymptomatic when transmission occurs. Visible symptoms usually occur within 2 weeks of transmission and include blisters or ulcers on or around the penis, vagina, and anus. Outbreaks of genital ulcers typically last 2 to 4 weeks during the first clinical episode and several weeks to months during subsequent episodes, which are often less severe. Although the infection remains in the body indefinitely, outbreaks tend to decrease over a period of years.

Clinical diagnosis of genital herpes is insensitive and nonspecific and thus should be confirmed through laboratory testing. Isolation of the HSV in a cell culture is the preferred virologic test in patients who present with genital ulcers or other lesions. Because false-negative HSV cultures are common, especially in patients with recurrent infection or with healing lesions, type-specific serologic tests are useful in confirming a clinical diagnosis of genital herpes. Laboratory testing can be used to diagnose persons with unrecognized infection, which in turns allows for the treatment of sex partners. Distinguishing between HSV serotypes is important and influences prognosis and treatment since HSV-1 (which is responsible for about 30% of first-episode outbreaks) causes less frequent recurrences.

There is no cure for genital herpes, and the virus may be transmitted during asymptomatic periods. The burden caused by the lack of a cure, recurrence of outbreaks, and the possibility of transmission to sexual partners causes some people with genital herpes to experience recurrent psychological distress. Antiviral chemotherapy is the mainstay of management and offers clinical benefits to most symptomatic patients. Systematic antiviral drugs partially control the symptoms and signs of herpes episodes when used to treat both first clinical and recurrent episodes, or when used as daily suppressive therapy. However, these drugs neither eradicate the latent virus nor affect the risk,

frequency, or severity of recurrences after the drug is discontinued. Consequently, counseling regarding the natural history of genital herpes, sexual and perinatal transmission, and methods of reducing transmission is integral to clinical management. Persons with genital herpes should be informed that sexual transmission of HSV could occur during asymptomatic periods and that it is important to abstain from sexual activity when lesions or prodromal symptoms are present. The risk for neonatal HSV infection should be explained to both men and women.

Syphilis

Syphilis is caused by the bacterium *Treponema pallidum*, which is transmitted by vaginal, anal, or oral contact with the infected individual's open ulcers during the primary stage and mucus membrane or other sores during the secondary and latent stages. Syphilis may also be passed from mother to baby during pregnancy or labor and delivery. The CDC and the U.S. Prevention Services Task Force strongly recommend that all pregnant women be screened for syphilis because of the severe health consequences of congenital syphilis. The development of syphilis occurs in four stages—primary, secondary, latent, and tertiary stages. During the *primary stage*, the individual develops an often painless ulcer at the transmission site within 10 to 90 days after exposure. Ulcers mainly occur on the external genitals (men), inner or outer part of the vagina (women), or rectum. Ulcers usually last from 28 to 48 days and heal without treatment, leaving a thin scar. The *secondary stage* may begin before ulcers occurring in the primary stage have healed and is characterized by a rash that appears between 4 and 10 weeks after the development of the ulcer and consists of reddish brown, small, solid, flat, or raised skin sores that may mirror common skin problems. Small, open sores that may contain pus or moist sores that look like warts may develop on the mucous membranes. The skin rash usually heals without scaring in 2 to 12 weeks. Symptoms occurring during this stage include fever, sore throat, physical weakness or discomfort, weight loss, patchy hair loss, swelling of the lymph nodes, and nervous system symptoms such as headaches, irritability, paralysis, unequal reflexes, and irregular pupils. During the primary and secondary stages, the person is highly contagious. After secondary-stage symptoms subside, the person enters the *latent stage* and has no symptoms for a period of time ranging from 1 year to 20 years, although about 20% to 30% of individuals have a relapse during this period. If the syphilis infection is left untreated, the latent stage is followed by the *tertiary stage*, during which serious blood vessel and heart problems, mental disorders, gummata (large sores inside the body or on the skin), blindness, nerve system problems, and death may occur. Complications during this stage also include cardiovascular syphilis, which affects the heart and blood vessels, and neurosyphilis, which affects the brain or the brain lining.

The two definitive methods for diagnosing early syphilis are dark-field examinations and direct fluorescent antibody tests of lesion exudiate or tissue in the adjacent area. A presumptive diagnosis is possible with the use of two types of serologic tests: nontreponemal tests (e.g., Venereal Disease Research Laboratory test and rapid plasma reagin test) and treponemal tests (e.g., fluorescent treponemal antibody absorbed test and *T. pallidum* particle agglutination test). The use of only one type of serologic test is insufficient for diagnosis, because false-positive nontreponemal test results may occur due to various medical conditions. Penicillin G, administered parenterally, is the preferred drug treatment of all stages of syphilis, though dosage and length of treatment depend on the stage and clinical manifestations of the disease. The efficacy of penicillin for the treatment of syphilis is well established, and almost all treatments have been supported by case series, clinical trials, and 50 years of clinical experience.

Human Papillomavirus

Human papillomavirus (HPV) is the name of a group of approximately 100 related strains of a virus, more than 30 of which are sexually transmitted. HPV can affect the vulva, lining of the vagina, and the cervix of women; the genital area—including the skin of the penis—of men; and the anus or rectum of both men and women, which in turn may result in anal warts (it is not necessary to have had anal sexual contact to contract anal warts). HPV affects approximately 20 million people in the United States, and at least 50% of sexually active men and women will acquire genital HPV at some point in their lives. By the age of 50 years, at least 80% of women will have acquired genital HPV infection. The virus itself lives in the skin or mucous membranes, and most infected people are asymptomatic and unaware of the infection. Some people develop visible genital warts or have

precancerous changes in the cervix, vulva, anus, or penis. Genital warts may appear around the anus, penis, scrotum, groin, or thighs of men and on the vulva, in or around the vagina, and on the cervix of women. Most infections clear up without medical intervention. Although no HPV test is currently available for men, there exists a test to detect HPV DNA in women. A Pap test is the primary cancer-screening tool for cervical cancer or precancerous changes in the cervix, many of which are related to HPV. All types of HPV can cause mild Pap test abnormalities, and approximately 10 to 30 identified HPV types can lead, in rare cases, to cervical cancer. Pap tests used in U.S. cervical cancer-screening programs are responsible for greatly reducing deaths from cervical cancer. For 2004, the American Cancer Society (ACS) estimated that about 10,520 women will develop invasive cervical cancer and about 3,900 women will die from the disease. Certain types of HPV have also been linked to cancer of the anus and penis in men. The ACS estimates that about 1,530 men will be diagnosed with penile cancer in the United States in 2006. This accounts for approximately 0.2% of all cancers in men. The ACS estimates that about 1,910 men will be diagnosed with anal cancer in 2006, and the risk for anal cancer is higher among gay and bisexual men and men with compromised immune systems, including those with HIV.

There is currently no cure for HPV. The recently developed HPV vaccine has been shown to be effective in protecting against four strains of HPV. The HPV vaccine is targeted toward preventing cervical cancer and has been approved for women aged 9 to 26 years. The vaccine is nearly 100% effective in preventing precancerous cervical changes caused by two strains of HPV. These two strains cause 70% of all cervical cancers and most vaginal and vulvar cancers. It is also somewhat effective in protecting against two other HPV strains that are responsible for 90% of all cases of genital warts. The vaccine is not effective in women who are already infected with HPV and is most effective if given before women become sexually active.

Lymphogranuloma Venereum

Lymphogranuloma venereum (LGV) is caused by a type of *Chlamydia trachomatis* and is transmitted by genital or oral contact with infected fluid. The *first stage* of the disease occurs 3 days to 3 weeks after infection, during which a small, painless vesicle appears on the genitalia, anus, or mouth. This vesicle soon develops into an ulcer and goes away in a few days without treatment. Some men develop a node or bubonulus ("bubos") at the base of the penis, which may rupture or form draining sinuses or fistulas. The next stage, which occurs 7 to 30 days after the primary lesion resolves, is called regional lymphangitis. The lymph nodes that drain the area of original infection become swollen, tender, and painful. With mouth or throat infections, LGV can produce bubos under the chin, in the neck, or in the clavicular region. The bubos transform from hard and firm to softer masses, accompanied by reddened skin, which will sometimes rupture or form draining sinuses or fistulas. The patient may also suffer from fever, chills, headache, stiff neck, loss of appetite, nausea, vomiting, muscle and joint pains, or skin rashes. Women may suffer from lower abdominal or back pain. Those with rectal infections may have mucoid discharge from the anus. Late complications of LGV include elephantiasis (grotesque swelling due to lymphatic blockages) of the genitals, other genital deformities, ulcerative lesions (especially of the vulva), perianal abscesses, rectal strictures, fistulas, and large perianal swellings known as lymphrrhoids. LGV causes ulcers, which can increase the risk of contracting HIV infection.

LGV rarely occurs in the United States and other industrialized countries. However, recent outbreaks of LGV proctitis in MSM in the Netherlands and other European countries in 2003, and a handful of cases in New York City, have raised concerns in the United States. White, HIV-infected MSM constitute the majority of patients diagnosed with LGV proctitis, and preliminary findings suggest that the main mode of transmission was unprotected anal intercourse. Men and women may be asymptomatic and unknowingly transmit LGV. Men can spread infection while only rarely suffering long-term health problems. Women are at high risk of severe complications of infection, including acute salpingitis and PID, which can lead to chronic pain, ectopic pregnancy, and infertility. LGV is sensitive to a number of antibiotics, including erythromycin, tetracyclins, doxycyclin, and sulfadiazine. Infected lymph nodes must often be drained through aspiration, and fistulas or other structural problems may require surgery. Technology for LGV testing is not currently commercially available. In U.S. states that lack laboratory capacity to perform LGV diagnostic testing, specimens may be submitted to the CDC.

Bacterial Vaginosis

Bacterial vaginosis (BV) is a condition where the normal balance of bacteria in the vagina is disrupted and there is an overgrowth of harmful bacteria. It is estimated that approximately 16% of pregnant women have BV, and it is the most common vaginal infection in women of childbearing age. Little is known about what causes BV or how it is transmitted. It is believed that some behaviors such as having a new sex partner, having multiple sex partners, douching, and using an intrauterine device for contraception can upset the normal balance of bacteria in the vagina and put women at increased risk. BV may spread between two female partners.

Symptoms of BV include a vaginal discharge—which is white-gray, thin, with an unpleasant odor—burning during urination, and itching around the outside of the vagina. Some women with BV report no signs or symptoms, and in most cases, BV causes no complications. Serious risks do exist, however, and include increased susceptibility to HIV infection if exposed to the HIV virus, increased chance of passing HIV to a sex partner if infected with the HIV virus, increased development of PID following surgical procedures such as a hysterectomy or an abortion, increased risk of complications during pregnancy, and increased susceptibility to other STDs. Although BV will sometimes clear up without treatment, all women with symptoms of BV should be treated to avoid complications such as PID. Treatment is especially important for pregnant women.

Trichomoniasis

Trichomoniasis is caused by the single-celled parasite *Trichomonas vaginalis* and affects the vagina in women and the urethra in men. Trichomoniasis is sexually transmitted through penis-to-vagina intercourse or vulva-to-vulva contact. Women can acquire the disease from infected men or women, while men usually contract it only from infected women. Most men with trichomoniasis are asymptomatic, although some may experience temporary irritation inside the penis, mild discharge, or slight burning after urination or ejaculation. Some women have symptoms, including vaginal discharge that is frothy, yellow-green, and has a strong odor; discomfort during intercourse and urination; irritation and itching of the genital area; and, in rare cases, lower abdominal pain. Symptoms usually appear in women within 5 to 28 days of exposure. For both sexes, trichomoniasis is diagnosed by physical examination and laboratory test, although the parasite is harder to detect in men. Trichomoniasis can usually be cured with a single dose of an orally administered antibiotic. Although the symptoms of trichomoniasis in men may disappear within a few weeks without treatment, an asymptomatic man can continue to infect or reinfect a female partner until he has been treated. Therefore, both partners should be treated concurrently to eliminate the parasite.

—Roy Jerome

See also HIV/AIDS; Hepatitis; Partner Notification; Sexual Risk Behavior

Further Readings

American Cancer Society. (2006, June 8). *HPV vaccine approved: Prevents cervical cancer*. American Cancer Society News Center. Retrieved June 19, 2006, from http://www.cancer.org/docroot/NWS/content/NWS_1_1x_HPV_Vaccine_Approved_Prevents_Cervical_Cancer.asp.

Centers for Disease Control and Prevention. (2002, May 10). Sexually transmitted disease treatment guidelines, 2002 [Electronic version]. *Morbidity and Mortality Weekly Report*, *51*(RR-6), 13–36, 44–47. Retrieved June 19, 2006, from http://www.cdc.gov/std/treatment/default-2002.htm.

Centers for Disease Control and Prevention. (2004). *Bacterial vaginosis: Fact sheet*. Retrieved May 15, 2006, from http://www.cdc.gov/std/bv/STDFact-Bacterial-Vaginosis.htm.

Centers for Disease Control and Prevention. (2004). *Genital herpes: Fact sheet*. Retrieved May 15, 2006, from http://www.cdc.gov/std/Herpes/STDFact-Herpes.htm.

Centers for Disease Control and Prevention. (2004). *Genital HPV infection: Fact sheet*. Retrieved May 15, 2006, from http://www.cdc.gov/std/HPV/STDFact-HPV.htm.

Centers for Disease Control and Prevention. (2004). *Trichomonas infection: Fact sheet*. Retrieved May 15, 2006, from http://www.cdc.gov/ncidod/dpd/parasites/trichomonas/factsht_trichomonas.htm.

Centers for Disease Control and Prevention. (2005, November). *Trends in reportable sexually transmitted diseases in the United States, 2004*. Retrieved May 15, 2006, from http://www.cdc.gov/std/stats04/trends2004.htm.

Centers for Disease Control and Prevention. (2006). *Lymphogranuloma Venereum (LGV) Project: Information sheet*. Retrieved June 19, 2006, from http://www.cdc.gov/std/lgv/LGVinf05–22–2006.pdf.

Centers for Disease Control and Prevention. (2006). *Viral hepatitis B: Fact sheet.* Retrieved May 15, 2006, from http://www.cdc.gov/ncidod/diseases/hepatitis/b/fact.htm.

Hook, E. W. (2004). Syphilis. In L. Golman, & D. Ausiello (Eds.), *Cecil textbook of medicine* (22nd ed., Vol. 2, pp. 1923–1932). Philadelphia: Saunders.

Low, N. (2004). Chlamydia (uncomplicated, genital). *Clinical Evidence*, *11*, 2064–2072.

Peipert, J. F. (2003). Genital chlamydial infections. *New England Journal of Medicine*, *349*, 2424–2430.

Quest Diagnostics. (2005). *Chlamydia*. Retrieved May 15, 2006, from http://www.questdiagnostics.com/kbase/topic/major/hw146904/descrip.htm.

Quest Diagnostics. (2005). *Gonorrhea*. Retrieved May 15, 2006, from http://www.questdiagnostics.com/kbase/topic/major/hw188975/descrip.htm.

Quest Diagnostics. (2005). *Syphilis*. Retrieved May 15, 2005, from http://www.questdiagnostics.com/kbase/topic/major/hw195071/descrip.htm.

van de Laar, M. J., Gotz, H. M., de Zwart, O., van der Meijden, W. I., Ossewaarde, J. M., Thio, H. B., et al. (2004). Lymphogranuloma venereum among men who have sex with men: Netherlands, 2003–2004 [Electronic version]. *Morbidity and Mortality Weekly Report*, *53*, 985–988.]

Weinstock, H., Berman, S., & Cates, W. (2004). Sexually transmitted diseases among American youth: Incidence and prevalence estimates, 2000. *Perspectives on Sexual and Reproductive Health*, *36*, 6–10.

World Health Organization. (1995). Sexually transmitted diseases. In *Teaching modules for continuing education in human sexuality: Vol. 8.* (pp. 5–40). Manila, Philippines: Author.

World Health Organization. (2000). *Guidelines for surveillance of sexually transmitted diseases*. New Delhi: Author.

Sexual Minorities, Health Issues of

The health of sexual minorities represents a broad and rapidly expanding area of public health research and concern. Encompassing people who are lesbian, gay, bisexual, and transgender (LGBT), sexual minorities are diverse populations who have struggled with issues of sexuality, identity, and gender amidst historic and continuing stigma, fear, and discrimination. Consequently, while possessing the same basic health needs as the general population, LGBT people face additional health issues related to social discrimination, behavioral risk factors, and unique medical conditions. The diversity of these populations spans race, ethnicity, age, education, socioeconomic position, geography, political affiliation, and the degree to which individuals identify and interact with other LGBT people. From all indications, sexual minorities exist wherever there are human societies. This entry examines health outcomes among LGBT people, including the impact of disparities in their access to health care. It also discusses methodological issues related to research on these populations.

Over the past 25 years, since the birth of the modern gay rights movement following the "Stonewall Riots" in New York City in 1969, LGBT people have experienced growing acknowledgment of their basic human rights throughout the world. Facilitated in large measure by the activism of LGBT people themselves, this has led to increased awareness of persistent health disparities among and between these populations, including rates of HIV/AIDS among gay and bisexual men and certain types of cancers among lesbians. However, conducting epidemiological research on LGBT people is complicated by several issues, including varying definitions of the most basic terms (such as *homosexual*), reluctance of LGBT people to identify themselves as such or participate in research due to stigma, and a history of differing priorities and conflicts between LGBT people and the medical and social science communities.

Definitions of Terms and Conceptual Issues

To consider the health issues of sexual minorities first requires a clarification of terms and an appreciation for the subject's conceptual complexity. While there is no complete agreement over language, broadly speaking, *sexual orientation* refers to a person's sexual and romantic attraction to other people. This term is increasingly favored over *sexual preference*, which implies that attraction is merely a choice and not an inherent personal characteristic. Individuals whose sexual orientation is to people of the opposite sex are *heterosexual* and those whose orientation is to people of the same sex are *homosexual*, with women who are primarily attracted to other women referred to as *lesbians* and men who are primarily attracted to other men referred to as *gay*. Individuals who are attracted to both men and women are *bisexual*, and depending on the person, this attraction may be felt equally toward both sexes or it may be stronger or different toward one or the other. It is important to note that sexual orientation is not the same as sexual identity or behavior, and one does not necessarily imply the other.

Transgender is an umbrella term referring to people who do not fit to traditional or customary notions

of *gender*—the sociocultural norms and beliefs that define what it means to be a man or a woman. In general, society assigns an individual's gender role at birth based on one's genitals, but some people identify more strongly with the opposite gender (e.g., natal females who identify as men, natal men who identify as women) or with a variance that falls outside dichotomous gender constructions (e.g., individuals who feel they possess both or neither genders, including those with ambiguous sex characteristics). Some transgender people are *transvestites* or cross-dressers, primarily men who dress in women's clothing for erotic or other personal interests; others may "do drag" for fun, entertainment, or personal expression. Many of these people are comfortable with their natal gender. Other transgender people may be *transsexuals*, people who pursue medical interventions such as the use of hormones of the opposite sex and/or surgeries to align their bodies more closely with their interior sense of self. Still others reject any categorization of their gender or sexuality. Transgender people may identify as male or female independent of their anatomy or physiology and may be heterosexual, homosexual, bisexual, or nonsexual.

At the individual level, sexuality and gender reflect attractions, behaviors, and identities that develop throughout the life course, influenced by both biological and psychosocial factors such as family and culture. For example, a man may engage repeatedly in sexual activities with other men but neither acknowledge participation in homosexual behavior nor identify as gay. A woman may enter into an emotionally fulfilling intimate relationship with another woman that involves temporary identification as a lesbian but at a future time enter into exclusive long-term relationships with men, at no time seeing herself as bisexual. Another individual may present as anatomically female but personally identify as a transgender gay man. These examples illustrate the need for health professionals to approach the subjects of gender and sexuality with few assumptions and perspectives open to the needs of the clients within their cultural contexts.

The health of LGBT people may be affected by social conditions characterized by stigma, fear, rejection, prejudice, discrimination, and violence. *Heterosexism* is unconscious or deliberate discrimination that favors heterosexuality, such as the assumption that everyone is heterosexual or the extension of social privileges or an elevated social status based on heterosexual identity, behaviors, or perceived heterosexual orientation. *Homophobia*, *biphobia*, and *transphobia* describe the respective hatred and fear of homosexuals, bisexuals, and transgender people. These fears and prejudices operate at all levels of society to varying degrees and at various times, interacting with and exacerbating the effects of racism, sexism, and class-based discrimination. These biases may account for a large degree of the disparity in health outcomes observed among LGBT populations, although research is currently insufficient to determine this conclusively.

Consequently, whether working with individuals or populations, many health professionals find it useful to apply an ecological framework to better comprehend LGBT health issues, taking time to identify, articulate, and address both needs and assets among interpersonal, community, and societal factors. For example, LGBT stigma and discrimination can influence the determination and funding of research priorities, the design and implementation of prevention and intervention programs, the development of standards of care, and the use of culturally appropriate services. This discrimination can be expressed directly through exposures to violence, humiliation, or suboptimal care and indirectly through invisibilization and marginalization of LGBT health concerns and treatment. However, taking an assets-based approach, these inequities can be mitigated by inclusive and supportive policies, education, research, and training; community empowerment; and by drawing on the resources, resilience, and participation of LGBT people themselves.

Outcomes

Prevalence estimates of same-sex attractions, identities, and behaviors range between approximately 1% and 13% of the general population. In a 1994 national study of U.S. adults, Laumann, Gagnon, Michael, and Michaels found 8% of respondents reported same-sex attractions; 2% identified as gay, lesbian, or bisexual; and 7% reported having engaged in same-sex behaviors. These data appeared congruent with data from national population-based samples of France and the United Kingdom the following year. Prevalence estimates of transgender individuals are much less certain; there are no reliable data for the U.S. population. Outside the United States, transgender estimates range from 0.002% to 0.005%, with some transgender activists estimating prevalence of "strong transgender feelings" (without sexual reassignment) at 0.5%, or 1 of

every 200 persons. LGBT estimates, where they exist, vary across studies due to a variety of factors, including definitions and use of terms, sampling methodologies, data analyses, age (youth vs. adults), location (rural vs. urban), and cultural contexts. These gaps and inconsistencies in knowledge suggest the need for more routine, rigorous, and inclusive population-based sampling, as done during censuses, risk-behavior surveys, and disease surveillance and reporting.

It may be useful to categorize LGBT health outcomes among three areas of concern. First, sexual minorities may experience risks or exposures unique to the population of interest, such as hormone use among transsexuals or anal sex among men who have sex with men. Second, the risk or exposure may not be unique to the population but exists at a prevalence higher than that found in the general population, such as substance abuse or depression. Third, the risk or exposure may neither be unique nor exist in greater proportion, but the issue warrants an approach that is particularly attentive or culturally sensitive, such as screening of sexually transmitted diseases (STDs) among bisexuals, reproductive services among lesbians, or routine medical examinations among transsexuals.

Stigmatizing social conditions, particularly among youth, racial/ethnic minorities, and transgender individuals, contribute to a number of health disparities shared to varying degrees among LGBT populations. These include access and use of programs and services, mental health issues, and exposures to violence. For example, the Women's Health Initiative, a U.S. sample of 96,000 older women, found that lesbians and bisexual women were significantly more likely to be uninsured compared with heterosexual women (10%, 12%, and 7%, respectively). Uninsured levels appear highest among transgender people, disproportionately so among people of color (21% to 52% among studies), and most health care related to transgender issues is not covered by insurance, making transgender health care personally very expensive. Adolescents are the most uninsured and underinsured among all age groups, and LGBT youths perhaps face the greatest barriers to appropriate, sensitive care.

Social stigma is a stressor with profound mental health consequences, producing inwardly directed feelings of shame and self-hatred that give rise to low self-esteem, suicidality, depression, anxiety, substance abuse, and feelings of powerlessness and despair that limit health-seeking behaviors. According to findings documented in the report, Healthy People 2010 Companion Document for LGBT Health, homosexually active men report higher rates of major depression and panic attack compared with men who report no homosexual behavior, and homosexually active women report higher rates of alcohol and drug dependence compared with women who report no homosexual behavior. In New Zealand, LGBT youths were found to be at higher risk for major depression, generalized anxiety disorder, and conduct disorders than were non-LGBT youths. Among 515 transsexuals sampled in San Francisco in 2001, Clements-Nolle and colleagues reported depression among 62% of the transgender women and 55% of the transgender men; 32% of the sample had attempted suicide. While the earliest studies of alcohol and other substance use among lesbians and gay men suggested alarmingly high rates, subsequent studies describe elevated rates among LGBT populations as a function of socially determined factors, including age, race/ethnicity, socioeconomic position, and prior exposures to trauma and violence.

These more recent studies demonstrate the focus and refinement that is now taking place in the field of LGBT public health research, providing unprecedented opportunities to comprehend health issues within specific populations. For example, studies suggest that lesbians and bisexual women are at higher risk for breast cancer compared with heterosexual women due to higher rates of risk factors, including obesity, alcohol consumption, having never given birth, and lower rates of breast cancer screenings. Similarly, lesbians and bisexual women may also be at higher risk for gynecologic cancers because they receive less frequent gynecologic care. Among men who have sex with men, in addition to higher rates of HIV/AIDS—particularly among racial/ethnic minorities—gay and bisexual men are at increased risks for other STDs, including syphilis, gonorrhea, chlamydia, human papillomavirus, and hepatitis A, B, and C. Cross-sex hormone use among transsexuals, including long-term use of estrogen and testosterone, presents potential health risks that are poorly documented at this time, including possible cancers of the breast and ovaries. Because of transphobia, transgender people may be reluctant to provide a full health history to their medical providers or they may be flagrantly denied care. These experiences contribute to absence of treatment or delayed treatments, such as cancer screenings, or to informal and medically unsupervised procedures, such as unregulated hormone therapy or the cosmetic use of injectable silicone.

Access

Access refers to the ways that LGBT health concerns are or are not addressed at various levels of society, from government and institutional policies and resources to the individual practices of health professionals. For example, following the removal of homosexuality from the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* in 1974, the mental health profession reversed its long-held position that regarded homosexuality as a psychopathology and emerged as one of the most important advocates for the normalization and acceptance of same-sex attraction. By the beginning of the 21st century, the American Public Health Association had acknowledged the special health concerns of LGBT populations with a policy statement on the need for research on gender identity and sexual orientation and a subsequent journal issue wholly dedicated to the topic in 2001. The U.S. government signaled similar support with publication of an Institute of Medicine report on lesbian health in 1999 and the inclusion of gays and lesbians in *Healthy People 2010*, the 10-year blueprint for public health produced by the U.S. Department of Health and Human Services. In turn, these policies potentially influence research, funding, and programs that directly affect the lives and well-being of LGBT people and their families.

However, to be effective, resources and a proactive political agenda must be mobilized to enact findings and recommendations. A provider motivated to do more for LGBT health can do little with insufficient funding or a hostile or indifferent environment. A client can't seek health services that don't exist, or he or she is less willing to do so if he or she has either experienced stigma or anticipates a stigmatizing environment. For these reasons, advocates have advanced guidelines and standards of care for LGBT people, including provider guidelines from the U.S. Gay and Lesbian Medical Association and the seminal transgender standards of care from the World Professional Association for Transgender Health (formerly the Harry Benjamin International Gender Dysphoria Association). At their most basic, these guidelines encourage providers to promote open, honest, and trusting relationships with LGBT individuals that facilitate optimal delivery of care and services. Recommendations include (1) *the creation of a welcoming environment*, including provider participation in LGBT referral programs and displays of media, brochures, and a nondiscrimination policy inclusive of sexual orientation and gender identity/expression in a multicultural context; (2) *use of inclusive forms, languages, and discussions* that does not assume the individual's identity, orientation, behavior, and relationship status; (3) *development of a written confidentiality policy* that outlines the types of information collected and how that information is protected and shared; and (4) *training and evaluation of staff* to maintain standards of respect, sensitivity, and confidentiality toward LGBT patients, clients, and personnel. Given the ubiquitous and diverse nature of LGBT populations and their varied health concerns, it is incumbent on the ethical health professional to anticipate and prepare for the presence of sexual minorities, remaining especially sensitive to the ways LGBT people may have experienced prior discrimination and trauma from the health care system.

Methodologies

Until very recently, little research specifically detailing LGBT health issues existed, as social stigma discouraged scientific careers dedicated to LGBT health, the allocation of resources, and the publication of findings in reputable sources. This maintained a cycle of silence: Without valid information, it was difficult to demonstrate need; with no demonstrated need, it was difficult to justify LGBT research.

Beginning in the early 1980s, with the advent of HIV/AIDS primarily among white, middle-class, gay-identified men in the United States and western Europe, LGBT individuals began to organize themselves into groups, such as the AIDS Coalition to Unleash Power, to protest seeming government indifference to this emerging epidemic. These actions spurred revolutionary changes in institutional research protocols, including improved access to clinical trials and faster approval of new treatments. As research agendas began to focus on HIV risks among men who have sex with men and other sexual minorities (albeit more slowly and to a lesser degree), government and academia established collaborative partnerships with community-based LGBT organizations, such as Gay Men's Health Crisis in New York City and the Howard Brown Health Center in Chicago. Over the succeeding 25 years, this catalyzed the development of a research infrastructure more amenable to address LGBT health issues than at any other time in history.

While HIV/AIDS remains one of the most widely studied health concerns among these populations, the field of research in LGBT health has expanded in breadth and depth to encompass the social, behavioral, biomedical, and policy dimensions across a wide range of issues. This diversity presents numerous ethical and methodological challenges that, in turn, provide rich opportunities for innovation, discussion, and refinement. Alluding to the complexities described previously, it is essential that researchers clarify terms during the collection, analysis, and reporting of data regarding sexual and gender orientations, identities, and behaviors. For example, behavioral risk surveys that ask "Are you a lesbian?" potentially miss same-sex behavior unless the question "Do you have sex with men, women, or both?" is specifically asked, and this same survey may miss past experience unless a time frame is specifically assigned. A form that provides only for "male" or "female" gender identities, or allows descriptions of relationship status based on heterosexual concepts of marriage, fails to capture potentially relevant data and tacitly devalues the contributions of LGBT participants.

Social stigma and the relatively small numbers of LGBT people create difficulties when data are collected using traditional probability sampling methods such as random household- or telephone-based surveys. Individuals may be reluctant disclosing such personal information when other household members may be present or they may not trust the researchers, potentially providing biased, unreliable responses. Additionally, securing a representative, statistically significant sample size may prove expensive. Consequently, LGBT research has historically relied on samples of convenience, using targeted, venue-based, or snowball sampling methods that are nonrandom and therefore difficult to generalize to the larger populations of interest. These limitations have led to the development of time-space (or time-location) sampling and respondent-driven sampling, two probability sampling methods specifically designed to reach hidden or rare populations.

Many advocates recommend a participatory research model (also known as participatory action research) that involves LGBT community members themselves in the development, implementation, and analysis of research conducted on these populations. The traditional researcher enters into a partnership with diverse representatives of the populations of interest, who work together to identify the research question, develop and implement the research plan, collect and analyze data, and disseminate results, including among the populations under investigation. In turn, the research partnership serves to educate and empower participating communities, as the experience builds capacity among community members and findings help guide policies, programs, and further research. As outlined in the 2007 groundbreaking text, *The Health of Sexual Minorities*, one such model is Fenway Community Health, a comprehensive, community-based health center that integrates care, education, and research for lesbian, gay, bisexual, and transgender populations throughout the greater Boston area.

—Carey V. Johnson and Matthew J. Mimiaga

See also Community-Based Participatory Research; Ethics in Health Care; Ethics in Public Health; Health Disparities; HIV/AIDS

Further Readings

Conway, L. (2002, December). *How frequently does transsexualism occur?* Retrieved February 11, 2007, from http://www.lynnconway.com.

Dean, L., Meyer, I. H., Robinson, K., Sell, R. L., Sember, R., Silenzio, V. M. B., et al. (2000). Lesbian, gay, bisexual, and transgender health: Findings and concerns. *Journal of the Gay and Lesbian Medical Association*, *4*, 102–151.

Dunn, P., Scout, W. J., & Taylor, J. S. (Eds.). (2005). *Guidelines for care of lesbian, gay, bisexual and transgender patients.* San Francisco: Gay and Lesbian Medical Association.

Gay and Lesbian Medical Association and LGBT Health Experts. (2001). *Healthy People 2010 companion document for lesbian, gay, bisexual and transgender (LGBT) health*. Retrieved January 24, 2007, from http://www.glma.org/_data/n_0001/resources/live/HealthyCompanionDoc3.pdf.

Gross, M. (Ed.). (2003). Special issue devoted to lesbian, gay, bisexual and transgender public health. *American Journal of Public Health*, *93*, 857–1012.

Harry Benjamin International Gender Dysphoria Association. (2001). *The standards of care for gender identity disorders* (6th version). Düsseldorf, Germany: Symposium Publishing.

Laumann, E. O., Gagnon, J., Michael, R. T., & Michaels, S. (1994). *The social organization of sexuality: Sexual practices in the United States.* Chicago: University of Chicago Press.

Meyer, I. H. (Ed.). (2001). Special issue devoted to lesbian, gay, bisexual, and transgender public health. *American Journal of Public Health*, *91*, 856–991.

Meyer, I. H., & Northridge, M. E. (Eds.). (2007). *The health of sexual minorities.* New York: Springer.
Solarz, A. (Ed.). (1999). *Lesbian health: Current assessment and directions for the future.* Washington, DC: National Academy Press.
U.S. Department of Health and Human Services. (2000). *Healthy People 2010: Understanding and improving health* (2nd ed.). Washington, DC: Government Printing Office.
Weitze, C., & Osburg, S. (1996). Transsexualism in Germany: Empirical data on epidemiology and application of the German Transsexuals' Act during its first ten years. *Archives of Sexual Behavior*, *25*, 409–425.

Sexual Risk Behavior

Sexual risk behaviors constitute a range of sexual actions that increase individuals' risk for bacterial and viral sexually transmitted infections (STIs), including the human immunodeficiency virus (HIV), and for unintended pregnancy. Increased sexual risk results from a combination of the specific sexual behavior and the level of protective action used. Contextual factors, such as drug and alcohol use, can also influence the level of risk involved. While abstinence and autoeroticism are the only truly effective methods of preventing unintended pregnancy and STIs, various risk-reduction strategies exist. Ultimately, sexual behaviors fall on a risk continuum, which depends on the sexual behavior and the protective action employed.

Sexual Behaviors and Associated Risk

Sexual behaviors comprise a wide array of acts ranging from minimal contact to penetration. Abstinence from vaginal intercourse is the only completely effective method for preventing unintended pregnancy. Abstinence from oral, vaginal, and anal intercourse, as well as autoeroticism (otherwise known as masturbation)—fulfilling individual sexual needs without a partner—is the only truly effective means for preventing STIs. When more than one individual is involved in a sexual act, sexual behaviors fall along a continuum of risk. At the low end of the risk continuum are behaviors that consist of minimal physical contact, including kissing, frottage (rubbing against the body of another individual), and fondling.

Oral sex, the act of orally stimulating the penis, vagina, or the anus (termed anilingus or rimming), is toward the middle of the risk continuum when no barrier method is used. While the risk for STI transmission during oral sex is lower than the risk associated with vaginal or anal sex, the risk is still present. The theoretical risk of STI transmission from *oral-penile* contact is present due to infected preejaculate or semen, penile fissures, open sores on the penis, bleeding gums, or open sores in the mouth. The theoretical risk from *oral-vaginal* contact is present due to infected vaginal fluid or blood from menstruation, open sores in the vulva or vagina, or bleeding gums or open sores in the mouth. The theoretical risk from *oral-anal* contact is present due to infected blood in fecal matter, anal fissures, open sores in the anal area, or if infected blood from the mouth enters the rectal lining.

Vaginal-penile intercourse is at the high end of the risk continuum when no barrier method is used. The consequences of unprotected vaginal intercourse include STIs, including HIV, and unintended pregnancy. The risk of HIV transmission from unprotected vaginal intercourse is present for either partner due to infected semen, infected vaginal fluid or menstrual blood, or open sores in the vulva or vagina. STI transmission also occurs through these pathways; however, certain STIs can be transmitted solely through contact with mucosal surfaces or infected skin.

While unprotected insertive anal intercourse is at the high end of the risk continuum, the riskiest sexual behavior is unprotected receptive anal intercourse. The risk consequences of unprotected anal intercourse are HIV and other STIs. The risk of infection is present in methods used without a barrier due to infected semen (including preejaculate), open sores in the anus, or tears in the lining of the anus.

Factors That Increase Sexual Risk

Several factors can place individuals at increased risk for the consequences of sexual behaviors, including substance use and sex with multiple or anonymous partners.

Individuals who use substances (including drugs and/or alcohol) prior to and/or during sexual activity place themselves at higher risk for engaging in sexual behaviors that may expose them to HIV and other STIs, including abandoning barrier methods. Furthermore, individuals addicted to mood-altering substances may trade sexual acts for money or drugs, increasing the associated risks.

Having multiple sexual partners also increases the risk for consequences associated with sexual behaviors because it increases the likelihood of having

a sexual encounter with an infected partner. This is also the case for individuals having sexual relations with anonymous partners or partners with unknown STI status without protection.

Prevalence of Sexual Risk Behaviors

Although various studies have examined sexual risk behaviors, the definitions and questions used to measure these behaviors have been inconsistent. Furthermore, nationally representative studies are dated. These issues make it difficult to assess prevalence on a national level and to ascertain how pervasive the problem currently is. The lack of such data may be partly attributable to the difficulty in defining and measuring these behaviors due to the sensitive nature of the topic.

Prevalence data on sexual risk behaviors have focused less on the general adult population and more so on specific populations, including men who have sex with men (MSM) and adolescents. National-level data from the United States have estimated the prevalence of alcohol and/or drug use directly prior to sexual encounters at, on average, between 52% and 85% among MSM and between 20% and 31% among high school students. Additionally, having multiple sexual partners (operationalized differently in different studies) has been estimated at 11% for the general population of adults, between 26% and 50% for MSM, and 14% for high school students. Finally, the estimated prevalence of condom use at last sexual intercourse has been estimated to be between 19% and 62% in the general adult population (dependent on level of commitment with the sexual partner), 55% to 77% of MSM living in urban areas, and 51% to 65% of sexually active high school students. It is important to note that these assessments include individuals involved in mutually exclusive relationships.

Physical Consequences of Sexual Risk Behavior

The physical consequences of unprotected vaginal intercourse are unintended pregnancy and the transmission of STIs. The consequences of unprotected oral and anal intercourse are STIs, which can be either bacterial or viral. Bacterial STIs can be both treated and cured. The most common bacterial STIs in the United States include bacterial vaginosis, chlamydia, gonorrhea, pelvic inflammatory disease, syphilis, and trichomoniasis.

Viral STIs can be treated but are incurable. Once infected with a viral STI, an individual will always be infected, and although they are not always symptomatic, they always remain at risk for infecting others. The most common viral STIs in the United States include HIV; herpes simplex viruses type 1 and type 2; hepatitis A, B, and C; and human papillomavirus (also known as genital warts).

Risk Reduction

For sexually active individuals, various methods exist to reduce risk associated with sexual behavior, including physical barrier methods, chemical methods, and monogamy. Physical barrier methods to prevent pregnancy and STIs include male condoms and female condoms. When used correctly (i.e., using a new condom for each sexual encounter) and consistently, male latex condoms have been demonstrated to be a highly effective form of protection against unintended pregnancy and HIV. Condoms have also been shown to provide high levels of protection against some STIs, including gonorrhea, chlamydia, and trichomoniasis. Condoms afford less protection against genital ulcer STIs, including herpes, chancroid, and syphilis, because they may be transmitted via contact with mucosal surfaces or infected skin not covered by a condom.

For individuals with allergies to latex, male condoms are also manufactured in polyurethane. These condoms have also been shown to be highly effective in preventing unintended pregnancy, HIV, and other STIs. However, the breakage and slippage rate among polyurethane male condoms has been shown to be significantly higher, making them a slightly inferior physical barrier method.

Natural skin condoms, which are made out of animal tissue, have been shown to be effective in preventing unintended pregnancy but not in preventing HIV or other STIs. Because these condoms are made of animal membrane, they may be porous, potentially allowing viruses to pass through.

The female condom is another form of physical barrier method that is manufactured out of polyurethane. These condoms line the vagina to form a protective barrier to prevent STI acquisition and unintended pregnancy. The female condom may also be

placed on the penis for use as a barrier method during vaginal or anal sex. The female condom has been shown to be as effective as male condoms in preventing unintended pregnancy and the transmission of HIV and other STIs.

For additional protection, condoms may be used with chemical barriers such as spermicide, a chemical agent that immobilizes sperm to prevent pregnancy, and/or microbicide, a chemical agent that prevents the transmission of HIV and other STIs. Nonoxynol-9, a spermicide that has in vitro activity against some STIs, is the active ingredient in the majority of over-the-counter contraceptive products. Research has demonstrated that Nonoxynol-9 is not effective alone as a preventive method for STIs. Furthermore, some evidence has shown that repeated use can create genital or anal lesions, thereby increasing the risk of STI acquisition.

A third method for risk reduction is mutual monogamy, when sexual partners only have sexual relations with one another. Because having sexual encounters with multiple partners dramatically increases the risk for STI transmission, mutual monogamy can eliminate the risk, if both partners are uninfected.

—Margie Skeer and Matthew J. Mimiaga

See also HIV/AIDS; Partner Notification; Sexually Transmitted Diseases

Further Readings

Centers for Disease Control and Prevention. (2000). *HIV/AIDS update: Preventing the sexual transmission of HIV, the virus that causes AIDS: What you should know about oral sex*. Retrieved June 20, 2006, from http://www.cdcnpin.org/Updates/oralsex.pdf.

Centers for Disease Control and Prevention. (2001). Prevalence of risk behaviors for HIV infection among adults: United States, 1997. *Morbidity and Mortality Weekly Report*, *50*, 262–265.

Centers for Disease Control and Prevention. (2002). Trends in sexual risk behaviors among high school students: United States, 1991–2001. *Morbidity and Mortality Weekly Report*, *51*, 856–859.

Holmes, K. K., Levine, R., & Weaver, M. (2004). Effectiveness of condoms in preventing sexually transmitted infections. *Bulletin of the World Health Organization*, *82*, 454–461.

Minnis, A. M., & Padian, N. S. (2005). Effectiveness of female controlled barrier methods in preventing sexually transmitted infections and HIV: Current evidence and future research directions. *Sexually Transmitted Infections*, *81*, 193–200.

Van Damme, L. (2004). Clinical microbicide research: An overview. *Tropical Medicine and International Health*, *9*, 1290–1296.

SF-36® Health Survey

The SF-36® Health Survey is a 36-item questionnaire used to assess patient-reported health-related quality-of-life outcomes. The SF-36® Health Survey measures eight domains of health:

1. Physical functioning
2. Role-Physical
3. Bodily pain
4. General health
5. Vitality
6. Social functioning
7. Role-Emotional
8. Mental health

It yields an eight-scale profile of norm-based scores (one for each of the eight domains of health) as well as physical and mental health component summary scores, a self-reported health transition rating, a response consistency index (RCI), and a preference-based health utility index (SF-6D).

The SF-36® Health Survey is a generic measure proven to be useful for comparing general and specific populations, comparing the relative burden of diseases, differentiating the health benefits produced by a wide range of different treatments, screening individual patients, and predicting health care costs, mortality, and other important outcomes. Adapted for use in more than 90 country/language versions, it is available in standard (4-week recall) or acute (1-week recall) forms. It has been successfully administered to persons 14 years and older using self-administration by paper and pencil, Internet, telephone, interactive voice response, and personal digital assistant, as well as interviewer-administered forms. Cited in more than 7,500 publications, including approximately 1,000 published randomized clinical trials, the SF-36® Health Survey is part of the "SF (or short-form) family" of instruments representing an international

benchmark for health outcomes measurement generally accepted by the Food and Drug Administration as valid measures of health outcomes that can be used in clinical studies.

Background

With roots in the landmark Health Insurance Experiment (HIE) and Medical Outcomes Study (MOS), the SF-36® Health Survey was constructed to be a comprehensive yet practical measure that achieves reductions in response burden without sacrificing psychometric standards of reliability, validity, and measurement precision.

The HIE's main goal was to construct the best possible scales for measuring a broad array of functional status and well-being concepts for group-level longitudinal analyses of data from children and adults. Results from data collected between 1974 and 1981 clearly demonstrated the potential of scales constructed from self-administered surveys to be reliable and valid tools yielding high-quality data for assessing changes in health status in the general population.

The MOS was a 4-year longitudinal observational study from 1986 through 1990 of the variations in practice styles and of health outcomes for more than 23,000 chronically ill patients. It provided a large-scale test of the feasibility of self-administered patient questionnaires and generic health scales among adults with chronic conditions, including the elderly, and attempted to answer two questions resulting from the HIE: (1) Could methods of data collection and scale construction such as those used in the HIE work in sicker and older populations? (2) Can more efficient scales be constructed?

The eight health domains represented in the SF profile were selected from 40 domains that were included in the MOS. The health domains chosen represented those most frequently measured in widely used health surveys and believed to be most affected by disease and health conditions. The SF-36® Health Survey was first made available in "developmental" form in 1988 and released in final original form in 1990 by its principal developer, John E. Ware Jr., Ph.D.

In 1991, the SF-36® Health Survey was selected for the International Quality of Life Assessment (IQOLA) Project, an organized effort to expand the use of health status instruments worldwide. The goal was to develop validated translations of a single health status questionnaire that could be used in multinational clinical studies and other international studies of health. By 1993, 14 countries were represented in the IQOLA Project. Interest in developing translations of the tool continued such that it had been translated for use in more than 90 country/language versions by 2006.

SF-36® Health Survey (Version 2)

Although the original SF-36® Health Survey form proved to be useful for many purposes, 10 years of experience revealed the potential for improvements. A need to improve item wording and response choices identified through the IQOLA Project, as well as a need to update normative data, led to development of the SF-36® Health Survey (Version 2).

In 1998, the SF-36® Health Survey (Version 2) was made available with the following improvements: (1) improved instructions and item wording; (2) improved layout of questions and answers; (3) increased comparability in relation to translations and cultural adaptations, and minimized ambiguity and bias in wording; (4) five-level response options in place of dichotomous choices for seven items in the Role-Physical and Role-Emotional scales; and (5) simplified response options for the Mental Health and Vitality scales. Without increasing the number of questions, improvements make the survey easier to understand and complete and substantially increase the reliability and validity of scores over a wider range, thereby reducing the extent of floor and ceiling effects in the role performance scales.

The SF-36® Health Survey (Version 2) is part of the "SF family" of patient-reported outcomes measures for adults—including the SF-8™ Health Survey, SF-12® Health Survey, SF-12® Health Survey (Version 2), SF-36® Health Survey, and DYNHA® Generic Health Assessment (a dynamic or "computerized adaptive" instrument)—which are all cross-calibrated and scored on the same norm-based metric to maximize their comparability.

Scales and Component Summaries

Physical Functioning (PF)

The content of the 10-item PF scale reflects the importance of distinct aspects of physical functioning and the necessity of sampling a range of severe and minor physical limitations. Items represent levels and

kinds of limitations between the extremes of physical activities, including lifting and carrying groceries; climbing stairs; bending, kneeling, or stooping; and walking moderate distances. One self-care item is included. The PF items capture both the presence and extent of physical limitations using a three-level response continuum.

Role-Physical (RP)

The four-item RP scale covers an array of physical-health-related role limitations in the kind and amount of time spent on work, difficulties performing work, and level of accomplishment associated with work or other usual activities.

Bodily Pain (BP)

The BP scale is composed of two items: one pertaining to the intensity of bodily pain and one measuring the extent of interference with normal work activities due to pain.

General Health (GH)

The GH scale consists of five items, including a general rating of health (“excellent” to “poor”) and four items addressing the respondent’s views and expectations of his or her health.

Vitality (VT)

This four-item measure of vitality was developed to capture ratings of energy level and fatigue.

Social Functioning (SF)

This two-item scale measures the effects of health on quantity and quality of social activities and, specifically, the impact of either physical or emotional problems on social activities.

Role-Emotional (RE)

The three-item RE scale covers mental-health-related role limitations assessing time spent on, level of accomplishment associated with, and level of care in performing work or other usual activities.

Mental Health (MH)

The five-item scale includes one or more items from each of four major mental health dimensions (anxiety, depression, loss of behavioral/emotional control, and psychological well-being).

Reported Health Transition (HT)

The survey includes a general health item that requires respondents to rate the amount of change they experienced in their health in general over a 1-year period. This item is not used to score any of the eight multi-item scales or component summary measures but provides information about perceived changes in health status that occurred during the year prior to the survey administration.

Physical and Mental Component Summary (PCS and MCS)

The aggregate of the scales is referred to as “component” summaries because the scales were derived and scored using principal components analysis. Although they reflect two broad components or aspects of health—physical and mental—*all* the eight scales are used to score *both* component summary measures.

All items, scales, and summary measures are scored so that a higher score indicates a better health state.

Norm-Based Scoring

The SF-36® Health Survey originally produced eight scales with scores ranging from 0 to 100 and norm-based PCS and MCS scores. The improved SF-36® Health Survey (Version 2) produces norm-based scores for all eight scales and the two component summaries, easing interpretation and score comparability.

Norm-based scoring linearly transforms the scales and summary measures to have a mean of 50 and a standard deviation of 10 in the 1998 U.S. general population. Thus, scores above and below 50 are above and below the average, respectively, in the 1998 U.S. general population. Also, because the standard deviation is 10, each one-point difference or change in scores has a direct interpretation; that is, it is one tenth of a standard deviation or an effect size of 0.10.

Scoring Software

Scoring instructions for the eight scales, the PCS and MCS measures, the reported HT item, and the optional RCI are published in the *User’s Manual for the SF-36® Health Survey (Version 2)*, Second

Edition. QualityMetric offers scoring services for Version 2 and the other SF instruments through the QualityMetric Health Outcomes™ scoring software. Among the features of the software is its ability to remove bias in estimates of scores for those having one or more missing responses and to enable score estimation for virtually all respondents regardless of the amount of missing data. In addition, the scoring software conducts data quality evaluations (i.e., data completeness, responses outside range, response consistency index, percentage of estimable scale scores, item internal consistency, item discriminant validity, scale reliability) and allows users of the SF-36® Health Survey (Versions 1 and 2) to make direct comparisons of scores across data sets that use different versions of the SF surveys and to published norms obtained on either form.

Reliability and Validity

Years of empirical research have demonstrated the reliability and validity of the SF-36® Health Survey, which is summarized in several user's manuals and thousands of articles. For the SF-36® Health Survey (Version 2), this tradition was continued by retaining item content from the original survey, making past empirical work on the reliability and validity of the tool generalizable to the SF-36® Health Survey (Version 2).

Evidence of the survey's internal, alternative forms and test-retest (Version 1) reliability has been documented in peer-reviewed articles and the *User's Manual for the SF-36® Health Survey (Version 2)*, Second Edition. To summarize, internal consistency (Cronbach's alpha) estimates using data from the 1998 U.S. general population ranged from .83 to .95 across the eight scales and summary component measures (internal consistency reliability estimates for the summary components take into account the reliability of and covariances among the scales), all exceeding the recommended minimum standard (.70) for group-level comparison of scores. Overall, reliability estimates for general population subgroups are also favorable and higher for component summary estimates than the eight scales. Studies of the reliability of alternative forms indicate that the SF-36® Health Survey (Version 2) is a comparable yet improved version of the original. Correlations (ranging from .76 to .93) between scales and related DYNHA domain item banks corrected for item overlap provide further evidence of alternative forms' reliability. Formal studies of the test-retest reliability of the SF-36® Health Survey (Version 2) have not yet been conducted, but studies of the original tool's scales indicate reliability that exceeds the recommended standard for measures used in group comparisons. Because four of the scales were improved and the other four scales remained unchanged during the development of the SF-36® Health Survey (Version 2), reliability estimates reported in the original tool's studies may be interpreted as representing the lower limits of scale reliabilities for the SF-36® Health Survey (Version 2).

Evidence of the tool's *construct validity* has been documented in studies involving factor analysis, item-scale correlations, interscale correlations, correlations of the scales with the component summary measures and the SF-6D, and known-groups comparisons. *Criterion validity* has been demonstrated through the correlations of each scale with the theta score for its associated DYNHA® item bank. Data on the likelihood of future events (e.g., job loss, psychiatric treatment) based on scale score ranges also provide evidence of criterion validity. *Content validity* has been shown through a comparison of the SF-36® Health Survey's (Version 2) coverage of health domains to the health domain coverage of other general health surveys. The validity of the tool is fully documented in the *User's Manual for the SF-36® Health Survey (Version 2)*, Second Edition, and further documented in peer-reviewed articles by the developer and in numerous studies from the research literature (Versions 1 and 2).

Interpretation

Generally, interpretation of the SF-36® Health Survey's profile begins by determining if the norm-based scores for the PCS and MCS measures deviate from what is considered the "average" range for the general U.S. population. This is followed by an examination of the scale scores to make a similar determination. Each of these decisions is based on separate, empirically based individual patient- and group-level guidelines. Unlike previous presentations of the SF profile, the current profile now begins with a presentation of the results of the PCS and MCS measures, emphasizing the importance of first considering findings from these more general measures of health status (see Figure 1). It also facilitates interpretation by immediately establishing what the general burden of illness or effects of treatment are (i.e., physical or mental) before examining the more specific scales. As

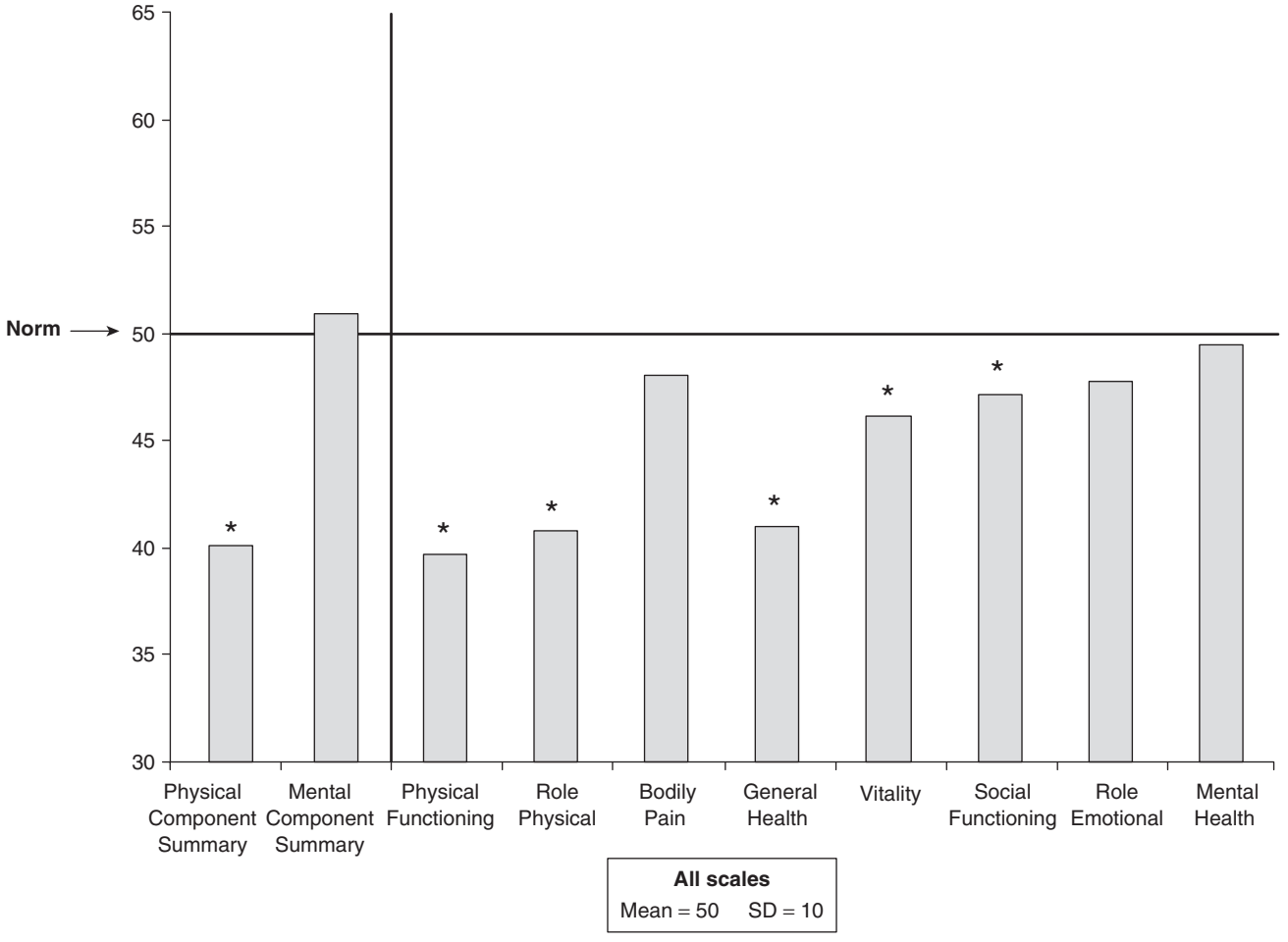


Figure 1 SF-36® Health Survey Profile of Norm-Based Scores (Adult Asthma Sample)

Source: Adapted from Okamoto, Noonan, DeBoisblanc, and Kellerman (1996).

Note: Norm significantly higher.

their labels suggest, the PCS and MCS scores provide a summary of the respondent's health status from both a broad physical health perspective and a broad mental health perspective, respectively.

In addition, the application of measure- or scale-specific standard errors of measurement allows one to determine, within specific levels of confidence, intervals in which the respondent's true score on every measure and scale falls. Guidelines for interpreting high and low scores on the PCS and MCS measures and on each scale, guidelines for determining the minimally important difference, score cutoffs for determining the likelihood of the presence of a physical or mental disorder, and U.S. general population norms for age, gender, age-by-gender, and combined groups for both the standard and acute forms are provided in the *User's Manual for the SF-36® Health Survey (Version 2)*, Second Edition.

Applications

Applications of the SF-36® Health Survey (Versions 1 and 2) include the following:

- Clinic-based evaluation and monitoring of individual patients
- Population monitoring

- Estimating the burden of disease (by standardizing questions, answers, and scoring, reliable and valid comparisons can be made to determine the relative burden of different conditions in several domains of health)
- Evaluating treatment effects in clinical trials
- Disease management and risk prediction (i.e., the ability to predict health outcomes, hospitalization, future medical expenditures, resource use, job loss and work productivity, future health, risk of depression, use of mental health care, future health, and mortality)
- Cost effectiveness
- Enhancing patient-provider relations
- Providing direct-to-consumer information (i.e., educating the public about medical conditions, their symptoms and effects, and potential treatment options; prompting recognition or detection of personal health problems that may benefit from clinical consultation, thereby encouraging more appropriate care seeking, case finding, and physician-patient dialogue; and promoting self-care and compliance with treatment regimens)

—Diane M. Turner-Bowker, Michael A. DeRosa, and John E. Ware Jr.

See also Functional Status; Global Burden of Disease Project; Measurement; Missing Data Methods; Quality of Life, Quantification of

Further Readings

Gandek, B., & Ware, J. E., Jr. (Eds.). (1998). Translating functional health and well-being: International quality of life assessment (IQOLA) project studies of the SF-36 Health Survey [Special issue]. *Journal of Clinical Epidemiology*, *51*(11).

Okamoto, L. J., Noonan, M., DeBoisblanc, B. P., & Kellerman, D. J. (1996). Fluticasonepropionate improves quality of life in patients with asthma requiring oral corticosteroids. *Annals of Allergy, Asthma and Immunology*, *76*, 455–461.

Ware, J. E., Jr. (2000). SF-36 Health Survey update. *Spine*, *25*, 3130–3139.

Ware, J. E., Jr., & Kosinski, M. (2001). *SF-36 Physical & Mental Health Summary Scales: A manual for users of version 1* (2nd ed.). Lincoln, RI: QualityMetric.

Ware, J. E., Jr., Kosinski, M., Bjorner, J. B., Turner-Bowker, D. M., & Maruish, M. E. (2006). *User's manual for the SF-36® Health Survey (Version 2)* (2nd ed.). Lincoln, RI: QualityMetric.

Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, *30*, 473–483.

Ware, J. E., Jr., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36® Health Survey: Manual and interpretation guide.* Boston: Health Institute.

Web Sites

Further information about tools from the SF family of instruments is available from http://www.qualitymetric.com or http://www.sf36.org.

The sf36.org Web site is a community forum for users of the SF tools and offers news, events, online discussion, and a searchable database of SF publications.

Sick Building Syndrome

Sick building syndrome (SBS) is a term applied to a situation in which some or all the people occupying a building (usually working or living in it) experience unpleasant health and comfort effects such as headache; dizziness; nausea; irritated eyes, nose, or throat; dry cough; or skin irritation. The term is sometimes applied to the symptoms themselves also. These effects may be localized to a part of the building or be present throughout. The definition of SBS requires that the symptoms disappear soon after leaving the building and that they cannot be ascribed to a specific cause or illness. SBS is differentiated from building-related illness, which describes diagnosable illness whose cause can be attributed to airborne contaminants within a building. SBS is usually assumed to be caused by poor indoor air quality (IAQ). It was first identified in the 1970s, and a 1984 report by the World Health Organization suggested that up to 30% of new and remodeled buildings may have problems with IAQ sufficient to cause health symptoms. Inadequate building ventilation is the most common cause; the appearance of SBS in the mid-1970s has often been attributed to

decreased ventilation standards for commercial buildings to increase energy efficiency during the oil embargo of 1973. Chemical contaminants are also potential contributors to SBS; these include volatile organic compounds emitted by carpeting, upholstery, cleaning agents, and other sources and combustion products including particulate matter and carbon monoxide produced from heating devices such as fireplaces and stoves. Biological contaminants such as molds, pollen, viruses, bacteria, and animal or bird droppings can also contribute to SBS.

Investigation of SBS requires first ascertaining whether the complaints are actually due to IAQ; if so, the investigation will gather information about the building's ventilation, heating and air conditioning system, possible sources of internal contaminants, and possible pathways for exterior pollutants to enter the building. Air sampling alone rarely provides sufficient information to solve the problem, because in SBS buildings contaminant concentration levels rarely exceed existing standards. The most common solutions to SBS include removing a known source of pollution, increasing ventilation rates and air distribution, and adding air cleaning devices.

SBS is a difficult condition to study because its symptoms are commonplace and could have many causes, such as allergies or stress, and may be influenced by psychological factors, such as dislike of a job or workplace. In addition, because many different aspects of the indoor environment can contribute to SBS, it is often difficult to identify the cause or causes for a particular case, and extensive renovations may fail to solve the problem. There is also a natural opposition between the interests of building owners and occupants in a case of suspected SBS. The occupants may believe SBS is causing their health symptoms and demand building inspections and modifications, while the owner may not believe that the building is the cause of their symptoms and may therefore be reluctant to pay for any inspections or alterations. Because of the aforementioned difficulties in verifying SBS and identifying its cause, the "truth of the matter" may never be unequivocally determined. In addition, some clinicians believe that SBS is not a meaningful term and should be abandoned, while others have argued that investigations into SBS should include evaluation of psychological and social as well as physical, environmental, and biomedical factors.

—Sarah Boslaugh

See also Environmental and Occupational Epidemiology; Pollution

Further Readings

Stenberg, B., & Wall, S. (1995). Why do women report "sick building symptoms" more often than men? *Social Science and Medicine*, *40*, 491–502.

Thorn, A. (1998). The sick building syndrome: A diagnostic dilemma. *Social Science and Medicine*, *47*, 1307–1312.

U.S. Environmental Protection Agency. (n.d.). *Indoor air facts no. 4 (revised): Sick building syndrome (SBS).* Retrieved August 9, 2007, from http://www.epa.gov/iaq/pubs/sbs.html.

Significance Testing

See Hypothesis Testing

Simpson's Paradox

Simpson's paradox is an extreme form of confounding, where the association between two variables in a full group is in the opposite direction of the association found within every subcategory of a third variable. This paradox was first described by G. U. Yule in 1903 and later developed and popularized by E. H. Simpson in 1951.

By way of example, consider a new drug treatment that initially appears to be effective, with 54% of treated patients recovering, as compared with 46% of patients receiving a placebo. However, when the sample is divided by gender, it is found that 20% of treated males recover compared with 25% of placebo males, and 75% of treated females recover as compared with 80% of placebo females. So the apparent paradox is that the drug is found to be more effective than the placebo in the full group but less effective than the placebo in each of the two gender-specific subgroups that fully comprise the combined group.

The key to unraveling this puzzle involves the gender confound—differing numbers of patients of each gender receiving the treatment versus placebo, combined with differing overall recovery rates for males versus females. Table 1 shows that in this example males are 1.6 times more likely to receive the placebo than the treatment, whereas females are 1.6 times

Table 1 Recovery Rates

	Treatment	*Placebo*
Male	10/50 (20%)	20/80 (25%)
Female	60/80 (75%)	40/50 (80%)
All patients	70/130 (54%)	60/130 (46%)

more likely to receive the treatment than the placebo. At the same time, females are more than three times as likely to recover as males within both the treatment group and the placebo group. In other words, females are relatively easy to cure. So the fact that the placebo is more effective than the treatment in both groups is obscured when the groups are combined, due to the disproportionate number of easy-to-cure females in the treatment group.

In this particular example, it would be commonly agreed that the correct conclusion involves the subgroup-specific results—the drug is not effective—and that the apparent effectiveness found in the combined group is merely a statistical artifact of the study design due to the gender confound.

Simpson's paradox can be problematic when not recognized, leading to naive and misleading conclusions regarding effectiveness or other relations studied. Perhaps more ominously, knowledge of Simpson's paradox can be intentionally used to present or emphasize results that support a desired conclusion, when that conclusion is not valid. More generally, Simpson's paradox has been shown to have implications for the philosophical study of causation and causal inference. In practical terms, it is prudent for both researchers and research consumers to be on guard for this potentially perilous paradox.

—Norman A. Constantine

See also Confounding; Ecological Fallacy

Further Readings

Malinas, G., & Bigelow, J. (2004, Spring). Simpson's paradox. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved August 30, 2006, from http://plato.stanford.edu/archives/spr2004/entries/paradox-simpson.

Reintjes, R., de Boer, A., van Pelt, W., & Mintjes-de Groot, J. (2000). Simpson's paradox: An example from hospital epidemiology. *Epidemiology*, *11*, 81–83.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Association*, *13*, 238–241.

Wainer, H., & Brown, L. M. (2004). Two statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. *American Statistician*, *58*, 117–123.

Skewness

If a person says that there is a car parked "askew" in a lined parking space, then we know the car is not parked straight and is closer to one side than the other. This common usage of the idea of skewness carries over into the statistical definition. If a statistical distribution is skewed, then more of the values appear in one end of the distribution than the other. Skewness, then, is a measure of the degree and direction of asymmetry in a distribution. Skewness is also called the third moment about the mean and is one of the two most common statistics used to describe the shape of a distribution (the other is kurtosis). Skewed distributions have values bunched at one end and values trailing off in the other direction. The most commonly used measure of skewness is the Pearson coefficient of skewness.

There are three types of skewness: right, left, or none. Often, these are referred to as positive, negative, or neutral skewness, respectively. Often, people are confused about what to call the skewness. The name of the type of skewness identifies the direction of the longer tail of a distribution, not the location of the

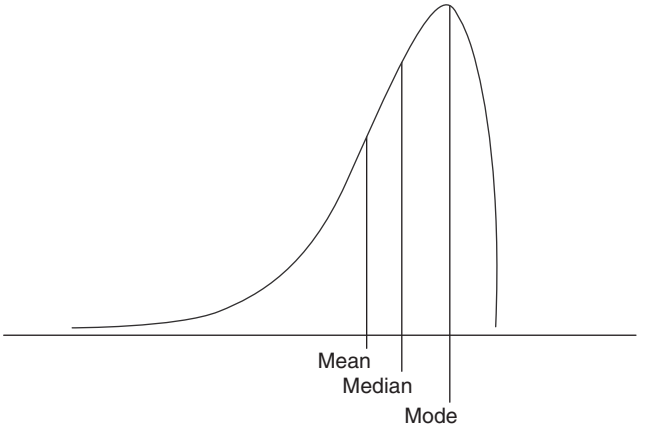


Figure 1 Negatively Skewed Distribution

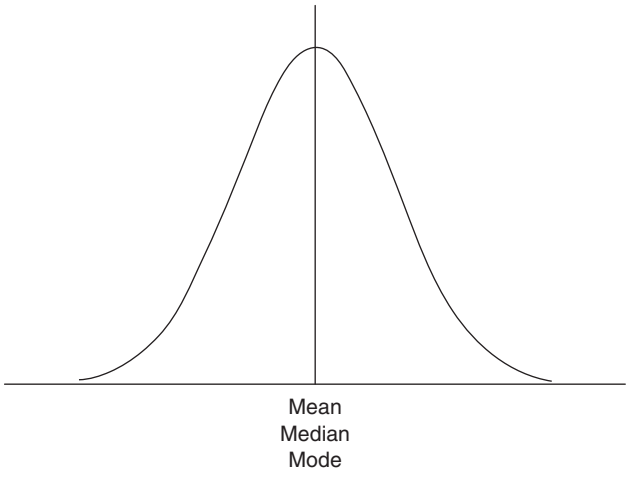


Figure 2 Neutrally Skewed Distribution

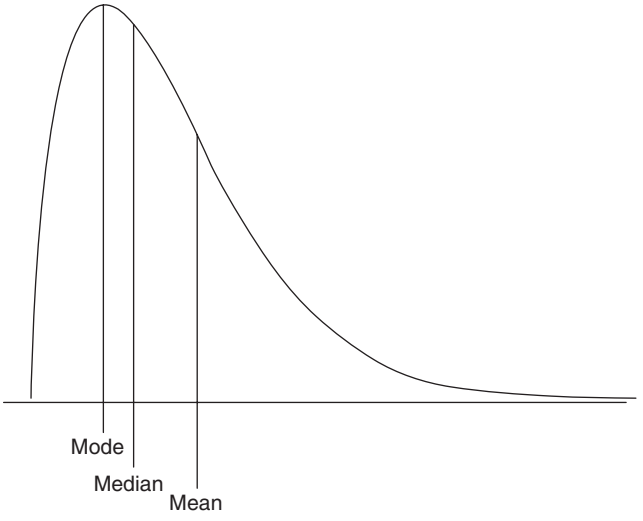


Figure 3 Positively Skewed Distribution

larger group of values. If a distribution is negatively or left skewed, then there are values bunched at the positive or right end of the distribution, and the values at the negative or left end of the distribution have a longer tail. If a distribution is positively or right skewed, then there are values bunched at the negative or left end of the distribution, and the values at the positive or right end have a longer tail.

The most commonly known distributions and their type of skewness are given in Table 1.

Often, skewness is used to help assess whether a distribution being studied meets the normality assumptions of most common parametric statistical tests. While the normal distribution has a skewness of 0, it is important to realize that, in practice, the skewness statistic for a sample from the population will not be exactly equal to 0. How far off can the statistic be from 0 and not violate the normality assumption? Provided the statistic is not grossly different from 0, then that decision is up to the researcher and his or her opinion of an acceptable difference. For most typically sized samples, values of the Pearson coefficient of skewness between -3 and $+3$ are considered reasonably close to 0. To accurately measure the skewness of a distribution, sample sizes of several hundred may be needed.

If a researcher determines that the distribution is skewed, then reporting the median rather than or along with the mean provides more information about the central tendency of the data. The mean is sensitive to extreme values (those skewed), while the median is robust (not as sensitive).

Table 1 Common Distributions With Type of Skewness

Distribution	*Type of Skewness*
Normal	Neutral
Student's *t*	Neutral
Uniform	Neutral
Exponential	Positive
Laplace	Neutral
Weibull	Depends on the parameter values; may be negative, neutral, or positive

—Stacie Ezelle Taylor

See also Inferential and Descriptive Statistics; Kurtosis; Measures of Central Tendency; Normal Distribution

Further Readings

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness & kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*, 183–189.

Keller, D. K. (2006). *The Tao of statistics: A path to understanding (with no math)*. Thousand Oaks, CA: Sage.

Sleep Disorders

An estimated 70 million people in the United States suffer from sleep problems, and more than half of those with sleep problems have a sleep disorder that is chronic. The four most prevalent sleep disorders are insomnia, obstructive sleep apnea, narcolepsy, and periodic limb movements in sleep, with sleep apnea accounting for nearly 80% of all sleep diagnoses in sleep centers in the United States. About 30 million American adults have frequent or chronic insomnia. Approximately 18 million have obstructive sleep apnea, but only 10% to 20% have been diagnosed. An estimated 250,000 people have narcolepsy, and more than 5% of adults are affected by periodic limb movements in sleep syndrome. Sleep disorders have major societal impacts. Each year, sleep disorders, sleep deprivation, and excessive daytime sleepiness (EDS) add approximately $16 billion annually to the cost of health care in the United States and result in $50 to $100 billion annually in lost productivity (in 1995 dollars). According to the National Highway Traffic and Safety Administration, 100,000 accidents and 1,500 traffic fatalities per year are related to drowsy driving. Nearly two thirds of older Americans have sleep difficulties, and the prevalence of sleep problems will increase as the older adult population increases. The 1990s has seen a significant increase in our awareness of the importance of diagnosing and treating sleep disorders. The prevalence rates, risk factors, and treatment options will be reviewed for each of the four major sleep disorders.

Insomnia

Insomnia is the most commonly reported sleep complaint across all stages of adulthood. An estimated 30 million American adults suffer from chronic insomnia, and up to 57% of noninstitutionalized elderly experience chronic insomnia. In the United States, total direct costs attributable to insomnia are estimated at $12 billion for health care services and $2 billion for medications. Emerging evidence suggests that being female and old age are two of the more common risk factors for the development of insomnia; other predisposing factors include excess worry about an existing health condition, lower educational level, unemployment, and separation or divorce. Insomnia is comorbid with anxiety and depressive disorders and may lead to the development of psychiatric disorders. Insomnia is correlated with high levels of medical use and increased drug use, as well as increased psychosocial disruption including poor work performance and poor memory.

Insomnia Treatments

Traditional management of insomnia includes both pharmacologic and nonpharmacologic treatments. Current guidelines suggest that chronic insomnia be treated with a combination of nonpharmacologic interventions, such as sleep hygiene training, relaxation training, stimulus control training, cognitive-behavioral therapy, or sleep restriction/sleep consolidation therapy, and pharmacologic interventions. Medications prescribed for insomnia range from newer agents such as zolpidem, zaleplon, and eszoplicone to older agents such as antidepressants (e.g., amitriptyline or trazodone) and benzodiazepines (e.g., clonazepam, lorazepam). Medications are not typically indicated for long-term treatment of insomnia, except for a medication recently approved by the Food and Drug Administration, eszoplicone.

Obstructive Sleep Apnea

Obstructive sleep apnea (OSA) is a medical condition characterized by repeated complete (apnea) or partial (hypopnea) obstructions of the upper airway during sleep. It is prevalent in 2% to 4% of working, middle-aged adults, and an increased prevalence is seen in the elderly ($\sim$24%), veterans ($\sim$16%), and African Americans. Being an obese male is the number one major risk factor for OSA. The risk of OSA increases significantly with increased weight, and more than 75% of OSA patients are reported to be more than 120% of ideal body weight. Other risk factors that can contribute to OSA include anatomical abnormalities of the upper airway (e.g., large uvula, enlarged tonsils, large neck circumference). Estimates of health care costs for OSA patients are approximately twice that of matched, healthy controls. This cost difference is evident several years prior to the diagnosis. OSA is associated with a higher mortality rate.

Consequences of OSA

OSA is associated with several cardiovascular diseases, most notably hypertension, ischemic heart disease, heart failure, stroke, cardiac arrhythmias, and

pulmonary hypertension. Compared with the general population, OSA patients have twice the risk for hypertension, three times the risk for ischemic heart disease, and four times the risk for cerebrovascular disease. The evidence supporting the link between OSA and hypertension is compelling, with OSA now officially recognized as an identifiable cause of hypertension. Alterations in sleep architecture cause sleep to be nonrestorative, resulting in mild to severe EDS. EDS and/or hypoxia due to OSA are associated with a number of neurocognitive, mood, and behavioral consequences, including lowered health-related quality of life, impaired cognitive performance, impaired driving ability (two to seven times increased risk of a motor vehicle accident), dysphoric mood, psychosocial disruption (e.g., more intensely impaired work performance and higher divorce rates), and disrupted sleep and impaired quality of life of spouses of OSA patients.

OSA Treatments

The goals of any OSA treatment are the elimination of breathing events and snoring, maintaining high blood oxygen levels, and improving symptoms. Categories of OSA treatments include medical devices (continuous positive airway pressure therapy and oral appliances), behavioral recommendations (weight loss, positional therapy), and surgical procedures. Nasal continuous positive airway pressure (CPAP) is the treatment of choice for this condition, with meta-analytic reports of numerous randomized controlled trials showing that CPAP improves both objectively and subjectively measured daytime sleepiness as well as health-related quality of life. CPAP has been shown to normalize sleep architecture and reduce blood pressure. Oral appliances (OAs) alter the oral cavity to increase airway size and improve patency. OAs reduce the number of apneas and hypopneas and reduce sleepiness levels. Weight loss helps reduce the number of apneas and hypopneas in obese OSA patients, reduces oxygen desaturations, and improves sleep architecture. Positional therapies, primarily indicated for mild sleep apnea, are based on the observation that most disordered breathing occurs in the supine (i.e., lying on the back) position, so the therapy encourages sleep in the prone (i.e., lying face downward) or side positions. There are a wide variety of surgical treatments that are now considered secondary treatments if other treatments do not work well or are not well tolerated.

Narcoplepsy

Narcolepsy is a chronic neurological disorder caused by the brain's inability to regulate sleep-wake cycles normally. It is estimated that approximately 250,000 adult Americans are affected by narcolepsy. Narcolepsy is the most common neurological cause of EDS. Direct medical costs for narcolepsy can cost the patient more than $15,000 per year. The impact of narcolepsy is often more severe than that of other chronic diseases, such as epilepsy. Genetics may play a large role, with first-degree relatives having a 40-fold increased risk for narcolepsy. Men and women appear to be at equal risk.

The two most common symptoms of narcolepsy are EDS and cataplexy. Cataplexy is a sudden loss of muscle tone and strength, usually caused by an extreme emotional stimulus.

Narcoleptic patients also can experience sleep paralysis, falling asleep at inappropriate times (conversations, dinner), psychosocial problems, and EDS. EDS comprises both a strong background feeling of sleepiness and sometimes an irresistible urge to sleep suddenly. These sudden naps associated with narcolepsy can last minutes to an hour and occur a few times each day. Furthermore, as a consequence of EDS, patients with narcolepsy often report problems with inattention, blurred vision, cataplexy, poor memory, and driving without awareness (automatic behaviors).

Narcolepsy Treatments

Because narcolepsy is a chronic condition, treatment focuses on long-term symptom management through medications and behavioral treatments. Medications for treatment of narcolepsy are aimed at managing the daytime symptoms of the disorder. EDS can be reduced by a newer, nonamphetamine "wake promoting agent" named modafinil and by amphetamine derivatives (dexamphetamine, methylphenidate). Side effects from the amphetamine-type drugs are common and include tolerance, irritability, and insomnia. Drugs suppressing rapid eye movement sleep can help in reducing cataplexy; the newest one is xyrem (gamma-hydroxybutyrate). The goals of behavioral therapies are to promote behaviors that can alleviate daytime symptoms. The primary therapy is planned research showing that scheduled daytime naps are effective in helping reduce daytime sleepiness.

Periodic Limb Movements in Sleep

Periodic limb movements in sleep (PLMS) is a sleep phenomenon characterized by periodic episodes of repetitive and highly stereotyped limb movements. Periodic limb movements are defined by their occurrence in a series (four or more) of similar movements with a wide range of periods and duration between 0.5 and 5.0 s. It has been estimated that 5% of those below the age of 50 years will have PLMS, while more than 30% of individuals aged above 65 years may have a significant number of PLMS. PLMS may begin at any age although prevalence increases markedly in elderly healthy people. In patients with periodic limb movement disorder, insomnia and EDS are common complaints. There is significant overlap between PLMS and restless legs syndrome (RLS), with more than 80% of RLS patients having PLMS as well.

PLMS Treatments

Treatment of PLMS consists primarily of pharmacological and secondarily of nonpharmacological interventions. Pharmacological agents recommended for use include dopaminergic agents, anticonvulsants, opioids, and sedatives/hypnotics. Nonpharmacological treatment of PLMS primarily consists of advising the patient of good sleep hygiene.

—Carl Stepnowsky and Joe Palau

See also Aging, Epidemiology of; Hypertension; Obesity; Vehicle-Related Injuries

Further Readings

Buscemi, N., Vandermeer, B., Friesen, C., Bialy, L., Tubman, M, Ospina, M., et al. (2005, June). *Manifestations and management of chronic insomnia in adults*. Evidence Report/Technology Assessment No. 125. AHRQ Publication No. 05-E021-2. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved August 9, 2007, from http://www.ahrq.gov/clinic/epcsums/insomnsum.pdf.

Chesson, A. L., Jr., Wise, M., Davila, D., Johnson, S., Littner, M., Anderson, W. M., et al. (1999). Practice parameters for the treatment of restless legs syndrome and periodic limb movement disorder. An American Academy of Sleep Medicine Report. Standards of Practice Committee of the American Academy of Sleep Medicine. *Sleep*, *22*, 961–968.

Kushida, C. A., Littner, M. R., Hirshkowitz, M., Morgenthaler, T. I., Alessi, C. A., Bailey, D., et al. (2006). Practice parameters for the use of continuous and bilevel positive airway pressure devices to treat adult patients with sleep-related breathing disorders. *Sleep*, *29*, 375–380.

National Center for Sleep Disorders Research and National Highway and Traffic Safety Expert Panel on Driver Fatigue and Sleepiness. (1998). *Drowsy driving and automobile crashes: Report and recommendations*. Washington, DC: Author.

Thorpy, M. (2001). Current concepts in the etiology, diagnosis and treatment of narcolepsy. *Sleep Medicine*, *2*, 5–17.

Smallpox

Smallpox, a contagious disease produced by the variola virus (genus *Orthopoxvirus*), was eradicated in 1977. The word *smallpox* is believed to come from the Latin word *pocca* meaning "pouch," and *variola* from *varius* or *varus* meaning "spotted pimple." There were three subspecies of variola: variola major, intermedius, and minor. The milder form of the disease, variola minor, had a case-fatality rate of less than 1%, whereas the rate for variola major was 25% to 50%. Ten percent of the smallpox cases involved hemorrhagic smallpox that was quickly fatal.

Smallpox was found only in humans and was usually transmitted through droplet nuclei, dust, and fomites (inanimate objects such as blankets that can transmit germs). The incubation period was between 12 and 14 days, with the respiratory tract as the main site of infection. The prodrome, or early symptom of the development of smallpox, was a distinct febrile illness that occurred 2 to 4 days before eruptive smallpox. Rashes usually developed 2 to 4 days after being infected. The smallpox rash was centrifugal, found more on the head, arms, and legs than on the trunk area of the body. Smallpox was very disfiguring because crusts on the skin would form from the fluid and pus-filled spots on the body. Many who survived smallpox had bad scars on their face or were blinded. Because of the rash that formed in the majority of smallpox cases, surveillance of the disease was less problematic.

Common symptoms of smallpox were fever, headache, backache, malaise, and abdominal pain. The very young and very old were at higher risk of dying of smallpox. Exposure to smallpox usually occurred in either the family or hospital setting. Compared with

chickenpox and measles, smallpox was not as infectious. Persons with smallpox were infectious from the time fever arose until the last scab separated; they were not infectious during the incubation period. A patient who survived smallpox was resistant to the infection.

The origin of the smallpox virus is thought to have been 3,000 years ago in India or Egypt. For a very long time, smallpox epidemics were quite common, annihilating populations. Smallpox wreaked havoc on the royal houses of Europe between 1694 and 1774, with Queen Mary II of England, Emperor Joseph I of Austria, King Luis I of Spain, Tsar Peter II of Russia, Queen Ulrike Elenora of Sweden, and King Louis XV of France all dying of the disease. In the 18th century, 1 out of every 10 children born in Sweden and France died from smallpox. Many Native American tribes were annihilated by the smallpox epidemic that occurred around 1837. The Native American tribes were introduced to smallpox through European settlement.

There were a number of ancient practices that were introduced to prevent smallpox. Worshippers could pray to a deity such as the Indian goddess of smallpox, Shitala Mata, or to the Chinese goddess of smallpox, T'ou-Shen Niang-Niang. Roman Catholic Europeans could pray to St. Nicaise, the patron saint of smallpox. There was also a widespread notion that red-colored objects could combat smallpox. The Red Treatment, as it was called, used red objects such as a red cloth hung in a room of smallpox victims in an attempt to prevent smallpox. Another method tried was inoculation of infectious matter from smallpox victims implanted into patients. It was not until 1796 that an English physician by the name of Edward Jenner (1749–1823) discovered a vaccination for smallpox.

Jenner discovered that immunity against smallpox was possible by injection of the cowpox virus into the system. He observed that a patient who had contracted coxpox by milking cows with cowpox lesions on their teats resisted variolation. Milkmaids had scars on their hands from previous cowpox infections, and Jenner noted that these women were immune from developing smallpox. He invented the vaccine and initiated a new field of medicine—preventative medicine. He successfully immunized an 8-year-old boy with a substance that was taken from a cowpox sore on the hand of a milkmaid, Sarah Nelmes. Jenner published his work *An Inquiry into the Causes and Effects of the Variolae Vaccinae, a Disease Known by the Name of Cox Pox* in 1798.

In 1966, the World Health Organization (WHO) led the Smallpox Eradication Program, an intensive effort to eradicate smallpox throughout the world. When the campaign started, smallpox was still found in 41 countries in the world. The eradication program focused on using a standardized vaccine and extensive public education. Furthermore, eradication was facilitated through active surveillance of cases that identified the location of the disease and through attempts to vaccinate all those who were in areas with widespread infection. Advances in technology helped the eradication efforts by making it possible to freeze-dry vaccines so that they did not need to be refrigerated. Other technological advances that aided in the eradication program included adaptation of the jet injector, which allowed for smallpox vaccine to be given intradermally starting in the early 1960s, and the development of the bifurcated needle in 1968.

In 1949, the last case of smallpox occurred in the United States. The last case of naturally occurring smallpox in the world was found in October 1977 in Somalia. Two cases associated with a virologic laboratory were reported in England in 1978. Since then, smallpox has no longer appeared in the world. In May 1980, at the 33rd World Health Assembly, the WHO declared victory in its fight for global eradication of smallpox. Smallpox was the first disease ever to be eradicated by humans.

There is growing concern that smallpox may be used by terrorists as a biologic weapon in warfare. Because of the threat, active duty service members continue to be vaccinated against smallpox. There are two remaining stocks of the smallpox virus: at the Centers for Disease Control and Prevention in Atlanta, Georgia, and at the Institute for Viral Preparations in Moscow, Russia. There is only a limited supply in the world of the smallpox vaccine, which protects against the smallpox virus for 10 years. Given that many of the historical cases of smallpox occurred at a time with significant population immunity from vaccination or having had the virus, the world population today is more susceptible to smallpox. Many countries are considering increasing their supply due to the threat of a biological terrorist attack using smallpox, for which there is no effective treatment.

—Britta Neugaard

See also Bioterrorism; Jenner, Edward; World Health Organization

Further Readings

Bazin, H. (2000). *The eradication of smallpox: Edward Jenner and the first and only eradication of a human infectious disease*. London: Academic Press.

Behbehani, A. M. (1988). *The smallpox story in words and pictures*. Kansas City: University of Kansas Medical Center.

Gorbach, S. L., Bartlett, J. G., & Blacklow, N. R. (Eds.). (2004). *Infectious diseases* (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.

Hoeprich, P. D., Jordan, M. C., & Ronald, A. R. (Eds.). (1994). *Infectious diseases: A treatise of infectious processes* (5th ed.). Philadelphia: Lippincott.

Hopkins, D. R. (1988). Smallpox: Ten years gone. *American Journal of Public Health*, *78*, 1589–1595.

Snow, John
(1813–1858)

John Snow has an unusual place in medical history because he is a seminal figure in two medical disciplines—anesthesiology and epidemiology. His contribution to the first field was to establish the chemical and biological principles underlying the administration of consistent dosages of anesthetic gases effectively and with minimal toxicity. In the latter field, he discovered how cholera—and, by extension, every form of intestinal infection—was transmitted. The process by which he discovered the fecal-oral and waterborne routes of disease communication was the first true model of epidemiologic investigation.

Snow's twin accomplishments were not unrelated. As the world's first practicing anesthesiologist, he was intimately familiar with the effects of gases on human physiology. This understanding made him skeptical of the then-reigning dogma that miasmas—hypothesized gaseous emanations from rotting material that were inhaled—could cause disease at a distance. As so often in science, the first step in developing a new hypothesis was recognition of the limitations of the old.

Snow had cared for cholera patients as a teenage apprentice in 1832. When cholera made its second appearance in Europe in 1848, he published a small pamphlet and a two-page paper on cholera transmission. He argued that cholera was fundamentally a disease of the intestinal system and that its major symptoms were the result of fluid loss. This led him to conclude that the "agent" of cholera was ingested. He further reasoned, from much circumstantial evidence, that the "agent" was transmitted by accidental soiling of the hands by the colorless evacuations of cholera and that this transmission could be greatly multiplied if the evacuations found their way into water supplies.

The work that most epidemiologists recognize as Snow's signature achievement was undertaken during the third European epidemic, from 1854 to 1855. Snow examined mortality from cholera in two regions of London with overlapping water supplies. One water company (Lambeth) used the rural Thames above London as its source, while the other (Southwark and Vauxhall) took its water from the Thames downstream of the city's sewage effluent. Snow visited hundreds of houses in the region to determine their water supplies, enlisting the help of medical colleagues, including the public health official William Farr, and linking the source of water in each house to the number of deaths from cholera among its residents. He found that in the first weeks of the epidemic, death rates were 14 times higher in houses with water from Southwark and Vauxhall than in houses supplied by Lambeth. He also showed that in a major local outbreak in Soho, the source was almost certainly a shallow well supplying a widely used public pump on Broad Street. He persuaded local officials to remove the pump handle, but by then the outbreak was almost over. The well was later shown to have been fecal contamination from a leak from a nearby cesspool. Decades before the work of Pasteur and Koch, Snow speculated that the cholera "agent" could reproduce, that the duration of reproduction accounted for the incubation period, and that the agent probably had a structure like a cell.

Although to modern readers Snow seems eminently persuasive, his views on cholera were not widely accepted in his day. Cholera investigations later in the century, however, convinced many British and American physicians that water supplies were a central feature of cholera transmission.

—Nigel Paneth

See also Waterborne Diseases

Further Readings

Snow, J. (1848). *On the mode of communication of cholera*. London: Churchill.

Snow, J. (1848). On the pathology and mode of communication of cholera. *London Medical Journal*, 44, 745–752, 923–929.
Snow, J. (1855). *On the mode of communication of cholera* (2nd ed.). London: Churchill.

Social Capital and Health

As a possible determinant of population health, social capital has emerged as a topic of growing interest in the epidemiologic literature. Epidemiologic studies have explored the potential protective effects of social capital on a variety of health outcomes. This entry highlights the conceptualization of social capital, hypothesized mechanisms for its health effects, and features of the empirical evidence on the relations between social capital and health to date, including the measurement of social capital.

Conceptualization

Unlike financial capital, which resides in people's banks and in property, and human capital, which is embedded in people's education and job skills, social capital has been conceptualized to exist in people's relations to one another—that is, within social networks.

Conceptualizations of social capital have ranged from definitions focusing on the resources within social networks that can be mobilized for purposeful actions to definitions that encompass both social structures and associated cognitive resources such as trust and reciprocity. In addition to being categorized according to structural and resource characteristics, social capital has been dichotomized into many forms, including formal versus informal social capital (e.g., participation in labor unions vs. family dinners), inward-looking versus outward-looking social capital (e.g., chambers of commerce vs. the Red Cross), and bonding versus bridging social capital.

The distinction between bonding social capital and bridging social capital has probably gained the most prominence in the social capital and health literature. Bonding social capital refers to social capital within relationships between individuals with shared identities such as race/ethnicity and gender, whereas bridging social capital corresponds to social capital in relationships between individuals who are dissimilar.

There is ongoing debate among social capital scholars as to the extent to which social capital is primarily an individual-level social network asset (i.e., dwelling within individuals' family and friend relationships), a collective or public good, or both. Furthermore, not all social capital can be considered an unqualified benefit for all. The sociologist Alejandro Portes recognized the potential for "negative externalities" of social capital that could harm individuals outside of a group, yet produce benefits, or "positive internalities," for group members. For example, among residents in a predominantly African American, racially segregated neighborhood, individuals may experience positive effects from their relationships with one another but suffer negative consequences from discriminatory practices by outside individuals of other races/ethnicities.

Hypothesized Mechanisms

Several mechanisms by which social capital may affect health have been proposed. These include the diffusion of knowledge about health promotion, influences on health-related behaviors through informal social control, the promotion of access to local services and amenities, and psychosocial processes that provide support and mutual respect. Each of these mechanisms is plausible based on separate theories and pathways. First, through the theory of diffusion of innovations, it has been suggested that innovative behaviors diffuse much faster in communities that are cohesive and high in trust. Informal social control may also exert influence over such health behaviors. In social-cognitive theory, one's belief in collective agency is tied to the efficacy of a group in meeting its needs. For example, a neighborhood high in social capital and trust would be expected to effectively lobby together for local services, such as adequate transportation and green spaces. Finally, psychosocial processes, including social support and trust, may buffer the harmful effects of stress or may have direct positive effects on health.

Empirical Evidence

With origins in political science and sociology, social capital was first introduced to the public health literature in the early 1990s. Interest on this topic subsequently expanded, and beginning in the late 1990s, there was a surge in the body of literature examining the associations between social capital and health outcomes.

This literature may be broadly classified according to the level at which the measure of social capital corresponds (i.e., the individual level, the collective level, or both), the domains captured in the measure, the health outcomes, and the study design applied (i.e., ecologic vs. multilevel).

To date, measures of social capital in epidemiologic studies have largely been based on individual-level indicators of interpersonal trust, norms of reciprocity, and associational memberships, as gathered through surveys administered to representative samples of individuals (primarily adults). To construct measures of social capital at the collective level, the level that has increasingly become the focus of much current research, researchers typically aggregate the individual-level measures by taking their mean value at the collective level of interest—that is, the neighborhood/community, metropolitan/municipal, state/provincial, or country level. Measures of collective social capital that are not derived from individual-level measures of social capital are more difficult to find and validate.

Early epidemiologic studies on social capital primarily focused on broad health outcomes, including life expectancy, all-cause mortality rates, and homicide rates. The more recent literature has explored associations with other outcomes, including general self-rated health and its components (e.g., physical and mental health), health behaviors such as physical activity and medication use, and specific diseases and conditions ranging from sexually transmitted diseases and obesity to behavioral problems in children and food security.

Many of the ecologic studies (i.e., studies in which only data at an area level and not individual level are compared) on social capital have determined significant and moderate associations between social capital and better population health outcomes, whereas "multilevel" analyses (incorporating both area- and individual-level characteristics) have generally found these associations to be more modest. A key disadvantage of ecological studies is the potential for ecological fallacy—that is, relationships between social capital and health at the collective level that may not necessarily translate to the individual level. *Multilevel analyses* can address this issue by modeling individual-level characteristics as well as area-level features simultaneously, and thus allowing for the distinction between population-level *contextual effects* of social capital from *compositional effects* (i.e., health effects due to the sociodemographic and socioeconomic composition of areas), while further taking into account similarities between individuals within the same areas. Multilevel analyses have also revealed the presence of significant interactions between collective and individual levels of social capital and between collective social capital and individual-level characteristics such as race/ethnicity, gender, and income in their health effects. Importantly, these interactions support the notion that the benefits of collective social capital do not necessarily occur uniformly across subgroups of populations.

—Daniel Kim

See also Diffusion of Innovation; Health Disparities; Multilevel Modeling; Social Epidemiology

Further Readings

Kawachi, I., & Berkman, L. F. (2000). Social cohesion, social capital and health. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 174–190). New York: Oxford University Press.

Kawachi, I., Kim, D., Coutts, A., & Subramanian, S. V. (2004). Commentary: Reconciling the three accounts of social capital. *International Journal of Epidemiology*, *33*, 682–690.

Kim, D., Subramanian, S. V., & Kawachi, I. (2006). Bonding versus bridging social capital and their associations with self-rated health: A multilevel analysis of 40 US communities. *Journal of Epidemiology and Community Health*, *60*, 116–122.

Lin, N. (2001). *Social capital: A theory of social structure and action.* New York: Cambridge University Press.

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Reviews of Sociology*, *24*, 1–24.

Putnam, R. D. (2000). *Bowling alone.* New York: Simon & Schuster.

Putnam, R. D., & Goss, K. A. (2002). Introduction. In R. D. Putnam (Ed.), *Democracies in flux* (pp. 3–19). New York: Oxford University Press.

Social-Cognitive Theory

Albert Bandura's social-cognitive theory (SCT) is the result of a revision and expansion of his social learning theory and advocates a model of triadic reciprocal determinism to explain a person's behavior in a particular context. That is, (1) external environment, (2) behavior,

and (3) cognitive/biological/other personal factors all influence each other bidirectionally. Bandura notes that this is a change from previous models that advocate unidirectional causation of behavior being influenced by internal dispositions and/or environmental variables and that SCT does not dictate that the different sources of influence are of equal strength nor does all the influence necessarily take place simultaneously. In summary, (1) internal dispositions (biology, cognition, emotion, etc.) may influence behavior, and behavior may influence internal dispositions; (2) internal dispositions may influence environmental events/reactions, and environmental events may influence internal dispositions; and (3) behavior may influence environmental events, and environmental events may influence behavior. SCT has been applied to study a wide range of public health issues, including medication compliance, alcohol abuse, and immunization behavior, and many public health interventions are based on SCT or selected aspects of it.

SCT recognizes the importance of modeling as an influence on human behavior: It explains how individuals may acquire attitudes from people in the media, as well as from those in their social network. *Direct modeling* refers to observing and possibly imitating people in our networks engaged in certain favorable or unfavorable behaviors, such as watching a father fasten his seat belt as soon as he enters the car. *Symbolic modeling* refers to observing and possibly imitating such behaviors portrayed in the media, such as seeing one's favorite film star smoke cigarettes in movies and magazine photos. Symbolic modeling forms the basis for many public health campaigns in which a celebrity spokesperson endorses or is seen performing a health behavior (such as drinking milk) or condemns a behavior (such as smoking).

Whether modeling leads to changed behavior on the part of the observers depends on many other variables, including individuals' perceptions of the favorable or unfavorable consequences of the behavior, their *outcome expectancies* (i.e., what they think will happen if they perform the behavior), and individuals' perceived ability to carry out the behavior—that is, their *self-efficacy.*

Self-Efficacy in Social-Cognitive Theory

Perhaps the most studied construct in SCT is self-efficacy. A PsycInfo search with self-efficacy as a keyword resulted in 11,530 citations from 1967 to 2006. When "health" as a keyword is combined with the previous search, PsycInfo lists 2,422 citations from 1967 to 2006. Bandura (1986) defines self-efficacy as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances. It is concerned not with the skills one has but with the judgments of what one can do with whatever skills one possesses" (p. 391). In his 1997 book *Self-Efficacy: The Exercise of Control*, he makes it clear that the power to make things happen is very different from the mechanics of how things are made to happen. He emphasizes the importance of *personal agency* (acts done intentionally) and comments that the power to originate behaviors toward a particular goal is the key to personal agency. Beliefs in personal efficacy are what make up the human agency. If people think they have no power to produce certain results, then they will not even try. This concept has clear implications for health behaviors since it may help explain why many people may not even attempt the health promotion behaviors recommended by their health professionals, family members, the media, and so on, or if they do, they do not effectively set short-term subgoals to help them reach their long-term goals.

Interestingly, Bandura suggests that self-efficacy influences the development of competencies as well as the regulation of action. An example of self-efficacy's influence on development might be the neophyte jogger using perceived self-efficacy to dictate the type of situations chosen while learning and/or perfecting his or her skills (e.g., only jogging with others who are just starting or only jogging alone and not with others who may jog faster). Therefore, self-efficacy influences which activities we choose to engage in, our motivational level, and our aspirations.

Bandura differentiates self-efficacy from several other popular constructs in the health and behavioral sciences—*self-esteem*, *locus of control*, and *outcome expectations* (sometimes referred to as *response efficacy*). Perceived self-efficacy is concerned with judgments of personal capacity, whereas self-esteem is concerned with judgments of self-worth; locus of control is concerned with the perception of whether one's actions affect outcomes; and outcome expectations, as noted above, are concerned with the consequences of behaviors that are performed. While self-efficacy asks, "Can I do it?" (e.g., Can I get up early each morning and jog 3 miles?), outcome expectation or response efficacy asks, "If I do it, will it work?" (e.g., If I jog 3 miles each morning, will I lose weight?).

SCT, Self-Efficacy, and Health Behaviors

Salovey, Rothman, and Rodin (1998) remark that self-efficacy may change over time, and in fact, media campaigns, social support groups, and other entities often target our self-efficacy regarding health behaviors to convince us that we can eat more fruits and vegetables, increase our physical activity, lose weight, see our doctor for regular exams, and so on. They suggest that changes in our perceived self-efficacy may be more important than the original baseline levels in motivating and maintaining health behaviors. Salovey et al. further state that, as social scientists are finding with a variety of dispositional constructs, self-efficacy is domain specific rather than a generalized expectation. One may be high in perceived self-efficacy when it comes to adding more fruit to his or her diet but still low on perceived self-efficacy when it comes to increasing his or her physical activity. As has been found in the attitude-behavior consistency research, perceived self-efficacy best predicts behavior when it is measured in the same domain and at the same level of abstraction as the behavior of interest.

There is an impressive amount of research supporting the importance of perceived self-efficacy in predicting behavior in laboratory and field research and in experimental and correlational research. Salovey et al. note that social-cognitive theorists believe that the more skill that a health behavior requires the larger the role played by self-efficacy.

Effects of Perceived Self-Efficacy on Biological Reactions

Bandura notes that social-cognitive theorists view biological reactions to stress (increased blood pressure, etc.) as a result of a low sense of efficacy to exert control over aversive environmental demands. Therefore, if these individuals believe that they can effectively cope with stressors, they are not troubled by them; however, without such a belief system they experience increased distress and impaired performance. Several studies support this view, including research in which phobics' perceived coping efficacy was raised to different levels by modeling or mastery experiences. Those with higher levels of mastery showed less autonomic activation when exposed to the phobia stressor. He also reports on research suggesting the importance of perceived self-efficacy in the management of pain and depression.

Self-Efficacy and Health Promotion

Bandura noted that research by DiClemente, Prochaska, Fairhurst, Velicer, Velasquez, and Rossi (1991) supported the reciprocal influence of behavior and cognition, as they found that perceived self-efficacy increases as people proceed from contemplation to initiation to maintenance of behavior change. He also emphasizes the importance of perceived self-efficacy in being resilient to the disheartening effects of relapses, a frequent problem with health enhancing and compromising behaviors (e.g., persons may start smoking again after abstaining for many weeks, or they may stop exercising after going to the gym on a regular basis for many weeks) and that these people may not even attempt preventive behaviors (e.g., breast self-exams to detect cancerous lumps early) or treatment (e.g., take their medication) since they do not believe they can be successful.

Bandura advocates incorporating self-efficacy-related arguments into health education/persuasive messages as preferable to trying to scare someone via the use of fear appeals into engaging in health-enhancing behavior or ceasing health-compromising behaviors. He observes that often public service campaigns address the efficacy of the method or treatment but ignore promoting personal efficacy. He advocates providing people with the knowledge about how to regulate their health behavior and helping them develop a strong belief in the personal efficacy to turn their concerns into preventive action. Both preexisting perceived self-efficacy as well as altered perceived self-efficacy predict health behavior. Again, part of this process is an emphasis on perseverance of effort and recovering from temporary relapses that often occur with entrenched health-compromising habitual behaviors such as smoking or overeating.

Self-Efficacy in Other Models of Health Behavior

Several issues are common to much research about self-efficacy and health behavior. Self-efficacy is similar to several constructs that have been used in other models of health behavior, including the construct of

perceived behavioral control in Ajzen's theory of planned behavior and the protection motivation theory, which includes both self-efficacy and response efficacy. Unfortunately, these related constructs are often confused with self-efficacy when research is designed and measures developed. This problem is increased because many researchers develop their own measures of self-efficacy, which may actually tap one or more of these other related constructs as well. A meta-analysis by Holden found that more than 75% of the studies they reviewed used their own measures of self-efficacy that had not been validated independently, leaving the question open of what they were actually measuring. Therefore, many studies, whether or not they support the predictive ability of self-efficacy on some health behavior, may not be measuring self-efficacy in the way Bandura originally conceptualized the construct but may instead be assessing one or more of the related constructs.

—Eddie M. Clark

See also Health Behavior; Self-Efficacy

Further Readings

Bandura, A. (1977). Social efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1992). Social cognitive theory. In R. Vasta (Ed.), *Six theories of child development: Revised formulations and current issues* (pp. 1–60). London: Jessica Kingsley.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

DiClemente, C. C., Prochaska, J. O., Fairhurst, S. K., Velicer, W. F., Velasquez, M. M., & Rossi, J. S. (1991). The process of smoking cessation: An analysis of precontemplation, contemplation, and preparation stages of change. *Journal of Consulting and Clinical Psychology*, *59*, 295–304.

Holden, G. (1991). The relationship between self-efficacy appraisals in subsequent health related outcomes: A meta-analysis. *Social Work in Health Care*, *16*, 53–93.

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *Journal of Psychology*, *91*, 93–114.

Rogers, R. W. (1983). Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In J. T. Cacioppo & R. E. Petty (Eds.), *Social psychophysiology* (pp. 153–156). New York: Guilford Press.

Salovey, P., Rothman, A. J., & Rodin, J. (1998). Health behaviors. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 633–683). Boston: McGraw-Hill.

Zanna, M. P., & Fazio, R. H. (1982). The attitude-behavior relation: Moving toward a third generation of research. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior: The Ontario symposium* (pp. 283–301). Hillsdale, NJ: Erlbaum.

Social Epidemiology

Social epidemiology is a field that primarily focuses on the investigation of the social determinants of population distributions of health, disease, and well-being. In contrast to many other fields in epidemiology, social epidemiology places emphasis on the causes of incidence of disease (i.e., the "causes of causes"), which may be very different from the causes of individual cases of disease. This entry describes several fundamental concepts within the field of social epidemiology, including socioeconomic status, social networks, race/ethnicity, residential segregation, social capital, income inequality, and working conditions, and details how these factors have been conceptually and empirically related to health. The entry also briefly discusses some of the core statistical methods that have been applied.

Socioeconomic Status (SES)

Individual-Level SES

The concept of SES is commonly used in the social epidemiologic literature to refer to the material and social resources and prestige that characterize individuals and that can allow individuals to be grouped according to relative socioeconomic position (although it should be noted that the term *socioeconomic status* is a bit of a misnomer, as it appears to emphasize status over material resources). Individual-level SES is typically measured through querying one's income, education, and occupation in surveys. Significant gradients in all-cause and cause-specific mortality for a number of diseases, including coronary heart disease, according to individual SES were established in the classic Whitehall study of British civil servants more than two

decades ago (with higher occupational grades being inversely associated with mortality). Similar relations between individual-level income and mortality have also been found among individuals in other countries, including the United States. Several possible mechanisms have been proposed for the presence of these gradients. These include material pathways (e.g., being able to afford more nutritious foods; having more knowledge about healthy behaviors through higher educational attainment; and having the ability to move into a richer neighborhood, which may provide a more conducive environment for healthy behaviors—as will be discussed further) and psychosocial pathways (e.g., fewer occupational demands relative to the degree of job control—as will also be later described).

Area-Level SES

There are conceptual reasons and empirical evidence to support the notion that the levels of socioeconomic resources and amenities across places in which people live affect the health of individuals, even after taking into account the SES of individuals. For instance, the availability of nutritious foods and green spaces plausibly vary across neighborhoods and, in turn, could influence individuals' diets and physical activity levels. Other characteristics of higher SES neighborhoods that might be relevant to health include the quality of housing and of health services; the presence or lack of "incivilities," such as graffiti and litter; and environmental hazards, such as air pollution and noise. Studies typically operationalize area-level SES by aggregating individual-level SES measures (e.g., by taking the median income of individual survey respondents within a neighborhood). A number of studies have found moderate yet statistically significant associations between neighborhood socioeconomic characteristics and one's risk of dying from cardiovascular disease and from any cause, with 1.1 to 1.8 times higher risks of these outcomes after controlling for one's SES. Other studies have reported significant inverse associations between neighborhood SES with chronic disease risk factors, including smoking, diet, physical activity, and hypertension, and with the incidence of coronary heart disease.

Social Networks

The importance of social networks (i.e., the web of social relationships surrounding an individual and the characteristics of individual ties) to health dates back to the late 19th century, when the sociologist Emile Durkheim showed that individual pathology was a consequence of social dynamics, by tying patterns of suicide to levels of social integration. During the 1950s, the anthropologists Elizabeth Bott and John Barnes developed the concept of "social networks" to understand social ties that extended beyond traditional categories such as kin groups, tribes, or villages. Later, in the mid-1970s, seminal work by the epidemiologists John Cassel and Sidney Cobb linked social resources and support to disease risk and was followed by a number of epidemiologic studies that consistently demonstrated that the lack of social ties predicted death from nearly every cause. These studies include the groundbreaking Alameda County Study in the late 1970s, which prospectively followed nearly 7,000 adults in Alameda County, California, over a 9-year period and found in both men and women a greater than two times greater risk of dying among those who lacked community and social ties, even after taking into account one's age, SES, and lifestyle risk factors such as smoking, physical activity, and obesity.

Psychosocial mechanisms by which social networks may produce its health effects include the provision of social support (e.g., emotional support or instrumental support such as money in times of need), social influence (i.e., interpersonal influence through the proximity of two individuals in a social network), social engagement (i.e., social participation that helps define and reinforce meaningful social roles), person-to-person contact (influencing exposure to infectious disease agents), and access to resources and material goods (e.g., job opportunities and access to health care and education). In turn, these psychosocial pathways have been hypothesized to affect health through health behavioral, psychological, and physiological pathways. For instance, individuals who are socially isolated may adopt unhealthy behaviors such as smoking, may develop negative emotional states such as poor self-esteem or depression, and may acquire prolonged stress responses such as an intermittently raised blood pressure, ultimately leading to hypertension.

Racial/Ethnic Disparities and Racial Residential Segregation

Race/ethnicity refers to the social categorization of individuals into groups, often according to shared

ancestry and cultural characteristics, as well as arbitrary physical features such as skin color. Disparities in health along racial/ethnic lines are well established. For example, in the United States, African Americans have a substantially higher risk of dying from coronary heart disease, stroke, cancer, and diabetes compared with whites. Possible reasons for these disparities relate largely to racism (an ideology used to justify the unequal treatment of racial/ethnic groups considered as inferior by individuals and institutions) and to residential segregation along racial/ethnic lines. Racial discriminatory practices have been shown to affect access to and quality of health care received and educational and employment opportunities and through perceived racial discrimination may contribute to higher levels of stress and unhealthy behaviors.

Residential segregation by race/ethnicity refers to the segregation of racial/ethnic groups along subunits of a residential area. Because of the range of opportunities and resources that different neighborhood socioeconomic environments may provide, as discussed, this segregation into different neighborhood contexts can propagate the socioeconomic deprivation among particular racial/ethnic groups that have historically been disadvantaged (e.g., in the United States, African Americans and Native Americans compared with whites). During the first half of the 20th century, racial discriminatory practices through federal housing policies, bank lending practices, and the real estate industry worked to physically separate blacks from whites in residential areas. More recently, there has been evidence showing the "targeting" and saturation of low-income African American and other minority neighborhoods with fast food restaurants and, prior to tobacco legislation, the targeted advertising of cigarettes within minority neighborhoods. These patterns likely contributed to poorer eating habits and other unhealthy lifestyle behaviors. Several measures of residential segregation by race/ethnicity exist, such as the index of dissimilarity, which captures the percentage of a particular racial/ethnic group that would have to move to evenly distribute the racial/ethnic groups across a residential area.

Other Contextual Determinants of Health

Apart from area-level SES and residential segregation, both *social capital* and *income inequality* have gained prominence in the social epidemiologic and public health literature as possible contextual determinants of population and individual health.

The application of *social capital* to the field of public health arose from prior theoretical and empirical work in the fields of sociology and political science. Definitions of the concept of social capital are varied, ranging from those focusing on the resources within social networks that can be mobilized for purposeful actions to definitions that include both social structures and associated cognitive resources (such as trust and reciprocity) to categorizations such as formal versus informal social capital (e.g., memberships in professional associations vs. outings with friends). A key distinction that cross-cuts these definitions is the level at which social capital exists—that is, the individual level (whereby social capital could take the form of individual-level networks and social support) versus the collective level (e.g., the neighborhood or state level). Several mechanisms by which collective social capital may affect health have been proposed. These include influencing the diffusion of knowledge about health promotion, affecting health-related behaviors through social norms, promoting access to local services and amenities, and psychosocial processes that provide support and mutual respect.

Over the last decade, the number of epidemiologic studies examining the associations between social capital and health outcomes has rapidly grown, primarily as a result of several ecologic studies (i.e., studies in which only data at an area level and not individual level are compared) that found significant inverse associations between social capital and broad health outcomes such as life expectancy, all-cause mortality rates, and homicide rates. More recently, the breadth of this literature has increased to explore associations with one's general self-rated health and its components (e.g., physical and mental health), health behaviors such as physical activity and medication use and specific diseases and conditions ranging from sexually transmitted diseases and obesity to behavioral problems in children and food security, in models additionally controlling for one's SES.

Like social capital, research interest and empirical work on *income inequality* has flourished over the past decade. Income inequality refers to inequality in the distribution of income within populations and has been postulated to have harmful effects on health. The original hypothesis arose from the inability of a country's gross domestic product to account for variations in

average life expectancy among rich nations. Mechanisms have since been put forth, including negative health effects resulting from individuals' feelings of relative deprivation, the erosion of social capital, and underinvestments in public goods such as education and health care, as the interests of the rich diverge from those of the poor. Although a number of measures of income inequality have been constructed (such as the 90/10 ratio, which compares the household income at the 90th percentile with that at the 10th percentile), the most widely applied measure is the Gini coefficient. The Gini coefficient equals half the arithmetic mean of the absolute differences between all pairs of incomes in a population and ranges from 0 (perfect equality) to 1 (perfect inequality).

As with the social capital and health literature, initial studies on income inequality and health were ecologic in design, and a number of these studies identified significant associations between income inequality and life expectancy, all-cause and cause-specific mortality, and self-rated health (in the anticipated directions). More recent investigations have applied a multilevel analytic framework and controlled for individual-level SES. Most of the studies supporting an association between income inequality and health have been in the context of the United States, a country with a comparatively high Gini coefficient among developed nations, whereas findings have generally been null in more egalitarian societies such as Japan and Sweden.

Working Conditions

The psychosocial work environment may also play an important role in determining levels of health among individuals. In support of this relation, one classic theoretical model is the *psychological demand-decision latitude model*, which consists of two dimensions: (1) emotional and psychological demands and (2) decision latitude, which corresponds to the degree of control an employee has over work-related tasks. Based on conceptualized interactions between these dimensions (each dichotomized as high or low), a worker may be assigned to one of four quadrants. In the quadrant of high psychological demands and low decision latitude, *job strain* is said to occur. Under these conditions, Robert Karasek hypothesized that the sympatho-adrenal system of the body is excessively activated while the body's ability to repair tissues is reduced, ultimately leading to illness. Job strain has been shown in some studies to predict the development of hypertension and coronary heart disease in both men and women, controlling for one's SES and other lifestyle factors.

A second classic model of the psychosocial work environment was developed by the sociologist Johannes Siegrist and is referred to as the *effort-reward imbalance model.* This model concerns the degree to which workers are rewarded (such as through financial compensation or improved self-esteem) for their efforts. When a high degree of effort is insufficiently met with the degree of reward, emotional stress and the risks of illnesses are hypothesized to increase. For example, in workplaces that offer disproportionately generous salaries and promotions in relation to employees' efforts, employees are expected to have lower levels of stress and better health status. Studies that have followed workers prospectively have found significant associations between effort-reward imbalance and higher risks of blood pressure and coronary heart disease, as well as reduced levels of physical, psychological, and social functioning, controlling for other factors.

Statistical Methods

As a field within the discipline of epidemiology, studies in social epidemiology often apply common statistical methods such as multiple linear regression and logistic regression. Issues of measurement are customary in social epidemiology due to the social constructs of interest, which for the most part cannot be directly observed (e.g., social capital, which is typically measured through multiple survey items on interpersonal trust, reciprocity, and/or associational memberships). In the case of area-level social factors (such as community social capital or neighborhood SES), measures are typically derived by aggregating individual-level measures. Methods such as factor analysis that are customarily applied in other disciplines (e.g., psychology and sociology) are frequently used to help in validating such measures.

When the research aim is to explore associations between contextual factors (such as neighborhood SES and income inequality) and individual-level health behaviors and outcomes, multilevel models aid in promoting validity. Such models account for individual-level observations within the same spatial area being potentially nonindependent and thereby reduce the likelihood of incorrectly concluding there is a true association when it is in fact due to chance (known as

a Type I error in statistics). In so doing, applying multilevel methods in social epidemiology can allow for the more valid estimation of *contextual effects* of features of the social environment (such as neighborhood social capital), while controlling for *compositional effects* of spatial areas (such as individual-level SES). These methods can be further extended to more validly assess whether contextual effects vary substantially across subgroups of a population.

—Daniel Kim

See also Multilevel Modeling; Social Capital and Health; Socioeconomic Classification

Further Readings

Berkman, L.F., & Glass, T. (2000). Social integration, social networks, social support, and health. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 137–173). New York: Oxford University Press.

Kawachi, I., & Berkman, L. F. (2000). Social cohesion, social capital, and health. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 174–190). New York: Oxford University Press.

Kawachi, I., Kennedy B. P., & Wilkinson, R. G. (Eds.). (1999). *Income inequality and health: A reader.* New York: New Press.

Kawachi, I., Kim, D., Coutts, A., & Subramanian, S. V. (2004). Commentary: Reconciling the three accounts of social capital. *International Journal of Epidemiology*, *33*, 682–690.

Krieger, N. (2001). A glossary for social epidemiology. *Journal of Epidemiology and Community Health*, *55*, 693–700.

Theorell, T. (2000). Working conditions and health. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 95–117). New York: Oxford University Press.

Williams, D. R. (1999). Race, socioeconomic status and health: The added effects of racism and discrimination. *Annals of the New York Academy of Sciences*, *896*, 173–188.

Social Hierarchy and Health

Social organization and population health are inextricably linked. Societies organize their affairs in different ways, and these differences, by means of various pathways, have an effect on the production of health and disease among individuals, as well as between and within communities. Although some of these pathways are not yet fully understood, an ever-mounting body of evidence persuasively supports the contention that the social gradient—the hierarchical organization of society's members along the social ladder, as defined by a number of socioeconomic classifications or representations of social position—is intimately mirrored by a corresponding health gradient. Almost invariably, those who rank lower in the socioeconomic scale have worse health status than those above them in the hierarchy—that is, the higher the social standing, the better the health. Central to the notion of this steadily observed *social gradient in health*—also known as the "status syndrome," the archetype of the relationship between social hierarchy and health—is the generation and persistence of inequalities in health. Current evidence points out the key role of the psychosocial impact of low position in social hierarchy on the generation of both ill health and health inequalities, issues of fundamental concern to both social epidemiology research and practice.

Hierarchy is a prominent ecological aspect of social organization that entails the establishment of a ranking among elements of a group—and hence asymmetrical relationships—based on power, coercion, and access to resources regardless of the needs of others. This hierarchy, which is institutionalized to minimize open conflict, contrasts with social affiliation by friendship, in which reciprocity, mutuality, and solidarity define a social system based on more egalitarian cooperation. Social hierarchy is the human equivalent of the pecking order or dominance hierarchy of nonhuman primates. Even the most egalitarian societies have some hierarchical structure based on distinctions with the result that some people are perceived as having higher social standing than others.

In 1978, a marked social gradient in coronary heart disease was identified along the six occupational classes defined by the British Registrar General's social class scale in the Whitehall study of British civil servants. Since then, even in fairly homogeneous populations, studies have repeatedly found a gradient in health by socioeconomic status: Those at lower socioeconomic positions have worse health status than those above them in the hierarchy. These findings have led researchers to postulate a relationship between position in the social hierarchy and health. Studies of social hierarchies in nonhuman primates have also identified this relationship. Disentangling the relative importance of economic versus pure social hierarchies in humans is, however, challenging due to their degree of overlap.

This profound relationship between social hierarchy and health inequality in human populations has been primarily revealed in studies of income and mortality: The risk of dying follows closely the social gradient defined by the level of income, and poorer societies and poorer population segments within societies have consistently higher mortality rates and lower life expectancy than their less poor counterparts. A solid set of indicators, such as the concentration index and the slope index of inequality, have been used to quantify the degree of inequality in health associated with the social hierarchy defined by a ridit scale (i.e., the succession of relative positions formed with interval midpoints, relative to an identified distribution, of discrete categories with a natural ordering) of income or other variables of socioeconomic status. Early on, the existence of a threshold effect with poverty as well as other egregious measures of material deprivation (such as illiteracy, lack of clean water and sanitation, famine, or even lack of health care access) was demonstrated, above which the association between social hierarchy and health is blurred. Far from denying a relation between hierarchy and health, this evidence suggests that income, and other *absolute* measures of *material* deprivation, may not always be a good proxy of social status and social differentiation and, more important, that nonincome aspects of social rankings operating in specific cultures and communities may overpower single economic measures such as income distribution.

Thus, when material deprivation is severe, a social gradient in mortality could arise from degrees of absolute deprivation. But the effects of social hierarchy in health are not confined to the poor: In rich societies with low levels of material deprivation, the social gradient in health changes the focus from absolute to *relative* deprivation and from material to *psychosocial* deprivation, which relates to a broader approach to social functioning and meeting of human needs. Realizing that social status is a relative—not absolute—concept, scholars have highlighted the significance of relative position for health: It is not what a person has that is important, but what he or she can do with what he or she has. In other words, it is not position in the hierarchy per se that is the culprit of social gradient in health and health inequalities but what position in the hierarchy means for what one can do in a given society. This realization brings the attention to two vital human needs: control over the circumstances in which people live and work and full social participation. The lower individuals are in the social hierarchy, the less likely it is that their fundamental human needs for autonomy and to be integrated into society will be met. This failure, in turn, is a potent cause of ill health in individuals and populations.

A growing body of evidence is being assembled with regard to the paramount importance of autonomy, human freedom to lead a life people have reason to value, and empowerment as determinants of the social gradient in health and socioeconomic inequalities in health. Poor social affiliation and low status carry high population attributable risks. More unequal societies not only suffer more relative deprivation but also tend to have lower rates of trust and of community involvement and social engagement. Interestingly enough, several studies have made the connection between social conditions and biological pathways that plausibly provide the link to violence, cardiovascular conditions, and other diseases. It has been shown that where income inequalities are greater and more people are denied access to the conventional sources of dignity and status in terms of jobs and money, people become increasingly vulnerable to signs of disrespect, shame, and social anxiety, consistently explaining the strong statistical relationship between violence, hierarchy, and inequality.

Likewise, low social position and lack of control are linked to less heart rate variability (i.e., a sign of low sympathetic tone), raised levels of blood cortisol, delayed heart rate recovery after exercise, and low exercise functional capacity (i.e., signs of impaired autonomic activity), all related to activity of the two main biological stress pathways: the sympatho-adreno-medullary axis and the hypothalamic-pituitary-adrenal axis. One plausible mechanism of action of these stress pathways is through an effect on the metabolic syndrome—that is, a cluster of risk factors (including abdominal obesity, atherogenic dyslipidemia, high blood pressure, insulin resistance and prothrombotic and proinflammatory states) that increases the risk of heart diseases and Type II diabetes. Stress at work has been shown to be strongly related to metabolic syndrome, which, in turn, exhibits a clear social gradient, and it is related to those biological stress pathways.

Life contains a series of critical transitions: emotional and material changes in early childhood, moving to successive schools, starting work, leaving home, starting a family, changing jobs, facing retirement, and

so on. Each of these changes can affect the ability of individuals to built and maintain social networks, influence their standing on the social ladder, and also impinge on their health by pushing people onto a more or less advantaged path, putting those who have been disadvantaged in the past at the greatest risk in each subsequent transition. The longer people live in stressful economic and social circumstances, the greater the physiological wear and tear they suffer, the greater the social divide, and the less likely they are to enjoy a healthy life.

Health and quality of social relations in a society seem to vary inversely with how pervasive hierarchy is within the society. The most important psychosocial determinant of population health may be the levels of the various forms of social anxiety in the population, which, in turn, are mainly determined by income distribution, early childhood experiences (including intergenerational nongenetically transmitted behaviors), and social networks. More hierarchical, unequal societies may be more differentiated by social rank into relations of dominance and subordination and less able to enjoy more egalitarian and inclusive relations consistent with higher social capital and less class and racial prejudice. In this analysis, far from being an epiphenomenon, social capital may emerge as an important element in the causation of health and illness in the population. The link between health and social capital (and egalitarianism) is emphasized by the epidemiological findings testifying to the importance of social status and social relations—that is, social cohesion as beneficial to societal health.

Social and economic resources shape the social organization and the health of individuals and communities: Different socioeconomic factors could affect health at different times in the life course, operating at different levels of organization and through different causal pathways. Moreover, socioeconomic factors can interact with other social characteristics, such as racial/ethnic group and gender, to produce different health effects and gradients across groups. The existence of wide—and widening—socioeconomic disparities in health shows how extraordinarily sensitive health remains to socioeconomic circumstances and, by consequence, to social hierarchy.

—Oscar J. Mujica

See also Determinants of Health Model; Health Disparities; Social Capital and Health; Social Epidemiology; Socioeconomic Classification

Further Readings

Bartley, M. (2004). *Health inequality: An introduction to theories, concepts and methods*. Cambridge, UK: Polity Press.

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in U.S. public health research: Concepts, methodologies, and guidelines. *Annual Review of Public Health*, *18*, 341–378.

Marmot, M. (2004). *The status syndrome: How social standing affects our health and longevity*. New York: Henry Holt.

Marmot, M. (2006). Harveian oration: Health in an unequal world. *Lancet*, *368*, 2081–2094.

Sapolski, R. M. (2005). The influence of social hierarchy on primate health. *Science*, *308*, 648–652.

Wilkinson, R. G. (1999). Health, hierarchy, and social anxiety. *Annals of the New York Academy of Sciences*, *896*, 48–63.

Social Marketing

Social marketing is the use of marketing principles and techniques to develop and promote socially beneficial programs, behaviors, and other products. In public health, social marketing has shown great promise as a strategic planning process for developing behavior change interventions and improving service delivery. This entry describes social marketing's distinctive features, steps, and major challenges.

Social Marketing's Distinctive Features

Social marketing is a data-driven strategic planning process that is characterized by its reliance on marketing's conceptual framework to bring about voluntary behavior change. The most distinctive features are a commitment to create satisfying exchanges, the use of the marketing mix to design interventions, segmentation of the target populations, and a data-based consumer orientation.

Satisfying Exchanges

Marketers believe people act largely out of self-interest, searching for ways to optimize the benefits they gain and minimize the costs they pay in their exchanges with others. In commercial transactions, consumers typically exchange money for tangible

products or services. In public health, people more often sacrifice comfort, time, and effort for the value gained from adopting a healthy behavior or participating in a program. Social marketing encourages public health practitioners to offer exchanges that satisfy customers' wants as well as their needs.

The Marketing Mix

Marketing also offers public health professionals a set of conceptual tools called the "4 Ps"—product, price, place, and promotion—for planning program interventions. Also, known as the marketing mix, these concepts are carefully considered from the consumers' points of view and used to develop integrated plans that guide all program activities.

The *product* refers to several critical features of an intervention. The actual product refers to the recommended or desired behavior—for example, a protective behavior being promoted, use of a public health program, or abandonment of a risky behavior. The core product refers to the benefits consumers gain from adopting the product. In some cases, tangible commodities, called augmented products, also are involved. For instance, in a program to decrease eye injuries among citrus pickers, the actual product is the use of safety glasses, reduction of daily irritation is the core product or benefit, and specific brands of safety eye wear that are comfortable to wear in Florida's groves are augmented products.

Price refers to monetary and other costs (e.g., embarrassment, hassle) that are exchanged for product benefits. In the eye safety project just described, intangible costs, such as discomfort and loss of productivity when glasses get dirty, were just as significant as the cash outlay to purchase them. Unless costs for public health products are lowered or made acceptable, even appealing offers may be rejected as unaffordable.

Place has several applications: the locations and times consumers perform the desired behavior, the distribution of augmented products and the point at which consumers obtain them, the actual physical location at which services are offered (attractiveness, comfort, and accessibility), and people and organizations that facilitate the exchange process (e.g., refer people to a program or reinforce behavioral recommendations).

Promotion includes a variety of activities intended to affect behavior change. In public health, an integrated set of activities are usually needed. Professional training, service delivery enhancements, community-based activities, and skill building are often combined with communications (e.g., consumer education, advertising, public relations, special events).

Audience Segmentation

Social marketers know that one intervention doesn't fit everyone's needs, so they identify subgroups in a population that respond differentially to marketing tactics (e.g., core benefits offered, spokespersons, information channels). To optimize resource allocation, marketers subdivide groups based on their current behavior (e.g., sedentary vs. moderately active), readiness to change, reasons people have not adopted the desired behavior, and other factors that affect their response to intervention strategies. They also select one or more segments to receive the greatest priority in planning their interventions.

Data-Based Consumer Orientation

Perhaps the most important element of social marketing is its reliance on consumer research to understand and address the respective audience's values, lifestyle, and preferences to make the key marketing decisions that comprise a marketing plan—that is, segments to give greatest priority, benefits to promise, costs to lower, and product placement and promotion requirements. Time and other resources are devoted to audience analysis; formative research; and pretesting of message concepts, prototype materials, and training approaches. Public health managers who use social marketing are constantly assessing target audience responses to all aspects of an intervention from the broad marketing strategy to specific messages and materials.

Steps in the Social Marketing Process

The social marketing process consists of five steps or tasks.

1. *Audience analysis.* The problem is analyzed to determine what is known about its causes and the audiences affected. Situational factors affecting the project are considered and formative research is conducted to understand the issue from the consumers' viewpoints. Of special interest are consumers' perceptions of product benefits, costs, placement, and potential promotional strategies.

2. *Strategy development.* Research findings are used to make key marketing decisions and develop a blueprint or marketing plan to guide program development. The strategy team determines the audience segments to target, the core product to offer, strategies for lowering costs or making them acceptable, places to offer products, partners to support product adoption, and ways to promote the product to select audience segments.
3. *Program development.* Interventions are developed and message concepts, prototype materials, and training and promotional activities are created and tested.
4. *Program implementation.* Social marketers carefully coordinate an integrated set of promotional activities and rely on the marketing plan to guide program implementation.
5. *Program monitoring and evaluation.* All aspects of program interventions are monitored to identify unforeseen problems that may require midcourse revisions to improve their effectiveness.

Social Marketing Applications

During the last 30 years, social marketing has been used in the United States and elsewhere to develop programs to promote family planning, breastfeeding, increased fruit and vegetable consumption, physical activity, immunization, environmental protection, a variety of safety practices, and other healthy behaviors. It also has been used to (re)design and promote programs, such as the food stamp program, the Special Supplemental Nutrition Program for Women, Infants, and Children, Medicaid, and others, with significant success. While its use has increased dramatically, social marketing has yet to realize its potential in public health because of a lack of training and widespread misunderstandings. The most challenging problems limiting its application include the following:

- An overreliance on communications, especially mass media. Many public health professionals still equate social marketing with social advertising and misuse the label to describe campaigns that rely exclusively on mass media messages to bring about change rather than a careful integration of the entire marketing mix.
- A reluctance to invest time and money on consumer research. While marketing research does not always have to be expensive or complex, it is essential to understand how consumers view the product benefits, costs, placement, and promotion.
- An overreliance on focus groups. Focus groups have many advantages in marketing research, but they can also be misleading if not conducted and analyzed carefully. Like all qualitative data collection methods, the results cannot be verified statistically or used to estimate the prevalence of views within a target population.
- An overreliance on demographic variables to segment audiences and reticence to select segments to target. Many public health professionals try to reach everyone with the same intervention, or if they segment, they do so exclusively with demographic variables.
- Failure to rigorously evaluate social marketing interventions.

Conclusions

Social marketing is widely accepted as a method for promoting healthy behaviors, programs, or policies. It is distinguished from other planning approaches by a commitment to offer satisfying exchanges, use of marketing's conceptual framework, segmentation and careful selection of target audiences, and close attention to consumers' aspirations, preferences, and needs. Although social marketing often uses mass media to communicate with its target audiences, it should not be confused with health communication, social advertising, or educational approaches, which simply create awareness of a behavior's health benefits and/or attempt to persuade people to change through motivational messages. Social marketers use research results to identify the product benefits most attractive to target audience segments, determine what costs will be acceptable and those that must be lowered, identify the best places to offer products, and communicate with consumers and design the best mix of promotional tactics to elicit behavior change.

—Carol Anne Bryant

See also Diffusion of Innovations; Health Behavior; Targeting and Tailoring

Further Readings

Donovan, R., & Henley, N. (2003). *Social marketing principles and practice.* Melbourne, Australia: IP Communications.

Grier, S., & Bryant, C. A. (2005). Social marketing in public health. *Annual Reviews in Public Health*, *26*, 6.1–6.21.

Kotler, P., Roberto, N., & Lee, N. (2002). *Social marketing: Strategies for changing public behavior.* Thousand Oaks, CA: Sage.

Kotler, P., & Zaltman, G. (1971). Social marketing: An approach to planned social change. *Journal of Marketing*, *35*, 3–12.

McKenzie-Mohr, D., & Smith, W. (1999). *Fostering sustainable behavior: An introduction to community-based social marketing.* British Columbia, Canada: New Society.

Socioeconomic Classification

Socioeconomic classification refers, in broad terms, to the arrangement, categorization, or assignment of individuals of a population (and, by extension, other population-based elements such as families, households, neighborhoods, geopolitical units, etc.) to predesignated classes, orders, subgroups, or continuous scale or gradient on the basis of perceived common social, societal, and/or economic attributes, characteristics, conditions, relations, or affinities. The goal of any socioeconomic classification is to provide a valid, relevant, and meaningful organization of the population into separate, discrete social classes or, conversely, all along a hierarchical continuum of socioeconomic position. Ample evidence supports the assertion that social and economic resources shape the health of individuals and communities; indeed, socioeconomic status is regarded as a fundamental macrodeterminant of population health. Socioeconomic classification is at the core of these considerations, and it can, therefore, critically affect epidemiological and public health research and practice, with direct implications for public health policy.

Social sciences, as well as social epidemiology, consistently recognize that behind any socioeconomic classification there is a multidimensional construct comprising diverse social and economic factors. It is increasingly acknowledged that a fundamental distinction between "social class," "social status," and measures of material living standards is needed to clarify definitions, measures, and interpretations associated with a given socioeconomic classification. This would include distinguishing between *income*, *assets*, and *wealth* (i.e., those based on individual and household ownership of goods), terms frequently used loosely and interchangeably despite their different theoretical foundations.

Social classes—hierarchical distinctions between individuals or groups in societies or cultures—are social groups arising from interdependent economic relationships among people. These relationships are governed by the social structure as expressed in the customs, values, and expectations concerning property distribution, ownership, and labor and their connections to production, distribution, and consumption of goods, services, and information. Hence, social classes are essentially shaped by the relationships and conditions of employment of people in the society and not by the characteristics of individuals. These class relationships are not symmetrical but include the ability of those with access to resources such as capital to economically exploit those who do not have access to those resources.

Unlike social class, *social status* involves the idea of a hierarchy or ranking based on the prestige, honor, and reputation accorded to persons in a society. Societal sources for attribution of status, that is, a relative position in the social ladder, are diverse but chiefly concern access to power, knowledge, and economic resources.

Both social class and social status can be regarded as representations of social position. Yet a growing body of knowledge from research on health inequalities indicates a need to consider a more comprehensive socioeconomic classification that can include class, status, and material asset measures, collectively referred to as *socioeconomic position*. This term is increasingly being used in epidemiology as a generic term that refers to the social and economic factors that influence which positions individuals or groups will hold within the structure of a society. Socioeconomic position is one dimension of social stratification and, as such, is an important mechanism through which societal resources and goods are distributed to and accumulated over time by different groups in the population.

From an analytical standpoint, social class involves categorical (usually nominal as opposed to ordinal), discontinuous variables. Social status, on the other hand, is considered as a continuous variable, although for the purposes of analysis it may be divided into categories using cutpoints or other divisions dependent on the data structure rather than a priori reference points. Characteristics of socioeconomic position pertaining to material resources (such as income, wealth, education attainment, and, by extension, poverty, deprivation,

etc.) can be modeled as ordinal or interval categorical variables. Another important implication for data analysis is that, unlike social class or status, socioeconomic position can be measured meaningfully at different levels of organization (such as individual, household, and neighborhood levels), as well as at different points in the lifespan (such as infancy, adolescence, adulthood).

The array of socioeconomic classification schemes and indicators of socioeconomic position includes both individual-level and area-level measures. Among others, there are those based on education; income, poverty, and material and social deprivation; occupation, working life, and exclusion from labor force; house tenure, housing conditions, and household amenities; social class position; proxy indicators; composite measures; and indices of deprivation. Among the best known socioeconomic classifications is the British Registrar General's Social Class (RGSC) scale, used since 1913. This scale, which is based on the occupation of the head of the household, defines six social classes: I, professional; II, managerial; III-NM, skilled nonmanual; III-M, skilled manual; IV, semiskilled manual; and V, unskilled manual. The RGSC scale is said to be based on either general standing in the community or occupational skill, and its categories broadly reflect social prestige, education level, and household income. Despite much criticism for its obvious class and gender biases, as well as for its exclusion of individuals outside the formal paid labor force, this schema has proven to be powerfully predictive of inequalities in morbidity and mortality. Wright class schema or socioeconomic classification is another well-known typology that—based on the contention that the essence of class distinctions can be seen in the tensions of a middle class simultaneously exploiting and being exploited (in terms of ownership, control, and possession of capital, organization, and credential assets)—ultimately distinguishes between four core class categories: wage laborers, petty bourgeois, small employers, and capitalists. Yet other standard socioeconomic classifications include the Erikson and Goldthorpe class schema, the Nam-Powers' occupational status score, the Duncan socioeconomic index, the Cambridge social interaction and stratification scale, the Hollingshead index of social position, the Warner index of status characteristics, and the Townsend deprivation index. These measures and others are discussed in detail in the Further Readings list provided at the end of this entry.

There are many criteria, schemes, and indicators to generate a classification of socioeconomic position; no single measure can be regarded as suitable for all purposes or settings. Ideally, this choice should be informed by consideration of the specific research question and the proposed mechanism linking socioeconomic position to the health outcome. In practice, however, the measures used tend to be driven by what is available or has been previously collected. To reflect on the potential of a given indicator of socioeconomic position to help understand, as opposed merely to describe, health, inequality should be taken as an overarching principle when choosing a socioeconomic classification measure.

—Oscar J. Mujica

See also Determinants of Health Model; Health Disparities; Social Capital and Health; Social Epidemiology; Social Hierarchy and Health

Further Readings

Bartley, M. (2004). Measuring socio-economic position. In M. Bartley (Ed.), *Health inequality: An introduction to theories, concepts and methods* (pp. 22–34). Cambridge, UK: Polity Press.

Galobardes, B., Shaw, M., Lawlor, D. A., Davey Smith, G., & Lynch, J. W. (2006). Indicators of socioeconomic position. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in social epidemiology* (pp. 47–85). San Francisco: Wiley.

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in U.S. public health research: Concepts, methodologies, and guidelines. *Annual Review of Public Health*, *18*, 341–378.

Liberatos, P., Link, B. G., & Kelsey, J. L. (1988). The measurement of social class in epidemiology. *Epidemiological Reviews*, *10*, 87–121.

Lynch, J., & Kaplan, G. (2000). Socioeconomic position. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 13–35). Oxford, UK: Oxford University Press.

Society for Epidemiologic Research

The mission of the Society for Epidemiologic Research (SER), established in 1968, is to create a forum for sharing the most up-to-date information in epidemiologic research and to keep epidemiologists at the vanguard of scientific developments.

SER is a membership organization governed by a four-member executive committee (president, president-elect, past president, and secretary-treasurer) and a five-member board including one student representative; the executive committee and the board members are elected by the SER membership. SER holds an annual scientific meeting and is one of the sponsors, along with the American College of Epidemiology and the Epidemiology Section of the American Public Health Association, of the North American Congress of Epidemiology, which is held every 5 years (most recently in 2006). In addition, SER sponsors publication of the professional journals *American Journal of Epidemiology* and *Epidemiologic Reviews* and publishes a semiannual newsletter concerning SER activities (available on the SER Web site). The SER office is located in Clearfield, Utah, USA.

The annual meeting of the SER is held in the United States or Canada. The 40th annual meeting was held in Boston, Massachusetts, in June 2007. The 41st annual meeting will be held in Chicago, Illinois, in June 2008. The meeting includes presentation of scientific papers and posters by SER members, roundtable discussions, and instructional workshops. The winner of the Abraham Lilienfeld Student Prize, which is awarded annually for the best paper describing research done as a student in an advanced degree program with a concentration in epidemiology, is invited to present his or her research during the plenary session of the annual meeting.

Epidemiologic Reviews is published once a year by Oxford University Press. It publishes review articles focused on a particular theme, which is changed annually: The 2006 issue focused on vaccines and public health, and the 2007 issue focuses on the epidemiology of obesity. The *American Journal of Epidemiology* is published 24 times a year by Oxford University Press and publishes original research articles, reviews, methodology articles, editorials, and letters to the editor. In 2005, the impact factor for *Epidemiologic Reviews* was 4.722, and the impact factor for the *American Journal of Epidemiology* was 5.068, ranking them second and fourth among 99 journals in public, environmental, and occupational health.

—Sarah Boslaugh

See also American College of Epidemiology; American Public Health Association; Journals, Epidemiological

Further Readings

Web Sites

American Journal of Epidemiology: http://aje.oxfordjournals.org.
Epidemiologic Reviews: http://epirev.oxfordjournals.org.
Society for Epidemiologic Research: http://www.epiresearch.org.

Specificity

See Sensitivity and Specificity

Spirituality and Health

Interest in the role of spirituality in health outcomes has increased in recent years in both scientific and lay circles, as evidenced by increases in published articles and in funding for research in this area. This entry provides a brief review of the research on spirituality and health, methodological challenges, and future research directions. This area of research not only has theoretical value but may also lead to applied knowledge relevant to health education and promotion.

Discussions of research in this area should generally begin with a brief definition of concepts, including what is meant by the term *spirituality*. These discussions typically also involve the concept of religion, as the two concepts are often used interchangeably in this literature even though they refer to distinct—yet potentially overlapping—constructs. Although there has been debate over the usage of these terms, *spirituality* is often used to refer to people's experience of what gives them meaning in life, which may include things such as nature or a higher power. The term *religion*, on the other hand, is often used to refer to an organized system of worship involving doctrine, beliefs, and a higher power often but not always referred to as "God." Research in spirituality and health is not as developed as in religion and health, largely due to the difficulty of assessing the construct of spirituality. There has been widespread disagreement on what is meant by spirituality. Although it poses its own set of conceptual challenges, religion is easier to define; as a result, the majority of research on "spirituality and health" has actually

focused on a concept more related to religion. Thus, for the purpose of this entry, the term *spirituality/religion* is used, with the recognition that these terms are not interchangeable.

Research Methodology

Early Work

Research in the area of spirituality/religion and health began with large population-based data sets examining the association between single indicators of religion (e.g., church attendance, religious affiliation) and health outcomes such as mortality. Because positive relationships were often found even with these crude indicators, interest in this area increased.

Emergence of Multidimensional Assessment

Researchers began to recognize that the single-item indicators of religion needed much improvement relative to the way that other psychological constructs were being assessed. Multiple-item scales began to be developed, and later came the recognition that religion and spirituality are indeed multidimensional constructs having several dimensions. These dimensions were reflected in instruments assessing dimensions such as public and private religiosity and religious beliefs and behaviors.

Other Methodological Challenges

Measurement was not the only methodological challenge to be overcome in spirituality/religion and health research. Even when positive associations were found between spirituality/religion and a health outcome (and this was not always the case), there were questions as to whether there was another variable, such as health status or age, that was confounding the relationship. Researchers in this area must be aware of the potential for confounding; for example, they must ask whether those who attend church are more likely to experience positive health outcomes for some reason other than their spirituality/religion. These variables must then be assessed and controlled for in research studies, as is being done in some research. But even when the confounding factors are controlled for, a cause-effect relationship cannot be demonstrated without longitudinal studies, and this is difficult to accomplish. Finally, most of the studies in the United States have focused on Christian populations. While this approach was taken to enhance generalizability by studying relatively large populations, it is inherently limited since it applies only to these populations. Much less is known about those of other faiths such as Buddhists and Muslims.

The Relationship Between Spirituality/Religion and Health

Is there a relationship between spirituality/religion and health? Although this is a complex question beyond the scope of this entry, several large-scale reviews of the literature in this area have concluded that the weight of the evidence for the relationship between spirituality/religion and health is generally, although not always, positive. Some studies have found negative relationships and others found no such relationship, but most do find a modest positive association with outcomes such as health-related behaviors, conditions, and general mortality.

After many studies attempting to answer the question of *whether* there is an association between spirituality/religion and health, the research in this area has generally moved on to ask *why* such an association might exist, for whom, and under what conditions. To address the ''why'' question, several excellent theoretical articles have proposed mechanisms for the relationship between health and spirituality/religion. These mechanisms include the hypotheses that spiritual/religious individuals experience health benefits because they have more social support, experience more positive affect, have a healthier lifestyle, engage in healthier behaviors, experience more social pressure to avoid unhealthy behaviors (e.g., not smoking in front of fellow church members), or cope better with stress than do less spiritual/religious individuals. Although there are little actual data to support many of these mechanisms to date, the future of the research points to studies that can provide data to support such mediational relationships and build the much needed theory in the area. It is also important, when examining an area such as spirituality/religion and health mediators, to begin this work with qualitative studies before making assumptions about what is certainly a very complex set of relationships. This will help ensure that the quantitative studies are asking the right questions.

Another question is for whom spirituality/religion might have a health benefit. Again, studies have begun to address this question but many more are needed to

identify population subgroups that are more or less apt to experience the connection. For example, it may be that particular racial/ethnic groups, those of different age groups, of different denominational affiliations, or of different socioeconomic strata differ in the strength of the spirituality/religion-health connection.

Finally, there is the question of under what conditions spirituality/religion might have a health benefit. It is possible that there may be a positive association for some health outcomes and not others. For example, individuals may view their spiritual/religious beliefs as a basis for avoiding behaviors such as tobacco, drug, or alcohol use but not for adopting health-promoting behaviors such as a healthy diet and physical activity. Additionally, it may be that the association holds only for individuals who have adequate social support systems in place or in other conditions of which researchers are unaware. Again, this is where qualitative methods such as in-depth interviews and focus groups can shed some light on these complex phenomena.

Why Study the Relationship Between Spirituality/Religion and Health?

Besides being a challenging area in terms of measurement and controlling for confounding variables, and an intriguing area in terms of the complexity of the spirituality/religion-health mediators, there is potential applied value in this area of research. For example, when researchers learn more about the nature of the association between spirituality/religion and health, this information can be used to improve the effectiveness of the many church-based health promotion programs that are now being used to better the health of these communities. Additionally, many patients are asking that their spiritual needs be addressed within the context of clinical care. This research may be able to inform how these situations are handled.

Future Research

There are many potential directions for research on the relationship between spirituality/religion and health, including answering the aforementioned questions (why, for whom, and under what conditions this relationship exists). In addition, another promising area of research that has applied value deals with ways to improve effectiveness of church-based health promotion programs by using a spiritually based approach to health education. Finally, it is important to bring together a multidisciplinary group of scholars to study this area—including theologians, who have not traditionally been included in past research. This is a dynamic area of research that has made significant advances in recent years and still has much opportunity for growth and discovery.

—Cheryl L. Holt

See also Cultural Sensitivity; Health Communication; Health Disparities; Locus of Control; Measurement

Further Readings

George, L. K., Larson, D. B., Koenig, H. G., & McCullough, M. E. (2000). Spirituality and health: What we know, what we need to know. *Journal of Social and Clinical Psychology*, *19*, 102–116.

Hill, P. C., & Hood, R. W. (1999). *Measures of religiosity.* Birmingham, AL: Religious Education Press.

Holt, C. L., Kyles, A., Wiehagen, T., & Casey, C. M. (2003). Development of a spiritually-based breast cancer educational booklet for African American women. *Cancer Control*, *10*, 37–44.

Holt, C. L., Lewellyn, L. A., & Rathweg, M. J. (2005). Exploring religion-health mechanisms among African American parishioners. *Journal of Health Psychology*, *10*, 511–527.

Koenig, H. G., McCullough, M. E., & Larson, D. B. (2001). *Handbook of religion and health.* New York: Oxford University Press.

Levin, J. S. (2001). *God, faith, and health: Exploring the spirituality-healing connection.* Hoboken, NJ: Wiley.

Levin, J. S., & Vanderpool, H. Y. (1989). Is religion therapeutically significant for hypertension? *Social Science and Medicine*, *29*, 69–78.

Spreadsheet

Spreadsheets are among the most common types of computer software used by people working in epidemiology and public health. When desktop computers were introduced in the late 1970s, the first "killer app" ("killer application," i.e., the software that everyone wants to have) was the spreadsheet. Visicalc, the first major spreadsheet application, was so useful that it justified the purchase of a computer. For people working in fields such as statistical analysis, scientific research, economics, and finance, the ability

to easily manipulate large amounts of numerical data presented an immense advantage over manual methods. The consequent time and cost savings easily paid back the investment in a desktop (or "personal") computer. In fact, spreadsheet software quickly became so popular that it helped establish the notion of the personal computer—a computer used primarily by a single individual and small enough to sit on a desktop, in contradistinction to the mainframe computers that were far more common at the time.

When the IBM PC was introduced in the early 1980s, Lotus 1-2-3 became its killer app. Lotus became the accepted standard for over a decade. Its success was coincidental with the runaway success of the PC. By the 1990s, with the introduction of the visual interface of Microsoft Windows and Apple's Macintosh, Microsoft Excel overtook Lotus 1-2-3 as the market leader and remains the standard to this day. Although there are other choices in spreadsheet software, Excel has a market share of more than 90%. Because Excel is bundled with the dominant word-processing program Microsoft Word in the ubiquitous Microsoft Office package and because many other programs can use Excel data files, it has become the accepted standard.

Spreadsheet software (the name is derived from the spreadsheet used by accountants to record financial information) is a computer program that presents a rectangular matrix of rows and columns to display data (see Figure 1). Each cell can contain numerical or textual data. Columns are defined by letters and rows by numbers. Cells are referenced as the intersection of those two criteria, A1 or D37, for example. In this figure, each row contains the information for one case, in this instance for one patient. Each column represents a variable (gender, date of birth, etc.) for that patient. In database terminology, each row is a record and each column is a field in that record.

Spreadsheets are most commonly used in epidemiology and public health for three purposes:

1. to create, store, and share electronic data files;
2. to perform basic calculations on data; and
3. to visually examine data and create reports, graphs, and charts based on the data in a spreadsheet.

The most common use of spreadsheets in epidemiology is to enter, store, and share electronic data files. Spreadsheets offer several advantages in data entry. They allow data to be copied and pasted, rearranged, and reused. Spreadsheets also have time-saving features such as the fill function, which will copy formulas and number series to other cells. Features such as sorting and filtering make it easy to look at data in a spreadsheet, and most statistical applications programs can easily import data stored in a spreadsheet. Numbers and text can be entered and displayed in a variety of formats, and columns and rows can be resized vertically and horizontally to accommodate varying lengths of entries.

Figure 1 shows an excerpt of a spreadsheet storing information about a medical study, in standard rectangular file format (in cells A5:D15: Rows 1 to 4 are used for descriptive information and would not be used in statistical analysis). Each row represents information about a single patient, while each column represents a single variable. Row 5 contains labels that can be preserved as variable names when we import these data into a statistical package. Therefore, the first patient (with ID #1) is a female born on May 4, 1956, and whose first office visit was on August 1, 2005; the second patient is a male born on March 15, 1953, and whose first office visit was on September 3, 2005.

Spreadsheets can also be used to process data. They include many built-in functions that automate tasks such as computing the sum of a column of numbers or the number of days between two dates. This allows the user to perform simple data manipulations without using dedicated statistics software such as SAS. Using these functions, epidemiologists can also

	A	B	C	D	E
1	**Dr. Lionel Schmerz**				
2					
3	Patient Records				
4					
5	**Patient _ID**	**DOB**	**Gender**	**1st Visit**	
6	1	04-May-56	F	01-Aug-05	
7	2	15-Mar-53	M	03-Sep-05	
8	3	27-Mar-53	M	10-Oct-04	
9	4	17-Jun-90	F	08-May-05	
10	5	20-Dec-63	F	08-Jan-05	
11	6	09-Jun-42	M	09-Jul-04	
12	7	08-Apr-62	F	07-Sep-05	
13	8	17-Mar-59	M	28-Jan-05	
14	9	08-May-62	F	19-Aug-05	
15	10	05-Nov-68	F	08-Dec-05	

Figure 1 A Typical Spreadsheet Where Each Row Contains Data About a Patient

test the results of "what if?" scenarios. For example, if one assumes that 2% of the population per year will become infected with a disease, how many cases will one have in 10 years? What if one assumes 3% or 4%? Using a spreadsheet, one can see the results of these different scenarios immediately.

Spreadsheets provide many options for displaying data: It can be sorted, rows and columns can be hidden, data can be filtered so that only particular cases are displayed, and so on. Database designers often use a spreadsheet to analyze the structure of a data set before incorporating it into another system. Modern spreadsheet software also offers the capability to create graphic representations directly from spreadsheet data, including pie charts, bar graphs, and scatterplots. Charts created from spreadsheets are instantly updated when the underlying data are changed. The resulting charts can be saved in a variety of formats or pasted into Word or PowerPoint files.

As spreadsheet software evolved, the features took on many of the capabilities of relational database software. Using the relational model, data in other spreadsheets can be incorporated into a single file. Data can also be searched, sorted, and extracted by criteria. For example, instead of just displaying the sum of a column of numbers, the software can show subtotals by specified categories of the data, even from other files. Specified data can also be extracted into another file and reused.

—Daniel Peck

See also Relational Database

Further Readings

Frye, C. (2004). *Microsoft Office Excel 2003 step by step*. Redmond, WA: Microsoft Press.

Keller, G. (2000). *Applied statistics with Microsoft Excel*. Pacific Grove, CA: Duxbury Press.

Langer, M. (2004). *Excel 2003 for Windows: Visual QuickStart guide*. Berkeley, CA: Peachpit Press.

Statistical and Clinical Significance

See Hypothesis Testing

Stem-and-Leaf Plot

The stem-and-leaf plot was developed by John Tukey and is used for continuous data during exploratory data analysis. It gives a detailed description of the distribution of the data and gives insight into the nature of the data. It is more informative than a simple tally of the numbers or a histogram because it retains individual data points. Using the information in a stem-and-leaf plot, the mean, median, mode, range, and percentiles can all be determined. In addition, when the stem-and-leaf plot is turned on its side, one is looking at a histogram of the data. From this, one can get an idea of how the data are distributed. For example, whether the data appear to be described by a normal curve or whether they are positively or negatively skewed. It also can point to unusual observations in the data that may be real or a result of reporting errors or data entry errors.

To make a stem-and-leaf plot by hand, the data should be ordered and made into categories or groups. If no logical categories exist for the data, a rough guide for the number of stems to use in the plot is two times the square root of the number of data points. When using a statistical program to create a stem-and-leaf plot, the number of categories is determined by the software. The leaf is generally the last digit of the number, and the stem includes the digits that come before the leaf. For example, if the data are whole numbers ranging from 20 to 75, the categories may be from 20 to 24, from 25 to 29, from 30 to 34, and so on. The stem is the digit in the tens position, valued from 2 to 7, and the leaf is the digit in the ones position, valued from 0 to 9. It is useful, especially when there is a place filler as in the example shown here, to add a key so that it is clear what number one's stem and leaf is portraying. See Table 1 for an illustration of a stem-and-leaf plot. By simply looking at this plot, it can be seen that the mode is 60 (the most frequent value) and the median, boldface in the table, is 59 (21 of the values fall above this number, and 21 of the values fall below this number).

The stem-and-leaf plot can also be used to compare data sets by using a side-by-side stem-and-leaf plot. In this case, the same stems are used for both data sets. The leaves of one data set will be on the right, and the leaves of the other will be on the left. When the leaves are side by side like this, the distributions, data ranges, and where the data points fall can be compared.

The stem-and-leaf plot becomes more difficult to create by hand as the amount of data increases. In addition,

Table 1 Height in Inches in U.S. Children Aged 0 to 17 Years

Stem	*Leaf*
2 *	0
2.	99
3 *	0
3.	556688
4 *	1224
4.	88889
5 *	34
5.	99
6 *	0000001234
6.	55677789
7 *	22

Key: 2 * = 20 to 24, 2. = 25 to 29, and so on.

Source: Data from a random selection of 43 points from Centers for Disease Control and Prevention, National Center for Health Statistics, State and Local Area Integrated Telephone Survey, National Survey of Children's Health, 2003.

with large amounts of data, the benefits of being able to see individual data values decrease. This is the case because it becomes increasingly difficult to determine summary measures, such as the median, which is part of the value of using a stem-and-leaf plot. A histogram or a box-and-whisker plot may be a better option to visually summarize the data when the data set is large.

—Rebecca Harrington

See also Box-and-Whisker Plot; Histogram; Measures of Central Tendency; Percentiles; Tukey, John

Further Readings

Tukey, J. (1977). Scratching down numbers (stem-and-leaf). In J. Tukey (Ed.), *Exploratory data analysis* (pp. 1–26). Reading, MA: Addison-Wesley.

Stratified Methods

Confounding is a major consideration in etiological investigation because it can result in biased estimation of exposure effects. Control of confounding in data analysis is achieved by stratified analysis or by multivariable analysis. (Control of confounding in research design stage is achieved by matching for observational studies and by randomization for experimental studies.) Stratified analysis is accomplished by stratifying the confounding variable into homogeneous categories and evaluating the association within these strata. Multivariable analysis, on the other hand, involves the use of a regression model and allows the researcher to control for all confounders at the same time while looking at the contribution of each risk factor to the outcome variable. Stratified analysis is a necessary preliminary step to performing regression modeling to control for confounding. Unlike regression models, stratified analysis requires few assumptions.

Here is a simple example of how the stratified method works. In comparing mortality statistics for Mexico and the United States in the 1990s, we observe that Mexico's crude death rate is lower than the crude death rate in the United States. Yet Mexico's age-specific death rates are higher than those of the United States for every age categories. The different age distributions of the two populations explain the direction and magnitude of the difference in the crude death rates between the two countries. The crude death rate may be expressed as $\sum_i m_i w_i$, which is a weighted average of the age-specific death rates m_i with age distribution w_i as weights. The population of Mexico is younger. Mexico has relatively more people in the younger age categories and less people in the older age categories than the United States—w_i differs as a function of i between the two countries. There is a strong positive association between age and mortality—m_i is an increasing function of i for both countries. Thus, the existence of the confounding variable age i leads to a lower sum of products of m_i and w_i, the crude death rate for Mexico in the unstratified analysis, whereas a stratified analysis with confounding variable age as the stratification variable provides the true picture—Mexico has higher death rates at every age. Consequently, comparison of directly standardized rates of the two countries will show higher mortality for Mexico.

Generally, epidemiologists consider stratified methods for controlling for confounding to include the following steps:

1. Perform an unstratified analysis by calculating the crude measure of association ignoring the confounding variable (depending on the study design, the measure of association could be risk

difference, rate difference, risk ratio, rate ratio, or odds ratio).

2. Stratify by the confounding variable.
3. Calculate the adjusted overall measure of association.
4. Compare the crude measure with the adjusted measure.

If the crude estimate differs from the adjusted estimate by 10% or more, there is confounding, and the adjusted estimate should then be calculated by stratifying the confounder. If the estimates differ by less than 10%, there is no confounding. If there is confounding, formal significance testing and the calculation of 95% confidence interval may then be carried out to determine the significance of the association between the risk factor and the outcome variable for the different strata.

Analysis of *n*-Way Contingency Tables

Stratified analysis is more rigorously performed using categorical data analysis. To do so, we need to categorize all variables to construct *n*-way contingency tables with *n* equal to or greater than 3. Analysis of such a contingency table is done by computing statistics for measures and tests of association between two variables (usually a risk factor and an outcome variable) for each stratum defined by the third variable (usually the confounding variable) as a stratification variable, as well as computing the adjusted overall measures and tests.

Suppose we are interested in the association between the two variables *y* and *z* but we know that a third variable, sex, is a potential confounder. For an unstratified analysis, we would ignore sex and compute an asymptotic chi-square or an exact test statistic to test for the significance of association between *y* and *z* (this may be done by using SAS procedure frequency with tables statement tables *y* * *z*/chisq;). To account for sex as a confounder, we need to perform a stratified analysis adjusting for the stratification variable sex. This is done by analyzing a three-way contingency table with sex defining the strata. We first compute the chi-square test of association for each stratum of sex and then pool the strata to produce the Cochran-Mantel-Haenszel test statistic to conclude on whether or not rows (*y*) and columns (*z*) are associated after controlling for the stratification variable sex (the tables statement for SAS procedure frequency is now changed to tables sex * *y* * *z*/chisq cmh;). Finally, the Mantel-Haenszel estimate provides the adjusted overall measure of association.

Suppose in the unstratified analysis we found no significant association between *y* and *z*. However, in the stratified analysis we found a significant association between *y* and *z* among males but not among females and the Cochran-Mantel-Haenszel statistic shows that there is a significant association between *y* and *z*. Thus, when we adjust for the effect of sex in these data, there is an association between *y* and *z* and the source of association is the male sex. But, if sex is ignored, no association is found.

If stratified analysis has ruled out confounding as a possible explanation for results and association is found to be not statistically significant, there are two possibilities. Either this is a true finding and there is actually no association between the suspected risk factor and the outcome, or the study did not have the power to show the difference even if it exists in the population because of insufficient sample size. In the latter case, another study may need to be conducted to exclude the possibility of confounding.

—John J. Hsieh

See also Confounding; Direct Standardization; Life Tables; Matching; Regression

Further Readings

Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic research: Principles and quantitative methods.* New York: Van Nostrand Reinhold.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology.* Philadelphia: Lippincott-Raven.

Silva, I. (1999). *Cancer epidemiology: Principles and methods.* Lyon, France: IARC.

Stress

Stress is one of the most talked about psychosocial constructs in popular discourse. We invoke the language of stress when we want sympathy, to convey that we feel inundated by demands, responsibility, or worry. The harried young mother in a store, a student at exam time, and the busy corporate executive are all familiar images of the stressed individual. Less prominent in the popular imagination is the stress of the

impoverished, the unemployed, those facing discrimination, and outcasts at the margins of society. Stress is central to the study of health disparities because the disadvantaged members of society bear it in disproportion.

In its epidemiological sense, stress is a way to characterize those aspects of experiencing the social and physical environment that influence the well-being of individuals. A variety of definitions have been put forth, but a prevailing theme is that stress results either from socioenvironmental demands that strain the adaptive capacity of the individual or from the absence of means for the individual to obtain desired ends. Stress therefore is not strictly an attribute of the environment but arises from discrepancies between social conditions and characteristics of the individual. Similar to stress in engineering, psychosocial stress can be thought of as a force on a resisting body that flexes within, but may exceed, a normal range. Social stress research differs from engineering in that it treats the capacity to resist as a separate construct, that of coping. Epidemiologists also distinguish stressors from distress: Stressors refer to the environmental stimulus, while distress is the psychological or behavioral response to the stressor. This entry will describe the origin and development of stress concepts, the continuum of stress, stress as a process, and social patterns of stress exposure.

Origin and Development of Stress Concepts

Early-20th-century investigations with laboratory animals suggested that emotion-provoking stimuli produce physiological changes related to the fight or flight response. It was soon recognized that persistent stimuli of this type could produce physical illness. Cases of clinical pathology in humans were noted to follow severe emotional trauma, and eventually, physicians were trained to use a life chart as a diagnostic tool. By the mid-20th century, the general adaptation syndrome was posited as a mechanism by which physical environmental stressors could lead to diseases of adaptation. This led the way for the investigation of psychosocial stimuli as potential stressors, and soon life stresses became accepted risk factors for disease, especially psychosomatic disease. Stress events represented a change in a person's life, and hence the need to adapt. The life event checklist, which typically provided a count of life change events over the preceding 6 months or year, was a standard tool to rate the level of stress in people's lives. By the late 20th century, it was accepted that only undesired change, and not change per se, constituted stress. The field of social epidemiology has generally not pursued the biophysiological mechanisms by which the experience of negative events can produce illness, although some scientists now study the related concepts of allostasis and allostatic load.

Work proceeded in the social sciences on the concept of role-related stress. This emphasized chronic or recurrent conditions rather than "eventful" stress. Stressfulness in the work role, for example, is characterized in terms of task demands, degree of control over the amount or pace of work, danger, and noise. Jobs that are high in demand and low in control represent a particularly stressful work environment. Stress in the marital role may arise from interpersonal conflict, lack of intimacy or reciprocity, or conflict with the demands of other roles. The parent role, as suggested at the beginning of this entry, can bring with it stressful conditions that may be enduring, as living with adolescent turmoil. The absence of roles is another source of social stress: Consider the strains associated with childlessness for those who aspire to parenthood, or the lack of a partner or job.

Stress also results from one's social identity as an immigrant or a minority group member. Newcomers to a society may experience acculturation stress as they adapt to new customs, places, and an unfamiliar language. Intergenerational conflict may arise within immigrant families as parents and their children do not acculturate at the same pace. Visible minority individuals are at risk for exposure to discrimination stress, as are members of other marginalized groups who are defined by age, sexual orientation, or religion. These forms of stress exposure are distinct from life event stress in that they are embedded in the social roles and status characteristics that individuals have and tend to be chronic or recurring.

A Continuum of Stress Exposure

Stress phenomena occur across a spectrum from discrete events to continuous. The most discrete stressors are sudden, unexpected traumatic events such as an automobile accident, natural disaster, or criminal victimization. Somewhat less discrete are events that take

some time to conclude—for example, a divorce, serious illness, or going on welfare. Near the middle of the spectrum is the category of daily hassles. While hassles are not major stressors individually, their accumulation from day to day may represent an important stress source. More continuous in nature is the ongoing absence of an expected or desired social role, or nonevent. Stressors of this type include joblessness and childlessness. Chronic stress is the most continuous type; examples include living in a dangerous neighborhood, poverty, and living with a disability.

Eventful and chronic stress may be related through a process of stress proliferation. One example is when a worker loses a job because macroeconomic conditions led to the closure of a plant. Soon the loss of the individual's worker role and his or her source of income precipitate a financial crisis and increased conflict in the marital and parent roles—the "event" of job loss has proliferated stressful experience in a whole constellation of life domains. Social roles, and hence role-related stressors, do not occur in isolation.

Sometimes, the stressful meaning of a normally undesirable life event is negated by the context within which it occurs. Consider the separation or divorce of a person whose marital role history had been fraught with disappointment, conflict, and unhappiness—the event in such a case does not demand the kind of adaptation that is a threat to the person's well-being.

Stress as a Social Process

Stress may be viewed as the central means by which the structural arrangements of society create differential health outcomes for the people who occupy different social statuses and roles. Stress theory does not treat stress exposure as a health determinant in isolation: Stress is one element within a process that is closely linked to the social system. The amount of stress experienced is largely determined by an individual's social location. So are the social and personal resources that are available to forestall or cope with stressful events and circumstances as they occur. Stressful experiences may motivate social support (if it is available), which can mitigate their deleterious consequences. Successful resolution of a stressful event, such the loss of a home in a natural disaster, can build confidence that one can cope with future losses. Or it could be devastating to a person who had limited access to coping resources in the first place. Social inequality in the exposure to stress and in the availability of protective factors amplifies the production and reproduction of health disparities. That is because social structural arrangements are systematically related both to the amount a person is exposed to stress and access to the resources needed to mitigate its ill-health effects.

Stress arises from the social context of people's lives. There is systematic variation in the level of stress and coping resources across social status dimensions. Greater exposure to stress is associated with low education and poverty, unmarried status, minority group membership, and youth. The existence of a gender difference is less clear. Since stress and coping resources are important determinants of health and illness outcomes, stress functions as an epidemiological link between a society's structure and the health outcomes of its members.

—Donald A. Lloyd

See also Geographical and Social Influences on Health; Health Disparities; Social Capital and Health; Social Epidemiology; Social Hierarchy and Health

Further Readings

Aneshensel, C. S. (1992). Social stress: Theory and research. *Annual Review of Sociology*, *18*, 15–38.

Cohen, S., Kessler, R. C., & Gordon, L. U. (1995). *Measuring stress: A guide for health and social scientists.* New York: Oxford University Press.

Dohrenwend, B. P. (1998). *Adversity, stress, and psychopathology.* New York: Oxford University Press.

Pearlin, L. I., Lieberman, M., Menaghan, E., & Mullan, J. T. (1981). The stress process. *Journal of Health and Social Behavior*, *22*, 337–356.

Thoits, P. A. (1995). Stress, coping, and social support processes: Where are we? What next? [Extra issue]. *Journal of Health and Social Behavior*, 53–79.

Turner, R. J., Wheaton, B., & Lloyd, D. A. (1995). The epidemiology of social stress. *American Sociological Review*, *60*, 104–125.

Structural Equation Modeling

The roots of structural equation modeling (SEM) begin with the invention of least squares about 200 years ago, the invention of factor analysis about 100 years ago, the invention of path analysis about 75 years ago, and the invention of simultaneous equation

models about 50 years ago. The primary focus with SEM is on testing causal processes inherent in our theories. Before SEM, measurement error was assessed separately and not explicitly included in tests of theory. This separation has been one of the primary obstacles to advancing theory. With SEM, measurement error is estimated and theoretical parameters are adjusted accordingly—that is, it is subtracted from parameter estimates. Thus, SEM is a fundamental advancement in theory construction because it integrates measurement with substantive theory. It is a general statistical methodology, extending correlation, regression, factor analysis, and path analysis.

SEM is sometimes referred to as "latent variable modeling" because it reconstructs relationships between observed variables to infer latent variables. Many variables in epidemiological research are observable and can be measured directly (e.g., weight, pathogens, mortality). However, many variables are also inherently unobservable or *latent*, such as well-being, health, socioeconomic status, addiction, and quality of life. Measuring and interpreting latent variables requires a measurement theory. Latent variables and its respective measurement theory can be tested using an SEM technique called "confirmatory factor analysis." This involves specifying which latent variables are affected by which observed variables and which latent variables are correlated with each other.

SEM also provides a way of systematically examining reliability and validity. Reliability is the consistency of measurement and represents the part of a measure that is free from random error. In SEM, reliability is assessed as the magnitude of the direct relations that all variables except random ones have on an observed variable. This capability of SEM to assess the reliability of each observed variable and simultaneously estimate theoretical and measurement parameters is a fundamental methodological advancement. The potential for distortion in theoretical parameters is high when measurement error is ignored, and the more complicated the model the more important it becomes to take measurement error into account. Validity is the degree of direct structural relations (invariant) between latent and measured variables. SEM offers several ways of assessing validity. Validity differs from reliability because we can have consistent invalid measures. The R^2 value of an observed variable offers a straightforward measure of reliability. This R^2 sets an upper limit for validity because the validity of a measure cannot exceed its reliability.

Major Assumptions

Like other kinds of analyses, SEM is based on a number of assumptions. For example, it assumes that data represent a population. Unlike traditional methods, however, SEM tests models by comparing sample data with the implied population parameters. This is particularly important because the distinction between sample and population parameters has been often ignored in practice. SEM generally assumes that variables are measured at the interval or ratio level, and ordinal variables, if used at all, are truncated versions of interval or ratio variables. Hypothesized relationships are assumed to be linear in their parameters. All variables in a model are assumed to have a multivariate Gaussian or normal distribution. Therefore, careful data screening and cleaning are essential to successfully work with SEM.

SEM shares many assumptions with ordinary least squares regression and factor analysis. For example, the error of endogenous latent variables is uncorrelated with exogenous variables. The error of the endogenous observed variables is uncorrelated with the latent endogenous variables. The error of the exogenous observed variables is uncorrelated with the latent exogenous variables. The error terms of the endogenous latent variables and the observed endogenous and exogenous variables are mutually uncorrelated. This is the result of combining factor analysis and regression in one overall simultaneous estimation.

Steps in SEM

Specification

Models are constructed by defining concepts, clarifying the dimensions of each concept, forming measures of the dimensions, and specifying the expected empirical relationships between the measures and the construct. The accuracy of parameter estimates is partly dependent on the correctness of the theory and partly dependent on the validity of the measurement. There is always more than one model that fits the data, and thinking about these alternative models and testing them helps refine theory. Depicted in Figure 1 is a path diagram—a common way to represent models. The circles represent latent variables, squares represent observed variables, double-headed arrows represent correlations, and single-headed arrows represent causal effects. The one-to-one correspondence between path diagrams and sets of structural equations facilitates communication and clarification of all

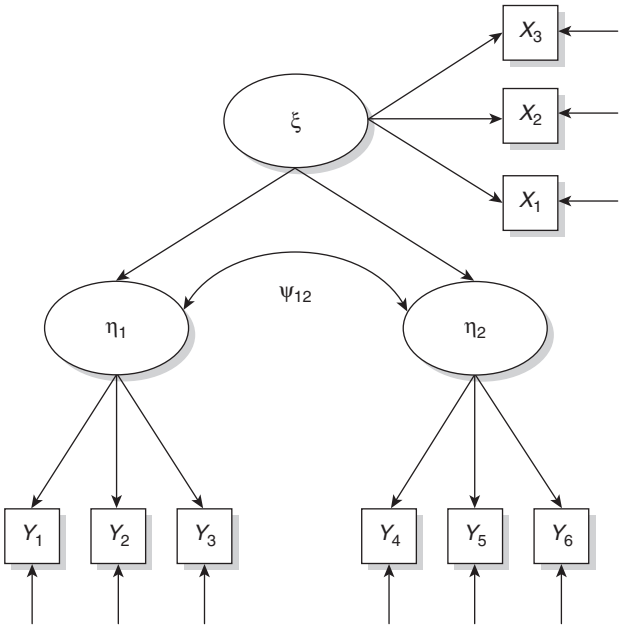


Figure 1 Example of a Path Diagram in SEM

parameters and their interrelationships. Model parameters are fully specified, which means stating a hypothesis for every parameter.

Identification

Models are composed of a set of equations with known and unknown parameters. Identification is the problem of determining whether there is a unique solution for each unknown parameter in the model. It is a mathematical problem involving population parameters, not sample size. A model can fail to be identified even with a large sample. There are a number of rules that if followed ensure identification. The most common is the *t* rule. The *t* in the *t* rule refers to the number of free parameters specified in the model. Specifically, a model is identified if the *t* value is equal to or smaller than half the number of observed variables multiplied by the number of observed variables plus 1 $[t \leq (1/2)(p)(p+1)]$. The *t* rule is a necessary but not sufficient condition for identification. Other rules are the scaling rule, three-indicator rule, null-β rule, recursive rule, and rank-and-order rules.

Estimation

SEM estimation procedures use a particular fitting function to minimize the difference between the population and the sample. Basically, this is a recipe to transform data into an estimate. The data matrix for SEM must be positive definite, a mathematical requirement for the estimation algorithms. Maximum likelihood is the default estimator in most SEM programs. Maximum likelihood is based on the idea that the sample is more likely to have come from a population of one particular set of parameter values than from a population of any other set of values. Maximum likelihood estimation is the vector of values that creates the greatest probability of having obtained the sample in question. This method of estimation is asymptotically unbiased, consistent, and asymptotically efficient, and its distribution asymptotically normal. If the sample is large, no other estimator has a smaller variance. There are two drawbacks with maximum likelihood. First, it assumes a normal distribution of error terms, which is problematic for many measures in the health and social sciences fields. Second, the assumption of multinormality is even more problematic, again because of the extensive use of crude measures.

In choosing estimators, the main choice is between maximum likelihood and weighted least squares. The weighted least squares estimator is used when multivariate normality is lacking and, especially, when some of the variables are ordinal. Although weighted least squares is computationally demanding, it is important to have a large sample size when some variables are ordinal. Other choices in estimators include generalized least squares and unweighted least squares. Maximum likelihood and generalized least squares are very similar. The generalized least squares estimator weights observations to correct for unequal variances or nonzero covariance of the disturbance terms. It is used when variable distributions are heteroscedastic or when there are autocorrelated error terms. An unweighted least square is used with variables that have low reliability. This estimator is less sensitive to measurement error than maximum likelihood or generalized least squares. Research shows estimates from the unweighted least square to be similar in models with and without error, while maximum likelihood estimates without and without errors are very different.

Fitting

After a model is estimated, its fit must be assessed. There are more than 20 different fit measures to assess misfit and goodness of fit. They are based on

six different criteria: (1) the discrepancy between the sample covariance matrix and the fitted (population) covariance, (2) accounting for observed variances and covariance, (3) maximizing the fit of a cross-validated model, (4) including a penalty for unnecessarily estimating parameters or creating fewer degrees of freedom, (5) the amount of improvement over a baseline model, and (6) separating the measurement model from the latent variable model.

Most of the existing fit measures are tied directly or indirectly to the chi-square ratio. This chi-square statistic is based on the same general idea as the familiar chi-square comparison between the observed and expected values. The difference is that in SEM our substantive interest or hypothesis is the null hypothesis. In traditional applications, we want to reject the hypothesis of no difference between observed and expected frequencies so that we can accept the alternative hypothesis of a difference. In contrast, with SEM, we want to find no difference between the expected and observed values. Therefore, the smaller the chi-square values the better because this leads to a failure to reject the null hypothesis, which is our substantive interest.

The chi-square statistic assumes that the variables in a model are multivariate normal, that the data are unstandardized (covariance as opposed to correlations matrixes), that sample sizes are at least $N > 100$ and preferably $N > 200$, and that the model holds exactly in the population. This chi-square has been found robust to skew violations but sensitive to Kurtosis violations. The interpretation of chi-square depends on adequate sample sizes. With large samples tiny deviations can lead one to reject the null hypothesis, which again in SEM is of substantive interest.

SEM fit measures are not applicable in exactly identified models. In exactly identified models (when degrees of freedom $= 0$), the sample variances and covariance always equal the estimates of the population variances and covariance because there is only enough information to calculate one estimate per parameter. A limitation of chi-square is that the closer the model is to being exactly identified, the higher the chi-square value. In other words, chi-square values always decrease when parameters are added to the model. With an overidentified model (degrees of freedom > 0), the overall fit can differ from the fit of different parts of the model. A poor overall fit does not help to detect areas of poor fit. The overall fit statistics also do not tell us how well the independent variables predict the dependent variables.

A good fit does not mean that model is "correct" or "best." Many models fit the same data well. Measurement parameters often outnumber theoretical parameters. Therefore, a "good fit" may reflect the measurement and not the theory. There is considerable discussion about fit measures. The best current advice in evaluating model fit is to seek a nonsignificant chi-square (at least $p > .05$ and preferably .10, .20, or better); an IFI (incremental fit index), RFI (relative fit index), or CFI (comparative fit index) greater than .90; low RMSR (root mean square residual) and RMSEA (root mean square of approximation) values, plus a 90% confidence interval for RMSEA $< .08$; and a parsimony index that show the proposed model as more parsimonious than alternative models.

Modification

It is not uncommon for models to exhibit a poor fit with the data. There are many potential sources of error, including an improperly specified theory, poor correspondence between the theory and the model, and causal heterogeneity in the sample. Modifications are typically made to poor-fitting models, and most SEM software packages provide modification indices that suggest which changes can improve model fit. However, using these indices in the absence of theory represents one of the main abuses of SEM. It is important that a systematic search for error is conducted and that modifications are based on theory or to generate new theory. A well-fitting respecified model does not represent a test. Respecified models must be tested on new data.

Specialized Techniques

SEM is a highly flexible methodology that allows for many special types of models to be examined. The most common models are those with unidirectional (recursive) causal effects, but SEM also allows for bidirectional (nonrecursive) effects to be tested. Stacked or multiple groups can also be examined, which facilitates interpretation and tests of interaction. Repeated measures designs can be analyzed using an SEM technique called "latent growth curves." This provides a way of examining both linear and nonlinear changes over time. Recent advances in software also provide a way of accounting for hierarchical or nested data structures, including survey weights.

Summary

SEM is a flexible and extensive method for testing theory. These models are best developed on the basis of substantive theory. Hypothesized theoretical relationships imply particular patterns of covariance or correlation. Statistical estimates of the hypothesized covariance indicate, within a margin of error, how well the models fit with data. The development and testing of these models advance theory by allowing latent variables, by including measurement error, by accepting multiple indicators, by accommodating reciprocal causation, and by estimating model parameters simultaneously. Structural equation models subsume factor analysis, multiple regression, and path analysis. The integration of these traditional types of analysis is an important advancement because it makes possible empirical specification of the linkages between imperfectly measured variables and theoretical constructs of interest.

The capabilities, technical features, and applications of SEM are continually expanding. Many of these advances are reported in the journal *Structural Equation Modeling* and communicated on the international and interdisciplinary SEM listserv called SEMNET. This listserv also archives its discussion and provides a forum for offering and receiving advice, which makes it an invaluable resource for epidemiologists and other social scientists learning and using SEM.

—David F. Gillespie and Brian Perron

See also Factor Analysis; Measurement; Regression

Further Readings

De Stavola, B. L., Nitsch, D., Silva, I. D. S., McCormack, V., Hardy, R., Mann, V., et al. (2005). Statistical issues in life course epidemiology. *American Journal of Epidemiology*, *163*, 84–96.

Der, G. (2002). Structural equation modelling in epidemiology: Some problems and prospects [Commentary]. *International Journal of Epidemiology*, *31*, 1199–1200.

Hoyle, R. H. (Ed.). (1995). *Structural equation modeling.* Thousand Oaks, CA: Sage.

Keller, S. D., Ware, J. E., Jr., Bentler, P. M., Aaronson, N. K., Alonso, J., Apolone, G., et al. (1998). Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: Results from the IQOLA project. *Journal of Clinical Epidemiology*, *51*, 1179–1188.

Kline, R. B. (2004). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Singh-Manous, A., Clarke, P., & Marmot, M. (2002). Multiple measures of socio-economic position and psychosocial health: proximal and distal measures. *International Journal* of Epidemiology, *31*, 1192–1199.

Study Design

Epidemiologic studies have traditionally been categorized as having "descriptive" or "analytic" designs. Descriptive studies are viewed primarily as hypothesis-generating studies and usually take advantage of routinely collected data to describe the distribution of a disease in a population in terms of the basic descriptors of person, place, and time. Analytic studies are further divided into "observational" and "experimental" study designs and are viewed as approaches suitable for testing specific hypotheses about disease etiology or the efficacy of disease prevention strategies. The main categories of observational studies are the cohort, case-control, nested case-control, case-cohort, case crossover, and cross-sectional designs. The most commonly employed experimental designs used in epidemiologic research include the classic randomized clinical trial and the quasi-experimental nonrandomized study design used to evaluate the effectiveness of population-based disease prevention approaches.

Descriptive Epidemiology

Data Sources

Descriptive epidemiologic studies are designed to determine the distribution of a disease in a population with regard to person, place, and time. The numbers of individuals in the population who are diagnosed with or die from various diseases are obtained from sources such as vital records files, disease registries, and surveys. Death certificates provide information on the underlying cause of death and provide basic sociodemographic data on the decedent such as age, gender, race/ethnicity, marital status, and place of residence at the time of death. Birth certificates are used to study the incidence of various birth outcomes such

as low birthweight (LBW) and its relationship to various parental factors such as maternal age. Birth defects registries also exist in some areas and combine data from birth certificates and reports from hospitals, physicians, and genetic testing laboratories.

Cancer registries now exist in all regions of the United States and are used to enumerate the number of total and specific forms of cancer that occur in a defined population over a specified time period. Cancer registries routinely collect information from hospital records and pathology laboratories regarding various clinical factors such as the cancer's anatomic location and histological type, clinical and pathologic stage of the disease, and information on the methods used to diagnosis the cancer. The cancer records also include data on various sociodemographic characteristics such as age, gender, race/ethnicity, marital status, and place of residence at the time of diagnosis. The address listed in the vital record or disease registry report at the time of death, birth, or diagnosis can be coded to the individual's census tract of residence using computer-based matching algorithms. The census tract reports contain data on various measures of socioeconomic status such as income and education for the geographic area in which the individual resided at the time of birth, death, or disease diagnosis. These numerator data are then combined with population denominator data to create disease incidence or death rates by person, place, or time.

Person, Place, and Time

Characteristics of persons include age, gender, race/ethnicity, marital status, and various measures of socioeconomic status such as education and income. Most diseases show distinct patterns of occurrence with regard to these personal characteristics. Breast cancer steadily increases with age until about the age of menopause at which time the age curve flattens. After the menopause, breast cancer incidence again increases with advancing age, albeit at a slower rate. These and other observations have led researchers to consider the possibility that pre- and postmenopausal breast cancer arise from separate etiologic processes. Numerous other examples exist regarding the relationship between disease incidence and mortality and age.

Descriptive epidemiologic studies have also shown that incidence and mortality rates for many diseases vary in relationship to personal characteristics. For example, mortality attributed to hypertension has been shown to occur more frequently among African Americans when compared with whites, while the risk for adult brain tumors is higher in men than women, most likely due to occupational exposures to chemicals.

Disease incidence and mortality patterns may also show distinctive geographic patterns. One of the earliest clues that early infection with hepatitis B virus (HBV) might be related to the development of liver cancer came from observations that countries with a high incidence of liver cancer were the same counties that reported high HBV infection rates. Researchers observed a strong concomitant geographic variation in liver cancer and HBV infection rates around the world. These data led to the development of case-control, cohort, and intervention studies that confirmed a strong association between HBV and liver cancer. The findings from the epidemiologic studies were further supported by strong evidence derived from animal models.

Variations in disease incidence and mortality over time have also provided insights into disease etiology and the effectiveness of intervention programs in a population-based setting. In the mid-1960s, approximately 60% to 65% of white males in the United States were cigarette smokers. Due in large part to the success of public health smoking prevention and cessation programs, the number of American males who currently smoke cigarettes is approximately 20%. An examination of age-adjusted lung cancer death rates for males in the United States between the years 1930 and 2005 show the waxing and waning of the cigarette-induced lung cancer epidemic. Between 1930 and 1980, male lung cancer death rates increased every year at a steady rate. However, by 1980 the success of the smoking prevention and cessation programs began to take effect, and the annual rate of increase in lung cancer mortality began to slow and then level off by 1990. Beginning in 1991, lung caner death rates began to fall for the first time in more than 60 years and have continued to decrease for the past 15 years.

Analytic Epidemiology

Analytic studies are further divided into "observational" and "experimental" designs. Observational studies include cohort, case-control, nested case-control, case-cohort, case crossover, and cross-sectional designs. These types of studies have also been referred to as natural experiments in that they are designed to take advantage of exposure/disease relationships that occur naturally in human populations.

Cohort Studies

The word *cohort* comes from Latin and originally referred to "one of 10 divisions of an ancient Roman legion [and later] a group of individuals having a statistical factor [such] as age or class" in common (*Merriam-Webster Online Dictionary*, www.m-w.com). The epidemiologic cohort represents a study base that is defined according to various characteristics. Figure 1 shows the basic structure of a typical cohort study. Note that baseline medical examinations or reviews of medical records are used to identify individuals who have already been diagnosed with the disease of interest. Since the purpose of the cohort study is to calculate incidence rates among exposed and nonexposed subgroups within the cohort, it is vital to start with a cohort that does not include prevalent cases. Various exposures of interest are collected at baseline through a variety of methods. A cohort study of factors related to heart disease might use questionnaires to collect baseline data on various lifestyle exposures such as diet, physical activity, and tobacco and alcohol use. Serum samples could also be drawn to measure levels of cholesterol, triglycerides, and other biochemical markers, while anthropometric methods could be used to categorize cohort members based on body fat deposition using measures such as body mass index and skin fold measurements. Since some of these measures may change over time, cohort members would be reexamined and reinterviewed on a periodic basis, perhaps annually. The annual follow-up surveys and clinic visits would also be designed to identify members of the cohort who have been diagnosed or died from heart disease during the past year. To confirm a diagnosis reported by a cohort member during the annual follow-up, every attempt is made to obtain medical records related to the diagnosis. Physicians trained in cardiology would review the available medical records to confirm the diagnosis. The reviewers should be blinded as to the exposure status of the cohort member and use standardized diagnostic criteria in making their decisions. Deaths that occur between annual surveys or clinic visits are obtained from family members or by matching cohort member information against the National Death Index. High-quality cohort studies maintain excellent follow-up rates of 90% or higher and obtain medical records for 98% of patients with a diagnosis of the disease of interest. Maintaining a small loss to

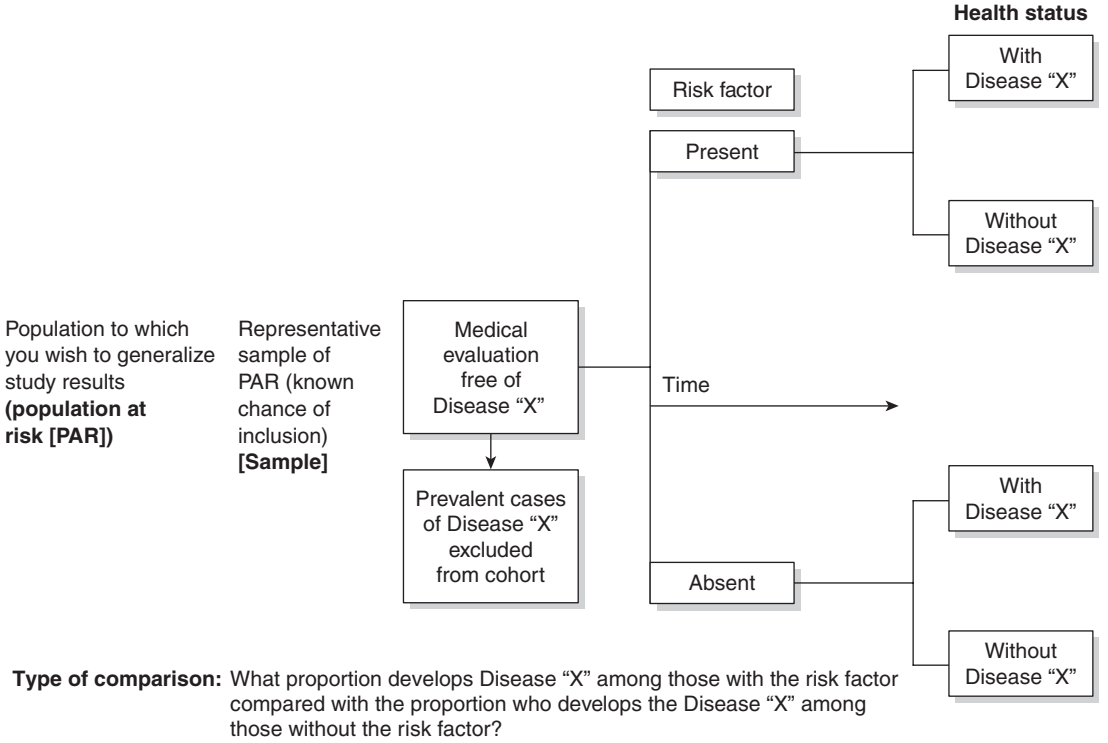


Figure 1 Basic Design of a Cohort Study: Flow Chart

follow-up rate is critical to ensuring that study results are not affected by selection bias where cohort members who are lost to follow-up are different from those who are not lost to follow-up with regard to both exposure and disease status.

Cohorts may be either "static" or "dynamic." In a static cohort study, all cohort members are enrolled at about the same time and are followed for a short period of time, thus minimizing loss to follow-up and the effects of competing causes of death. An example would be a cohort study of maternal factors related to LBW where a cohort of women are all registered in a prenatal clinic during their third month of pregnancy at about the same calendar time and are then followed to term. Let's assume that the exposure of interest is maternal cigarette smoking during the pregnancy and that LBW ($< 2{,}500$ g) serves as the outcome variable. This is a static cohort given that all cohort members are enrolled at the same time and the number of young, healthy mothers who are lost to follow-up or who die during the short 6-month period of the study is likely to be small. The task is to calculate a cumulative incidence of LBW children among smoking and nonsmoking mothers. Dividing the cumulative incidence of LBW among smoking mothers by the cumulative incidence among nonsmoking mothers provides the relative risk of a LBW child among smoking mothers when compared with nonsmoking mothers. The calculations are as follows:

$$\frac{\text{Number of LBW children among smoking mothers}}{\text{Number of smoking mothers initially enrolled}} = \frac{30}{1000} = 30 \text{ per } 1000,$$

$$\frac{\text{Number of LBW children among nonsmoking mothers}}{\text{Number of nonsmoking mothers initially enrolled}} = \frac{15}{1000} = 15 \text{ per } 1000,$$

$$\text{Relative risk} = \frac{\text{Cumulative incidence in smoking mothers}}{\text{Cumulative incidence in nonsmoking mothers}} = \frac{30/1000}{15/1000} = 2.0.$$

If the cohort structure is dynamic, cohort members are enrolled over a longer period of time and the follow-up time is usually measured in years. The long follow-up time leads to more cohort members being lost to follow-up or dying before they develop the disease of interest. These follow-up issues and the staggered enrollment period call for the use of a different approach to calculating disease occurrence among the exposed and nonexposed. This type of design calls for the calculation of an incidence density rate (IDR) where the numerator of the rate is still the number of cohort members diagnosed with the disease of interest, over a specified period of time. On the other hand, since each cohort member will enter and leave the cohort at different times, the investigators need to calculate the person-years of risk (PYR) for each cohort member. The sum of the individual PYR forms the denominator for calculating the IDR. Dividing the IDR among the exposed by the IDR among the nonexposed give us the rate ratio as an estimator of the relative risk.

In a dynamic cohort where the number of cohort members diagnosed with the disease during the follow-up period is small relative to a large number of PYR, the rate ratio is an excellent estimator of the relative risk. The calculations of the IDR and the incidence density rate ratio (IDRR) for a hypothetical cohort study of the risk of developing lung cancer among smokers versus nonsmokers are as follows:

$$\frac{\text{Number of smoking cohort members diagnosed with lung cancer}}{\text{PYR among nonsmoking cohort members}} = \frac{60}{5800} = 10.3 \text{ per } 1000 \text{ PYR},$$

$$\frac{\text{Number of nonsmoking cohort members diagnosed with lung cancer}}{\text{PYR among nonsmoking cohort members}} = \frac{15}{10100} = 1.5 \text{ per } 1000 \text{ PYR},$$

$$\text{IDRR} = \frac{\text{Incidence density rate among smokers}}{\text{Incidence density rate among nonsmokers}} = \frac{10.3}{1.5} = 6.9.$$

Case-Control Studies

The basic idea of the case-control design is to select a group of cases (individuals with the disease

of interest) and a group of controls (individuals without the disease of interest) and then measure the extent of exposure in the two groups. The basic outline of the case-control design is shown in Figure 2. A more detailed discussion of the issues covered in this section can be found in the series of three articles published by Wacholder et al., which are listed below in the "Further Readings" section. In theory, all case-control studies are embedded within a cohort, although the nature and structure of the underlying cohort is not always easy to discern. Cases and controls may be sampled from either a primary or a secondary study base. A primary study base requires a mechanism for ensuring complete or nearly complete ascertainment of all eligible cases in a defined population and a mechanism for assuring that all noncases have an equal chance of being selected from the same underlying population that produced the cases. The basic principle is that there is a reciprocal relationship between the case and control selection procedures such that if a case had not been diagnosed with the disease, the individual would have the opportunity to be selected as a control and vice versa. Let us assume that we want to study the relationship between estrogen replacement therapy in postmenopausal women and the risk of developing breast cancer. Let us further assume that we decide to conduct this study in a small country that has a high-quality national cancer registry the existence of which ensures that we can ascertain all or nearly all eligible cases. Furthermore, the country maintains a complete registry of the population from which we can select control women without breast cancer. The use of the case-control design in the presence of these resources would help to ensure that we have satisfied the reciprocity principle and that we are dealing with a primary study base. Access to a primary study base helps reduce the potential for creating selection bias at the design phase by minimizing the chances that cases and controls are selected for inclusion in the study based on the presence or absence of the exposure of interest.

However, these conditions frequently do not exist and investigators are often forced to use a secondary study base to select cases and controls. A common example involves selecting cases from one or more hospitals or medical care practices. The exact nature and structure of the underlying study base that gave rise to the hospitalized cases and controls is not always clear. In addition, the chances of being selected as a case or control depends on the extent to which the case and

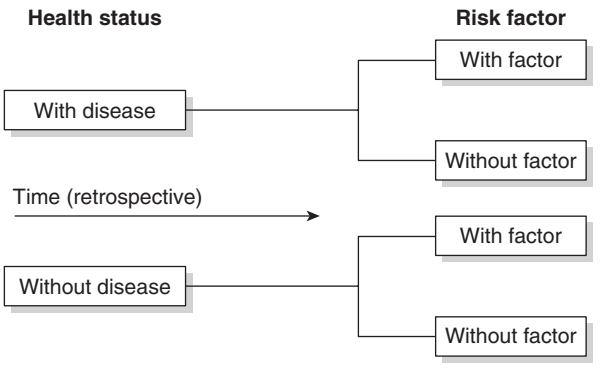


Figure 2 Basic Design of a Case-Control Study: Flow Chart

control diseases usually result in hospitalization and the hospital referral patterns for the case and control diagnoses in the underlying target population. If referral of the case and control diseases to the hospital is based on the presence or absence of the exposure, then there is a clear chance that study findings will be affected by design-based selection bias. Selection bias may also occur in case-control studies when not all eligible controls agree to be interviewed, tested, or examined and exposure information is absent for these nonparticipating study subjects and if the patterns of nonparticipation are related to both disease and exposure status.

The investigator needs to pay careful attention to methods and materials used to confirm the existence of the disease in cases. A standard and well-recognized set of diagnostic criteria should be employed. Diagnoses may be based on information from a variety of sources, including disease registry, hospital, and pathology reports. In some cases, it may be necessary to have an expert panel of pathologists conduct a blinded review of the original biopsy slides to accurately classify the disease, a step that is necessary, for instance, when attempting to classify subtypes of non-Hodgkin's lymphoma.

Another major threat to the internal validity of the case-control study is the difficulty encountered when attempting to assess exposure status retrospectively. The extent to which valid measurement of prior exposure can be obtained depends on the nature of the exposure, the amount of details required, and the length of time that has existed between the interview and the exposure event. Cigarette smoking and

alcohol consumption patterns can be assessed with a fair degree of accuracy through the use of structured questionnaires. Other lifestyle behaviors such as diet and physical activity are more difficult to measure. Using similar methods to collect exposure data for cases and controls is vitally important.

Investigators also employ biomarkers of internal dose as a means of determining exposures. Although these biomarkers can be quite useful, the laboratory values may only reflect recent rather than long-term exposures. In addition, the investigator is often forced to use a surrogate biologic source such as sera in place of another tissue source, which may lead to more accurate results but is highly invasive and therefore impractical in a population-based study. The collection of biomarker specimens may also add to respondent burden, increasing the nonresponse rate and thus increasing the potential for selection bias.

Nested Case-Control and Case-Cohort Studies

This variant of the basic case-control study is used to select study subjects from within an established cohort study. Cases are defined as individuals who have been ascertained through the cohort follow-up procedures to have the disease of interest. Controls are selected from among the remaining cohort members who have not developed the disease. One or more control subjects are selected for each case from among the nondiseased cohort members who were at risk at the time the case was diagnosed. This "matching" procedure provides close control over time as a potential confounding variable. If the follow-up rate for cohort members is high, then the cases and controls are derived from the same primary study base and participation rates are a nonissue. The quality of the exposure information is improved because these data are collected closer in time to the actual events. The temporal relationship between the exposure and disease is also assured because the exposure data are collected within the original cohort structure prior to the development of the disease. Nested case-control studies can be especially useful when the exposure measurement involves an expensive laboratory assay. Assume, for example, that we want to assess the relationship between various hormone patterns and breast cancer risk in a cohort of 10,000 women followed for 7 years. While the baseline blood draws might be reasonable to obtain, the cost of 10,000 hormone assays would be prohibitive. Another approach is to obtain and freeze the baseline serum samples. Once a sufficient number of cases are identified through the cohort follow-up procedures, the investigators can then thaw and test the samples for the much smaller number of cases and controls. The case-cohort and nested case-control studies are similar in design and vary only with respect to the control sampling procedures employed. In the case-cohort study, the controls are selected at random from among all nondiseased cohort members without regard to matching at the design phase by time at risk. Rather, information is collected on time at risk and included as a potential confounding variable at the time of analysis.

Case-Crossover Studies

The selection of an appropriate control group is particularly difficult when attempting to identify risk factors for acute disease outcomes. One example would be a study attempting to determine the events that immediately preceded a sudden nonfatal myocardial infarction. Another example would include a study of events that occurred close in time to an occupational injury. In both examples, the investigators would be hard pressed to select a separate control group to assess these proximal risk factors. The case-crossover design has been developed for these types of situations. In the case-crossover design, the case serves as his or her own control. The basic idea is to determine if a particular exposure occurred close in time to the acute event and how frequently this exposure occurred during a "control" time period. Let us assume a hypothesis that acute nonfatal myocardial infarctions may be triggered by heavy physical exertion just prior to the event. Let us further assume that we know the days and times when the acute nonfatal myocardial infarctions occurred. A key event exposure could be defined as heavy physical exertion occurring within 30 min of the acute nonfatal myocardial infarction. The "control" period could be chosen as the same day and time during the previous week. One approach to the analysis of data from a case-crossover design involves a matched pair design. Using a case as his or her own control also has a number of other advantages, including savings in time and money that result from not having to interview members of a separate control group. Another is that personal characteristics that do not change over time, such as gender and race/ethnicity, are controlled at the design phase. Data on other potential confounders that are more transient by nature need to be collected and included in the analysis.

Cross-Sectional Studies

The cross-sectional study is usually considered to be a hypothesis-generating design. In cross-sectional studies, a population of individuals is cross-classified with regard to a disease and potential risk factors at one point in time. The basic cross-sectional design is shown in Figure 3. Because data on the disease and the exposure are collected at a point in time, this approach cannot provide estimates of disease incidence but instead produces an estimate of disease prevalence with regard to the possible risk factors. In addition, the lack of a time dimension also prohibits the investigator from drawing any firm conclusions regarding the temporal relationship between the disease and the exposures. However, a series of cross-sectional surveys embedded within a community-based intervention study can help to test a hypothesis. Conducting multiple surveys in intervention and comparison communities can help to show that the prevalence of the desired behavioral changes is occurring more readily in the intervention community rather than in the comparison community.

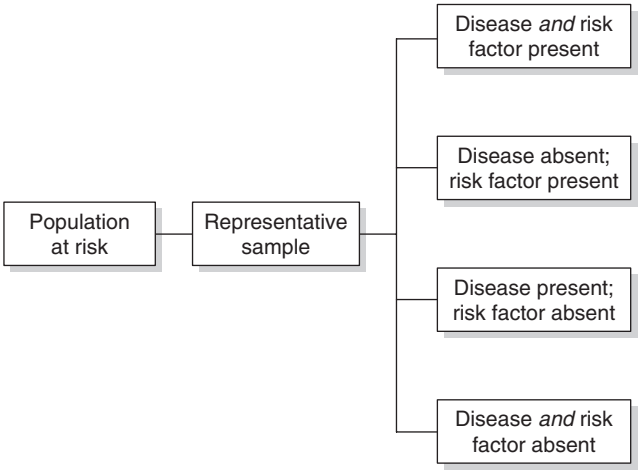


Figure 3 Cross-Sectional Studies: Flow Chart

Experimental and Quasi-Experimental Designs

Randomized Clinical Trials

Randomized trials are conducted to assess the efficacy and safety of a clinical intervention such as surgery, medical devices, and drug therapies. An outline of the basic design is shown in Figure 4. In some instances, a new intervention is tested against an existing form of treatment, such as testing a standard coronary artery metal stent against a new medicated stent. The study could be designed to assess a short-term outcome such as restenosis of the treated artery or long-term effects such as patient survival. In situations where a suitable treatment for the disease does not exist, the investigators might consider testing the new treatment against a "placebo" group, which typically receives no treatment or no effective treatment. Randomized trials are also conducted to assess the efficacy of various disease-screening techniques, the efficacy of natural products such as beta carotene for cancer chemoprevention, or the efficacy of various counseling approaches to effect desired behavioral changes.

The first phase of a randomized clinical trial involves developing inclusion criteria for patients. These criteria include sociodemographic factors such as age and clinical parameters that measure the patient's current health status as related to the disease under study and to other comorbid conditions. Individuals who meet the inclusion criteria are then randomized to either the intervention or comparison group. The randomization helps ensure that the intervention or treatment group and the comparison group will be similar with regard to baseline characteristics (potential confounders) that are strongly associated with the outcome under study. The extent to which randomization has equalized the distribution of potential confounding variables among the intervention and comparison groups can be determined by creating a table that compares the characteristics of the two groups at baseline following randomization.

Where feasible, a procedure called double-blinding is also used. For example, we might want to test the extent to which a new analgesic relieves headache better than buffered aspirin. Given that the outcome "pain relief" is somewhat subjective, the investigator might be concerned that if patients were aware of which drug they received, those treated with the "new" drug might be inclined to report more relief than those who knew they were treated with aspirin. To avoid this potential problem, a double-blind design is used where the patients and their primary care givers are not told whether the patient has been randomized to aspirin or the new drug. In addition, the individuals charged with interviewing the patients about their perceptions of pain relief are also blinded to the patients' treatment assignment. As an aid to maintaining patient blinding,

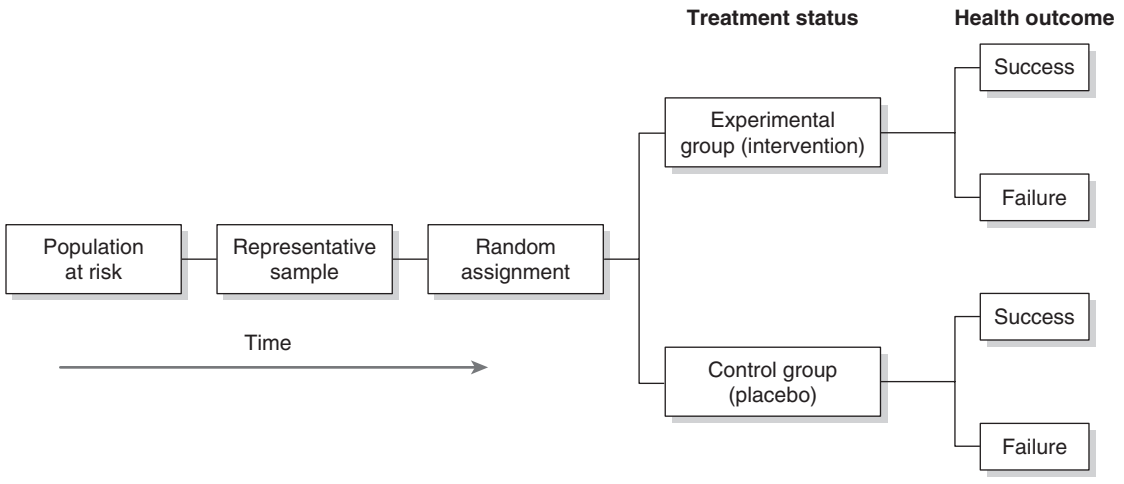


Figure 4 Experimental Design: Flow Chart

aspirin and the new drug would be designed to have the same shape, color, taste, and packaging. It is more difficult but not impossible to implement double-blinding in studies of interventions involving surgical or medical device interventions. In acupuncture treatment, the needles are often inserted using a small plastic tube as a guide. In studies of the potential therapeutic benefits of acupuncture, patients have been randomized to receive either real acupuncture treatments or to a "sham" acupuncture treatment group. The touch of the plastic guide tube on the patient's skin has been used to create a false sense that needles are actually being inserted. The procedures are applied to the patients' back in such a way that the patient is blinded to whether they are receiving the real or sham acupuncture treatment. The individuals charged with assessing the patients' perceptions of pain relief could also be blinded as to which group the patient had been randomized to.

Conduct of randomized trials should involve periodic analysis of the study data to determine if a preset level of statistical significance has been reached that would indicate that the new treatment has been shown to be superior to the old treatment or has more negative side effects than the old treatment. If either of these outcomes is observed, then the trial should be stopped on ethical grounds. These decisions are the main prevue of the safety monitoring board, which works closely with the analysis group to determine on an ongoing basis the observed benefits and risk associated with the new treatment.

Another key issue involves the extent to which patients randomized to the intervention group can be maintained on the intervention, particularly over long periods of time. Patient adherence is affected by the burden imposed by the intervention, including the amount of discomfort or negative side effects caused by the intervention. Adherence to the treatment regimen is measured in a number of ways, such as counting pills or weighing pill bottles at periodic visits to the clinic or through measuring the drug or its metabolites in blood or urine. The number of dropouts from the intervention group needs to be kept as low as possible and measured against the number of patients in the comparison group who adopt the intervention independently of the randomization process. Patients who switch groups during the course of the trial create some interesting issues with regard to data analysis. The most common approach to calculating an outcome measure such as the relative risk or the risk ratio involved has been referred to as the "intent to treat" approach. This approach analyzes the frequency of the outcome in the groups as originally randomized irrespective of whether patients failed to be maintained in their original groups.

Studies involving community-based interventions often involve what is described as a quasi-experimental design. Let us suppose that we are interested in determining if a media-based intervention can increase the number of adults in a population who are screened for hypertension and placed on an appropriate treatment regimen. Since the intervention will be community

based, it will not be feasible to randomize individuals to intervention or comparison group. The unit of analysis now becomes the community and not the individual. One approach would be to stratify the communities within a target area according to various sociodemographic characteristics and then randomly assign communities within each stratum to either the intervention group or the comparison group. To measure the impact of the intervention, the investigator could conduct special surveys of the population both prior to and after the intervention has been in operation. These cross-sectional surveys could provide a measure of increased medical surveillance for hypertension over time in both the intervention and comparison communities. Other design and measurement approaches that follow the general quasi-experimental design could, of course, be employed in these community-based studies.

—Philip C. Nasca

See also Bias; Causal Diagrams; Causation and Causal Inference; Descriptive and Analytic Epidemiology; Randomization; Sequential Analysis

Further Readings

Journal Articles

Wacholder, S., McLaughlin, J. K., Silverman, D. T., & Mandel, J. (1992). Selection of controls in case-control studies: 1. Principles. *American Journal of Epidemiology*, *135*, 1019–1028.

Wacholder, S., Silverman, D. T., McLaughlin, J. K., & Mandel, J. (1992). Selection of controls in case-control studies: 2. Types of controls. *American Journal of Epidemiology*, *135*, 1029–1041.

Wacholder, S., Silverman, D. T., McLaughlin, J. K., & Mandel, J. (1992). Selection of controls in case-control studies: 3. Design options. *American Journal of Epidemiology*, *135*, 1042–1050.

Introductory Epidemiologic Methods

Aschengrau, A., & Seage, G. R. (2007). Essentials *of epidemiology in public health* (2nd ed.). Boston: Jones & Bartlett.

Bhopal, R. (2002). *Concepts of epidemiology: An integrated introduction to the ideas, theories, principles, and methods of epidemiology.* New York: Oxford University Press.

Friis, R. H., & Sellers, T. (2004). *Epidemiology for public health practice* (3rd ed.). Boston: Jones & Bartlett.

Kleinbaum, D. G., Sullivan, K. M., & Barker, M. D. (2002). *ActivEpi* (CD-ROM and Companion Textbook). New York: Springer.

Koepsell, T. D., & Weiss, N. S. (2003). *Epidemiologic methods: Studying the occurrence of illness.* New York: Oxford University Press.

Merrill, R. M., & Timmreck, T. C. (2006). *Introduction to epidemiology* (4th ed.). Boston: Jones & Bartlett.

Oleckno, W. A. (2002). *Essential epidemiology: Principles and applications*. Long Grove, IL: Waveland Press.

Rothman, K. J. (2002). *Epidemiology: An introduction*. New York: Oxford University Press.

Intermediate to Advanced Epidemiologic Methods

Brownson, R. C., & Petitti, D. B. (2006). *Applied epidemiology: Theory to practice* (2nd ed.). New York: Oxford University Press.

Elwood, J. M. (2007). *Critical appraisal of epidemiological studies and clinical trials*. New York: Oxford University Press.

Rothman, K. J., Greenland, S., & Lash, T. L. (2007). *Modern epidemiology* (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.

Szklo, M., & Nieto, J. (2007). *Epidemiology: Beyond the basics* (2nd ed.). Boston: Jones & Bartlett.

Woodward, M. (2004). *Epidemiology: Study design and data analysis* (2nd ed.). New York: Chapman & Hall.

Suicide

The term *suicide* refers to deliberately ending one's own life and may also refer to someone who ends his or her life. Because the actor's intention is part of the definition of suicide, it is often a matter of judgment whether a particular death was due to suicide, was accidental, or was caused by a third party. Studies of suicide are complicated by inconsistent reporting due to the fact that different religions and cultures judge the act of suicide differently. In ambiguous cases, a death may be classified as suicide in a society for which that is a morally defensible choice and classified as accidental in a society in which suicide is considered shameful and in which consequences such as the inability to be buried among one's ancestors may follow. In addition, legal or practical considerations such as difficulty in collecting on a life insurance policy after the insured person has committed suicide may also influence whether a death is classified as suicide or not. It is even more difficult to get an accurate estimation of the number of people who attempted suicide and did not die: If they do not seek

medical treatment, they will not be counted; and if they do seek treatment, there is no guarantee that attempted suicide will be recorded as a cause.

Suicide in the United States

According to the Centers for Disease Control and Prevention, in the United States in 2001, 30,622 people died by suicide and about nine times that number were hospitalized or treated in emergency departments for attempted suicide. Suicide rates in the United States are highest in the spring and lowest in the winter, contrary to popular belief, and rates are higher in the western states as opposed to the eastern and Midwestern states. Women are about three times more likely than men to report attempting suicide in their lifetime, but men are four times more likely to die from suicide. Suicide rates are highest among Caucasians, followed by Native Americans and Alaska Natives.

Although rates of suicide among young people have declined in recent years, it is the third leading cause of death among people aged 15 to 24 years. In this age group, suicide rates are highest among Native Americans and Alaskan Natives and about six times as high for males as for females. Suicide rates increase with age and are highest among those aged 65 years and older; the male/female ratio of suicides in this age group is similar to that among persons aged 15 to 24 years.

A number of risk factors have been identified for suicide. Personal history factors associated with increased suicide risk include previous suicide attempts, a history of mental disorders, a history of alcohol and substance abuse, a family history of child abuse, and a family history of suicide. Concurrent medical and psychological risk factors include impulsive or aggressive tendencies, feelings of hopelessness or isolation, and physical illness. Other risk factors include lack of access to mental health care, recent traumatic events, access to lethal methods of suicide such as firearms, and local epidemics of suicide. Protective factors include access to medical and psychological care, family and community support, maintaining ongoing relationships with medical and mental health care providers, and personal skills in problem solving and conflict resolution.

Suicide in a Global Context

Statistics concerning suicide and related topics must be interpreted with even greater care when comparing rates across countries or estimating the global incidence of suicide, because the cultural and reporting issues discussed previously are that much greater when dealing with information from different countries with widely varying cultural attitudes and reporting systems. In addition, when making global comparisons over time, researchers must consider that the estimates for different years may not include data from the same set of countries.

The World Health Organization reports that in 2000 approximately 1 million people died from suicide, a mortality rate of 16 per 100,000. Suicide rates have increased by 60% worldwide over the last 45 years. Traditionally, suicide rates have been highest among elderly males, but they have been increasing among young people; in 1950, most cases of suicide (60%) were older than 45 years, while in 2000, the majority of cases (55%) were among people aged 5 to 44 years. Suicide rates for individual countries, as reported by the World Health Organization, are generally higher for men than for women and show considerable range: For some countries, the rate is less than 1.0 per 100,000 while some of the highest rates are for men in countries of the erstwhile Soviet Union, including Lithuania (74.3 per 100,000), the Russian Federation (69.3 per 100,000), Belarus (63.3 per 100,000), Kazakhstan (50.2 per 100,000), Estonia (47.7 per 100,000), and Ukraine (46.7 per 100,000). Reported suicide rates for women are highest in Asian and Eastern European countries and Cuba, including Sri Lanka (16.8 per 100,000), China (14.8 per 100,000), Lithuania (13.9 per 100,000), Japan (12.8 per 100,000), Cuba (12.0 per 100,000), the Russian Federation (11.9 per 100,000), and the Republic of Korea (11.2 per 100,000).

Physician-Assisted Suicide

Physician-assisted suicide or physician-assisted dying refers to a practice in which a doctor provides a means, such as an injection or prescription drugs, intended to hasten the death of a patient on the request of that patient. It is part of the larger category of *euthanasia*, a term formed by the Greek words for "good" and "death," which is sometimes translated as "the good death" or "the merciful death." Usage of the term *euthanasia* differs, and some include within this category the involuntary killing of people who have not requested to die and the hastening of death through the withdrawal of life-support systems. Physician-assisted suicide is a controversial topic

because of the ethical issues involved, the problems of protecting vulnerable people from abuses, and the association of euthanasia with the Nazi regime in Germany and similar historical abuses. Laws regarding physician-assisted suicide are constantly changing, but as of 2006 some form of physician-assisted suicide was legal in several European countries, including Switzerland, the Netherlands, and Belgium. Within the United States, in 2006 only Oregon had legislation allowing physician-assisted suicide.

—Sarah Boslaugh

See also Ethics in Health Care; Eugenics; Psychiatric Epidemiology; Violence as a Public Health Issue

Further Readings

Anderson, R. M., & Smith, B. L. (2003). Deaths: Leading causes for 2001. *National Vital Statistics Report*, *52*, 1–86.

Jamison, K. R. (1999). *Night falls fast: Understanding suicide*. New York: Vintage Books.

National Center for Injury Prevention and Control. (n.d.).*Suicide: Fact sheet*. Retrieved November 24, 2006, from http://www.cdc.gov/ncipc/factsheets/suifacts.htm.

Oregon Department of Human Services. (n.d.). *FAQs about the Death with Dignity Act*. Retrieved November 24, 2006, from http://www.oregon.gov/DHS/ph/pas/faqs.shtml.

Surgeon General, U.S.

The term *Surgeon General* is used in the United States to denote the supervising medical officer of the Public Health Commissioned Corps within the U.S. Department of Health and Human Services. The term is also applicable when referring to the senior medical officer within the U.S. Army and Air Force. Foreign governments also have equivalent positions but do not make use of this particular title.

The position of U.S. Surgeon General was created as a result of the reorganization and recognition of the Marine Hospital Service, a group of hospitals originally constructed to provide health services at key sea and river ports to merchant marines. Expansion of the military and growth in the science of public health led to the need for a national hospital system with centralized administration. The newly reconstructed Marine Hospital system was overseen by a Supervising Surgeon from 1871 to 1872, a Supervising Surgeon General from 1873 to 1901, and a Surgeon General from 1902 to date. Dr. John Woodworth was appointed the first Supervising Surgeon of the U.S. Marine Hospital Service, predecessor of today's U.S. Public Health Service, and Walter Wyman (1891–1911) was the first surgeon to hold the title of Surgeon General.

The U.S. Surgeon General oversees more than 6,000 members of the Public Health Commissioned Corps, holds the rank of Admiral of the Commissioned Corps, and ex officio is the spokesperson on matters of national public health. The U.S. Surgeon General conducts duties under the direction of the Assistant Secretary for Health and the Secretary of Health and Human Services. The Office of the Surgeon General is part of the office of Public Health and Science, Office of the Secretary, U.S. Department of Health and Human Services.

The U.S. Surgeon General is appointed to a 4-year term. This appointment is made after a recommendation from the President of the United States and the endorsement of the U.S. Senate. Official duties of this office are the following:

- Oversee the Commissioned Corps, a diverse collection of health professionals considered experts in public health.
- Provide leadership and direct response to public health matters, current and long term, and provide direction in matters of emergency preparedness and response.
- Establish, protect, and represent a commitment to national health through education and endorsement of empirically supported disease prevention and health promotion programs for the nation.
- Carry out communicative, advisory, analytical, and evaluative roles in matters of domestic and international scientific health policy with both governmental and nongovernmental agencies.
- Ensure quality in existing and planned public health practice throughout the professions by establishing research priorities and appropriate standards.
- Participate in various traditional and statutory federal boards, governing entities, and nongovernmental health organizations such as Board of Regents of the Uniformed Services University of the Health Sciences, the National Library of Medicine, the Armed Forces Institute of Pathology, the Association of Military Surgeons of the United States, and the American Medical Association.

—Floyd Hutchison

See also Governmental Role in Public Health; U.S. Public Health Service

Web Sites

U.S. Department of Health and Human Services, Office of the Surgeon General: http://www.surgeongeneral.gov.

Survey Research Methods

See Sampling Techniques

Survival Analysis

Survival analysis is a collection of methods for the analysis of data that involve the time to occurrence of some event and, more generally, to multiple durations between occurrences of events. Apart from their extensive use in studies of survival times in clinical and health-related studies, these methods have found application in several other fields, including industrial engineering (e.g., reliability studies and analyses of equipment failure times), demography (e.g., analyses of time intervals between successive child births), sociology (e.g., studies of recidivism, duration of marriages), and labor economics (e.g., analysis of spells of unemployment, duration of strikes). The terms *duration analysis*, *event-history analysis*, *failure-time analysis*, *reliability analysis*, and *transition analysis* refer essentially to the same group of techniques, although the emphases in certain modeling aspects could differ across disciplines.

The time to event T is a positive random variable with distribution function $F(t) = P[T \leq t], t \geq 0$. In biostatistics and epidemiology, it is more common to use the survivorship function or survival function $S(t) = 1 - F(t)$. Thus, $S(t)$ is the probability of being free of events at time t. In clinical studies, where T is the time of death of a patient, one refers to T as the *survival time* and $S(t)$ as the *probability of survival beyond t*. Since time and duration must have an origin, the specific context determines an appropriate starting point from which T is measured. For example, consider a clinical trial of competing treatments in which patients entering the study are randomized to treatment conditions. The time origin is the time of randomization or initiation of treatment, and T is the time until the primary endpoint (e.g., death) is reached.

Censoring

In clinical studies, patients enter the study at different points in time. For example, a 5-year study might be planned with a 2-year recruitment phase in which patients enter randomly. Patient follow-up begins at entry and ends at death (or some terminal endpoint) if observed before the end of year 5. The survival time T is then known. If the terminal event is not reached by the end of study, T is not observed but we know that $T > U$, where U is the follow-up time from entry to the end of study. The survival times of these patients are censored, and U is called the *censoring time*. Censoring would also occur if a patient died from causes unrelated to the endpoint under study or withdrew from the study for reasons not related to the endpoint. Such patients are lost to follow-up.

The type of censoring described above is called *right censoring*. If the true event time T was not observed but is known to be less than or equal to V, we have a case of *left censoring*. If all that is known about T is that it lies between two observed times U and V $(U > V)$, we say it is *interval censored*. For example, when periodic observations are made for the time to seroconversion in patients exposed to the human immunodeficiency virus, if seroconversion is observed, the time of conversion lies in the interval between the previous negative assessment and the first positive assessment. Right censoring occurs if seroconversion is not observed by the end of study, while left censoring is the case if the patient tests positive at the very first assessment.

Hazard Function

A useful concept in survival analysis is the hazard function h, defined mathematically by $h(t) = \lim_{\Delta t \to 0} P[T < t + \Delta t | t \geq t]/\Delta t$. The quantity $h(t)$ is not a probability. However, because $h(t)\Delta t \approx P[T < t + \Delta t | T \geq t]$, we can safely interpret $h(t)\Delta t$ as the conditional probability that the event in question occurs before $t + \Delta t$, given that it has not occurred before t. For this reason, $h(t)$ has been referred to as the *instantaneous risk* of the event happening at time t.

If the distribution of T is continuous, $h(t) = -d(\log S(t))/dt$. Then, $S(t) = \exp(-H(t))$ and $H(t) = \int_0^t h(u)\,du$ is the cumulative hazard function. (The relationship between S and H is more subtle when the distribution T is not continuous.) Other useful concepts in survival analysis are

Mean survival time	$\mu = E(T) = \int_0^\infty S(t)dt$
Mean survival restricted to time L	$\mu_L = E(\min(T, L)) = \int_0^\infty S(t)dt$
Percentiles of survival distribution	$t_p = \inf\{t > 0 : S(t) \le 1 - p\}$, $0 < p < 1$
Mean residual life at time t	$r(t) = E(T - t \mid T > t)$ $= \int_0^\infty S(u)du/S(t)$

Just as the H determines S, there is a unique relationship between r and S given by

$$S(t) = \frac{r(0)}{r(t)} \exp\left(- \int_0^t \frac{du}{r(u)} \right).$$

Because survival data are often quite skewed with long right tails, the restricted mean survival μ_L or the median survival time $t_{0.5}$ are generally preferred as summary statistics.

Parametric Distributions

Some common distributions for an event time T used in survival analysis are given below. The parameters α and λ have different interpretations, and $\alpha > 0$ and $\lambda > 0$.

1. *Exponential distribution:* $h(t) = \lambda, H(t) = \lambda,\ S(t) = \exp(-\lambda t)$, and $E(T) = \lambda^{-1}$.
2. *Weibull distribution:* $h(t) = \lambda t^{\alpha - 1}$, $H(t) = (\lambda t)^\alpha$, $S(t) = \exp(-(\lambda t)^\alpha)$, and $E(T) = \lambda^{-1}\Gamma(1 + \alpha^{-1})$, where Γ is the gamma function.
3. *Log-normal distribution:* Here, $\log T$ has a normal distribution with mean μ and standard deviation σ. Then,

$$S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma} \right),$$

$$h(t) = \frac{1}{\sigma t} \frac{\phi\left(\frac{\log t - \mu}{\sigma}\right)}{\left[1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)\right]},$$

and $E(T) = \exp(\mu + 1/2\sigma^2)$, where ϕ and Φ are, respectively, the standard normal density and cumulative distribution functions.

4. *Log-logistic distribution:*

$$h(t) = \frac{\lambda^\alpha \alpha t^{\alpha - 1}}{1 + (\lambda t)^\alpha},$$
$$H(t) = \log(1 + (\lambda t)^\alpha),$$
$$S(t) = \frac{1}{1 + (\lambda t)^\alpha},$$

and $E(T) = \lambda^{-1}\Gamma(1 + \alpha^{-1})\Gamma(1 - \alpha^{-1})$, if $\alpha > 1$, where Γ is the gamma function.

5. *Pareto distribution*:

$$h(t) = \frac{\lambda\alpha}{1 + \lambda t},$$
$$H(t) = \alpha \log(1 + \lambda t),$$
$$S(t) = \frac{1}{(1 + \lambda t)^\alpha},$$
$$E(T) = \lambda^{-1}(\alpha - 1)^{-1},\ \alpha > 1.$$

These distributions provide considerable flexibility in fitting a parametric distribution to survival data. For example, the Weibull distribution with $\alpha > 1$ has increasing hazard function, with $\alpha < 1$ the hazard function is decreasing, and with $\alpha = 1$ the hazard is constant. This last case results in the exponential distribution. For the log-logistic distribution with $\alpha \le 1$, the hazard is decreasing. With $\alpha > 1$, the hazard is increasing up to time $\lambda^{-1}(\alpha - 1)^{1/\alpha}$ and then decreases.

Generally, survival data exhibit skewness. To mitigate this, one might consider a log transformation. A general class of distributions in the location-scale family is specified by $\log T = \mu + \sigma\varepsilon$, where the constants μ and σ are called, respectively, the *location* and *scale* parameters. By specifying a distribution for the random variable εs, one induces a distribution on T. Clearly, the log-normal distribution is in this class. The Weibull distribution also belongs to this class where ε has a standard extreme value distribution $P[\varepsilon > y] = \exp(-e^y)$, $-\infty < y < \infty$, with $\sigma = \alpha^{-1}$, $\mu = -\log\lambda$. The log-logistic distribution is in this class with ε having a standard logistic distribution,

$P[\varepsilon > y] = 1/(1+e^{y})$, $-\infty < y < \infty$, with $\sigma = \alpha^{-1}$, $\mu = -\log\lambda$. However, the Pareto distribution does not belong to the location-scale family.

The location-scale family of survival distributions may also be expressed as $T = e^{\mu} T_0^{\sigma}$, with $\log T_0 = \varepsilon$. The survival distribution of T takes the form $S(t) = P[T_0 > (te^{-\mu})^{1/\sigma}] = S_0((te^{-\mu})^{1/\sigma})$, where S_0 denotes the survival distribution of T_0. Therefore, $S(t)$ is obtained from S_0 by either accelerating or decelerating the time. For example, if $\sigma = 1$ and $\mu < 0$, the time is accelerated. Distributions in the location-scale family are also called *accelerated failure time* (AFT) distributions.

Modeling Survival Data

Survival data obtained from a sample of patients often exhibit variation that sometimes can be explained by incorporating patient characteristics in survival models. For each patient, the observable data consist of the survival time T or the follow-up time U and a vector $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ of p covariates. For example, in clinical studies these covariates included indicator variables for treatment group; patient demographic variables such as age, race/ethnicity, level of educational attainment, and so on; and clinical variables such as comorbidities, laboratory assessments, and so on.

In longitudinal studies, some of these covariates might be assessed at several time points throughout the study making them dependent on time. The observed time is $X = \min(T, U)$ and an indicator δ of whether $X = T$ or $X = U$. The observable data on n patients are $\{(X_i, \delta_i, \mathbf{x}_i) : 1 \le i \le n\}$ with the subscript i indexing the patient. We generally assume that observations across patients are independent, and survival times are independent of the censoring times, given the covariates, which for now we assume are also fixed (i.e., are independent of time). The independence of T and U is important. A patient who is still at risk at time t, meaning that $X \ge t$ does not provide more information on survival than a patient with $T \ge t$. Censoring is therefore noninformative: It does not remove patients under study because they are at a higher or lower risk of death.

In the location-scale family of distributions, $\log T = \mu + \sigma\varepsilon$, we can incorporate covariate effects by modeling μ in the linear form $\mu = \mathbf{x}'\beta = \beta_0 + x_1\beta_1 + \cdots + x_p\beta_p$ with an intercept β_0. Additional patient heterogeneity could be accommodated by modeling the scale parameter σ, usually on the log scale as $\log\sigma = \mathbf{z}'\gamma$, where $\mathbf{z}$ is a subset of $\mathbf{x}$. The underlying model for our survival data looks like the familiar linear regression model, $Y_i \equiv \log T_i = \mathbf{x}_i'\beta + \sigma\varepsilon_i$, except that $E(\varepsilon_i|\mathbf{x}_i)$ is not necessarily zero, although it is a constant. Furthermore, we will not have complete observations on Y_i since some patients will have censored survival times. For inference, we must maintain the assumption of a parametric distribution for ε that is independent of $\mathbf{x}$. Because of this assumption and that of noninformative censoring, estimation of all regression parameters can be accomplished via maximum likelihood estimation using the censored sample $\{(X_i, \delta_i, \mathbf{x}_i) : 1 \le i \le n\}$. Standard tests such as the Wald test and the likelihood ratio test could be used to assess the statistical significance of components of (β, γ). In the interest of parsimony, one might choose to retain only significant covariate effects in the model.

An easy interpretation of the β coefficients is directly from $E(\log T|\mathbf{x}) = \mathbf{x}'\beta + (\varepsilon)$. For example, consider a placebo-controlled treatment study with a single treated group. Patients are randomized to treatment $(x_1 = 1)$ or placebo $(x_1 = 0)$. If σ is constant, the effect of treatment (vs. placebo) is $\beta_1 = E(\log T|x_1 = 1) - E(\log T|x_1 = 0)$. The aforementioned maximum likelihood estimation allows one to test the hypothesis of no treatment effect—namely, $H_0 : \beta_1 = 0$. If our model has additional covariates, $\mathbf{x}^* = (x_2, \ldots, x_p)$, that have no interactions with the treatment indicator x_1, then β_1 is called the adjusted effect of treatment because $\beta_1 = E(\log T|x_1 = 1, \mathbf{x}^*) - E(\log T_1 = 0, \mathbf{x}^*)$—that is, the treatment effect keeping all other covariates $\mathbf{x}^*$ held fixed in the treated and placebo groups. For a continuous covariate x_1 (e.g., age), the interpretation of β_1 is the partial effect—the effect of one unit increase in x_1 on the expected log survival, holding all other covariates fixed (and assuming no interactions with x_1).

The effect of a covariate on any summary measure, such as the survival function, mean survival time $E(T|\mathbf{x})$, or the median survival time (or other percentiles), may be used to provide comparisons on the original untransformed scale. For example, suppose the underlying survival distribution is Weibull and we want to assess the effect of treatment on mean survival time. From $E(T|\mathbf{x}) = \exp(\mathbf{x}'\beta + \log\Gamma(\sigma + 1))$, we get $E(T|x_1 = 1, \mathbf{x}^*)/E(T|x_1 = 0, \mathbf{x}^*) = \exp(\beta_1)$; that is, $\exp(\beta_1)$ is the adjusted effect of treatment, relative to placebo, on mean survival. This interpretation carries over to all the previously mentioned location-scale models. Similarly, a common structure applies

to the percentiles of the survival distributions in location-scale models. The upper $100p$th percentile t_p of the survival distribution is given by $t_p = \exp(\mathbf{x}'\beta + \sigma z_p)$, where z_p is the corresponding percentile of the distribution of ε. For example, in the log-normal distribution the median survival is $t_{0.5} = \exp(\mathbf{x}'\beta)$ and therefore in the aforementioned scenario we obtain an entirely equivalent interpretation of $\exp(\beta_1)$ as the adjusted effect of treatment relative to placebo, on median survival time. It is worthy of note that these structural similarities do not carry over to the survival function $S(t|\mathbf{x}) = S_0((te^{-\mathbf{x}'\beta})^{1/\sigma})$, which has very different functional forms for different survival distributions.

Nonparametric Methods

In the absence of a plausible parametric assumption for the survival distribution, its estimation may be based on the relationship $S(t) = \exp(-H(t))$. For right-censored survival data $\{(X_i, \delta_i, \mathbf{x}_i) : 1 \le i \le n\}$, we define two basic quantities: $N(t)$ is the number of events observed up to time t, and $Y(t)$ is the number of patients at risk of the event just prior to time t. While $N(t)$ is a step function that increases only at event times and remains constant between event times, $Y(t)$ decreases just after event and censoring times. Hence, $\Delta N(t)/Y(t)$ is the proportion of patients with events at t among those who are at risk and not censored just prior to t. This provides an estimator $\hat{S}(t) = \exp(-\hat{H}(t))$, where $\hat{H}(t) = \sum_{u \le t} \Delta N(u)/Y(u)$. These are called the Nelson-Aalen estimators. Because $Y(t) = 0$ if t is larger than the largest observed time (called $X_{(n)}$) in the sample, whether it is an event or censoring time, we cannot define these estimators for $t > X_{(n)}$. Generally, the Nelson-Aalen estimator $\hat{S}(t)$ is close to the Kaplan-Meier estimator $\tilde{S}(t)$ given by $\tilde{S}(t) = \prod_{u \le t}(1 - \Delta N(u)/Y(u))$. If $X_{(n)}$ is an event time, then $\tilde{S}(t) = 0$ for $t \ge X_{(n)}$. If not, $\tilde{S}(t)$ is left undefined for $t > X_{(n)}$.

The two estimators $\hat{S}(t)$ and $\tilde{S}(t)$ have equivalent large-sample properties. The fact that they are consistent and asymptotically normal estimators of $S(t)$ allows one to compute pointwise confidence intervals for $S(t)$. For example, a $100(1-\alpha)\%$ confidence interval for $S(t)$ has approximate confidence limits given by $\tilde{S}(t) \pm z_{1-\alpha/2}\tilde{\sigma}(t)$, where

$$\sigma^2(t) = \{\tilde{S}^2(t)\} \sum_{u \le t} \frac{\Delta N(u)}{Y(u)(Y(u) - \Delta N(u))}$$

is an estimate of the large sample variance of $\tilde{S}(t)$. However, better approximations may be obtained by first computing a confidence interval for a transformed $S(t)$, such as $\log[S(t)]$ or $\log[\text{tlog } S(t)]$, and then retransforming back to the original scale.

Another innovation is to compute a simultaneous confidence band for $S(t)$ over the interval $t_L \le t \le t_U$, where t_L and t_U are appropriately specified time points. Practically, these points should be between the smallest and largest failure times in the observed sample. A simultaneous confidence band will preserve the confidence level 1 tgα by deriving statistics $\hat{L}(t)$ and $\hat{U}(t)$ such that the interval $[\hat{L}(t), \hat{U}(t)]$ contains $S(t)$ for all $t_L \le t \le t_U$ with probability $1 - \alpha$—that is, $P[[\hat{L}(t), \hat{U}(t)]\ S(t), t_L \le t \le t_U] = 1 - \alpha$. In contrast, a pointwise confidence interval will only guarantee $P[[\hat{L}(t), \hat{U}(t)]\ S(t)] = 1 - \alpha$ for each t. When confidence intervals for the survival function are desired at several time points, using a simultaneous band is recommended.

Semiparametric Models

Many applications require comparison of survival across groups after controlling for other covariates that might have influence on survival. For parametric models within the location-scale family, we have seen how comparisons can be made after the estimation of regression parameters via maximum likelihood estimation. A widely used model is the proportional hazards model (PHM) in which the hazard function $h(t|\mathbf{x})$ is given by $h(t|\mathbf{x}) = h_0(t)\exp(\mathbf{x}'\beta)$, where $h_0(t)$ is a baseline hazard function that is left completely unspecified. Because this part of the model is nonparametric, the term *semiparametric* is used. The model was introduced by D. R. Cox in 1972 and has since received unprecedented attention. Because inference is often focused on the parameter β, Cox suggested an approach to its estimation by using a partial likelihood function that is independent of h_0. After the estimation of βs, we can estimate the cumulative hazard $H(t|\mathbf{x}_0) = H_0(t)\exp(\mathbf{x}_0'\beta)$ and the survival function $S(t|\mathbf{x}_0) = \exp(-H(t|\mathbf{x}_0))$ at a specified covariate profile $\mathbf{x}_0$.

The PHM and AFT models are distinct. In fact, the Weibull distribution is the only member of both classes. For the Weibull, $h(t|\mathbf{x}) = \alpha t^{\alpha-1}\lambda_0^{\alpha}(t)\exp(-\alpha\mathbf{x}'\beta)$, where $h_0(t) = \alpha t^{\alpha-1}$. The two parameterizations differ slightly.

To interpret the PHM model, consider once again a treatment study of survival with a single treated

group ($x_1 = 1$) and a placebo group ($x_1 = 0$). Because $h_0(t)$ is arbitrary, there is no need for an intercept parameter. The hazard at time t in the treated group is $h(t|x_1 = 1) = h_0(t)\exp(\beta_1)$, whereas in the placebo group it is $h(t|x_1 = 0) = h_0(t)$ that provides an interpretation for $h_0(t)$. Therefore, the relative hazard (treated vs. placebo) is $h(t|x_1 = 1)/h(t|x_1 = 0) = \exp(\beta_1)$. When $\beta_1 > 0$, the treated group has at each time t a higher probability of the event (e.g., death) than the placebo group. When $\beta_1 < 0$, this is reversed. Indeed, $S(t|x_1 = 1) = \{S(t|x_1 = 0)\}^{\exp(\beta_1)}$, $t > 0$, and so the two survival functions are ordered. With additional covariates, $\mathbf{x}^* = (x_2, \ldots, x_p)$, which have no interactions with x_1, $\exp(\beta_1)$ is the adjusted relative hazard for treated versus placebo when $\mathbf{x}^*$ is held fixed in both groups. Similarly, for a continuous covariate the interpretation of $\exp(\beta_1)$ is the relative hazard for a unit increase in x_1 while holding all other covariates fixed (and assuming no interactions with x_1).

In multicenter studies where survival data are collected from different sites, it is appropriate to consider a stratified PHM in which the hazard in site j is specified by $h_j(t|\mathbf{x}) = h_{j0}(t)\exp(\mathbf{x}'\beta)$, that is, we incorporate a separate baseline hazard for each site (stratum). Stratification could also be considered if the standard proportion hazard assumption might seem inapplicable. For example, instead of using age within the covariate mix $\mathbf{x}$ we could form age strata, and within each age stratum, the proportional hazards assumption applies to all other covariates under consideration.

Another extension of the PHM is to incorporate time-dependent covariates. This is often necessary in longitudinal studies with periodic measurements of important factors related to survival. The time-dependent model is $h(t|\mathbf{x}(t)) = h_0(t)\exp(\mathbf{x}'(t)\beta)$, where $\mathbf{x}(t)$ denotes the cumulative covariate history up to time t. Although the regression parameter β can still be estimated using the partial likelihood function, inference on the survival function will require some assumptions on the relationship between the stochastic development of $\{\mathbf{x}(t) : t \geq 0\}$ and the event time T. To retain the expression $S(t|\mathbf{x}(t)) = \exp(-\int_0^t h(u|\mathbf{x}(u)\,du)$, strict exogeneity of $\{\mathbf{x}(t) : t \geq 0\}$ with respect to T is sufficient. This means $P[\mathbf{x}(t + \Delta t)|T \geq t + \Delta t, \mathbf{x}(t)] = P[\mathbf{x}(t + \Delta t)|\mathbf{x}(t)]$, that is, given $\mathbf{x}(t)$, future values of the covariates are not influenced by future values of T. Time-fixed covariates are obviously strictly exogenous. A patient's age at any time t is independent of the event $T \geq t$ and is therefore strictly exogenous. There are other types of time-dependent covariates for which inference on the relative hazard parameter β is still valid. However, for more general considerations we need to model jointly the covariate process $\{\mathbf{x}(t) : t \geq 0\}$ and event time T.

Other Extensions

Several important developments that extend the basic survival analytic technique have arisen to support applications in many fields. One important extension is to multistate models and multiple failure times. The finite-state Markov process in continuous time is ideally suited to consider transitions from one state to another. The concepts of hazard functions are replaced by intensity functions and survival probabilities by transition probabilities. For example, in follow-up studies in cancer patients, we might consider periods of remission (State 0) and relapse (State 1) before death (State 2). A simple three-state model examines the impact of covariates on the transitions $0 \to 1$, $1 \to 0$, $0 \to 2$, and $1 \to 2$ by modeling the intensities $\alpha_{jk}(t|\mathbf{x}) = \alpha_{jk0}(t)\exp(\mathbf{x}'_{jk}\beta)$ just like in a proportional hazard model. Here, α_{jk} is the transition intensity for $j \to k$. The interpretation of these intensities is as follows: from entry into State j, $\alpha_{jj} = \sum_{k \neq j} \alpha_{jk}$ is the hazard function for the duration in state j; given that exit from state j occurs at time t, the exit state k is selected with probability α_{jk}/α_{jj}.

The survival model is a special case of a multistate model with just two states, alive and dead. A competing risks model has one origination state (alive, State 0) and several destination states (States $k = 1, \ldots, K$) corresponding to causes of death. The intensities α_{0k} are the cause-specific hazard functions, and $\alpha_{00} = \sum_k \alpha_{0k}$ is the overall hazard for survival.

Multiple failure times arise when an individual can potentially experience several events or when there is a natural clustering of event times of units within a cluster (e.g., failure times of animals within litters, survival times of zygotic twins). Often, this situation can also be cast in the framework of a multistate model if interest lies in modeling only the intensities of each event type and not on the multivariate joint distribution of the event times. Other important developments in survival analysis include Bayesian survival analysis and frailty models.

—Joseph C. Gardiner and Zhehui Luo

See also Censored Data; Cox Model; Hazard Rate; Kaplan-Meier Method

Further Readings

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer-Verlag.
Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling—A guide for social scientists.* Cambridge, UK: Cambridge University Press.
Collett, D. (2003). *Modelling survival data for medical research* (2nd ed.). London: Chapman & Hall.
Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. London: Chapman & Hall.
Elandt-Johnson, R. C., & Johnson, N. L. (1980). *Survival models and data analysis.* New York: Wiley.
Fleming, T. R., & Harrington, D. P. (1991). *Counting processes and survival analysis*. New York: Wiley.
Heckman, J. J., & Singer, B. (Eds.). (1985). *Longitudinal analysis of labor market data*. Cambridge, UK: Cambridge University Press.
Hosmer, D. W., & Lemeshow, S. (1999). *Applied survival analysis*: *Regression modeling of time to event data*. New York: Wiley.
Hougaard, P. (2000). *Analysis of multivariate survival data.* New York: Springer-Verlag.
Ibrahim, J. G., Chen, M.-H., & Sinha, D. (2001). *Bayesian survival analysis.* New York: Springer-Verlag.
Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.
Klein, J. P., & Moeschberger, M. L. (1997). *Survival analysis: Techniques for censored and truncated data*. New York: Springer-Verlag.
Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge, UK: Cambridge University Press.
Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. New York: Wiley.
Lee, E. T. (1992). *Statistical methods for survival data analysis.* New York: Wiley.
Marubini, E., & Valsecchi, M. G. (1995). *Analysing survival data from clinical trials and observational studies*. New York: Wiley.
Nelson, W. (1982). *Applied lifetime data analysis*. New York: Wiley.
Thernau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model.* New York: Springer-Verlag.

Syndemics

A syndemics model of health and illness focuses attention on the multiple interconnections that occur between copresent diseases and other health-related problems in a population as well as within individual sufferers at both the biological and social levels. This orientation, which developed initially within medical anthropology and diffused to epidemiology and public health, emerged in response to the dominant biomedical conceptualization of diseases as distinct entities in nature, separate from other diseases, and independent of the social contexts in which they are found. While isolating diseases, assigning them unique labels (e.g., AIDS, TB), and narrowly focusing on their immediate causes and expressions laid the foundation for the development of modern pharmaceutical and other biomedical approaches to sickness, it has become increasingly clear that diseases and other health conditions (e.g., nutritional status) interact synergistically in various and consequential ways and that the social conditions of disease sufferers are critical to understanding the impact of such conditions on the health of both individuals and groups. A syndemics approach examines disease concentrations (i.e., multiple diseases affecting individuals and groups), the pathways through which they interact biologically within individual bodies and within populations and thereby multiplying their overall disease burden, and the ways in which social environments, especially conditions of social inequality and injustice, contribute to disease clustering and interaction.

Disease Interactions

Interest in a syndemics approach has been driven by growing evidence of significant interactions among comorbid diseases. One such interaction has been found, for example, between type 2 diabetes mellitus and various infections, such as hepatitis C viral infection. Several factors are known to contribute to the onset of type 2 diabetes, including diet, obesity, and aging. The role of infection, however, is only beginning to be understood. Already, it is known that risk for serious infections of various kinds increases significantly with poor diabetes control, but appreciation of more complex relationships between infection and type 2 diabetes is now emerging as well. The Third National Health and Nutritional Examination Survey (NHANES III) found that the frequency of type 2 diabetes increases among people who have been infected with the hepatitis C virus. Similarly, several health reports note that diabetes is present in as many as 37% of those who are critically ill with severe acute respiratory syndrome.

The nature of interaction among diseases may vary and need not require direct physical interaction to produce new or amplified health consequences (e.g., as in AIDS, changes in biochemistry, or damage to organ systems caused by one pathogenic agent may facilitate the spread or impact of another agent). Direct interaction, however, including gene mixing among different types of pathogenic agents, has also been described, such as the molecular in vivo integration of the avian leukosis virus and Markek's disease virus (MDV) in domestic fowl. Both these cancer-causing viruses are known to infect the same poultry flock, the same chicken, and even the same anatomic cell. In coinfected cells, the retroviral DNA of the avian leukosis virus can integrate into the MDV genome, producing altered biological properties compared with the parental MDV. In studies of human populations, a lethal synergism has been identified between influenza virus and pneumococcus, a likely cause of excess mortality from secondary bacterial pneumonia during influenza epidemics. It is disease interactions of this sort that are a central biological component in syndemics. Syndemic theory seeks to draw attention to and provide a framework for the analysis of these interactions.

Social Origins of Syndemics

Beyond disease clustering and interaction, the term *syndemic* also points to the importance of social conditions in disease concentrations, interactions, and outcomes. In syndemics, the interaction of diseases or other health-related problems commonly occurs because of adverse social conditions (e.g., poverty, stigmatization, oppressive social relationships, health care disparities) that put socially devalued and subjugated groups at heightened risk for disease or limit access to timely and effective care. With reference to tuberculosis, for example, it is impossible to understand its persistence in poor countries as well as its recent resurgence among the poor in industrialized countries without assessing how social forces, such as political violence and racism, come to be embodied and expressed as individual pathology.

Identified Syndemics

Various syndemics (although not always labeled as such) have been described in the literature, including the SAVA syndemic (substance abuse, violence, and AIDS); the hookworm, malaria, and HIV/AIDS syndemic; the Chagas disease, rheumatic heart disease, and congestive heart failure syndemic; the asthma and infectious disease syndemic; the malnutrition and depression syndemic; the mental illness and HIV/AIDS syndemic; and the sexually transmitted diseases syndemic. Additional syndemics are being identified around the world as public health officials, researchers, and service providers begin to focus on the connections among diseases and the social context factors that foster disease interactions.

Syndemic Research

Several lines of future syndemics inquiry have been described. First, there is a need for studies that examine the processes by which syndemics emerge, including the specific sets of health and social conditions that foster the occurrence of multiple epidemics in a population and how syndemics function to produce specific kinds of health outcomes in populations. Second, there is a need to better understand processes of interaction between specific diseases with each other and with health-related factors such as malnutrition, structural violence, discrimination, stigmatization, and toxic environmental exposure that reflect oppressive social relationships. Finally, there is a need for a better understanding of how the public health systems and communities can best respond to and limit the health consequences of syndemics. Systems are needed to monitor the emergence of syndemics and to allow "early-bird" medical and public health responses designed to lessen their impact. Systematic ethno-epidemiological surveillance with populations subject to multiple social stressors must be one component of such a monitoring system. Current efforts by researchers at the Centers for Disease Control to expand the discussion of syndemics in public health discourse is an important step in the development of a funded research agenda that addresses these research needs. Given the nature of syndemics, this research requires a biocultural/social approach that attends to both clinical and social processes.

—Merrill Singer

See also Comorbidity; Geographical and Social Influence on Health; Global Burden of Disease Project; Health Disparities; Medical Anthropology

Further Readings

Freudenberg, N., Fahs, M., Galea, S., & Greenberg, A. (2006). The impact of New York City's 1975 fiscal crisis on the tuberculosis, HIV, and homicide syndemic. *American Journal of Public Health*, *96*, 424–434.

Singer, M. (1996). A dose of drugs, a touch of violence, a case of AIDS: Conceptualizing the SAVA syndemic. *Free Inquiry in Creative Sociology*, *24*, 99–110.

Singer, M., & Clair, S. (2003). Syndemics and public health: Reconceptualizing disease in bio-social context. *Medical Anthropology Quarterly*, *17*, 423–441.

Singer, M., Erickson, P., Badiane, L., Diaz, R., Ortiz, D., Abraham, T., et al. (2006). Syndemics, sex and the city: Understanding sexually transmitted disease in social and cultural context. *Social Science and Medicine*, *63*, 2010–2021.

Stall, R., Mills, T., Williamson, J., & Hart, T. (2003). Association of co-occurring psychosocial health problems and increased vulnerability to HIV/AIDS among urban men who have sex with men. *American Journal of Public Health*, *93*, 939–942.

Synergism

See Effect Modification and Interaction

T

Targeting and Tailoring

Information and communication are important and powerful tools for helping enhance population health. Generally speaking, health information that is carefully designed for a specific group or individual is more effective in capturing attention and motivating changes in health-related attitudes and behaviors than information designed for a generalized audience or with no particular audience in mind. The two most common types of health information customized for specific audiences are *targeted* and *tailored* communication.

Both targeted and tailored health communication are audience-centered approaches driven by a careful analysis of intended recipients. Both approaches use what is learned from this analysis to customize health messages, sources of information, and channels of information delivery to maximize the reach and effectiveness of a health communication to a particular audience. In targeted communication, the unit of audience analysis and customization is a particular *group*, while in tailored communication the unit of audience analysis and customization is one specific *individual.* Thus, these approaches are often referred to as "group targeted" and "individually tailored" health communication.

Targeted Health Communication

The rationale for group-targeted health communication is summarized in three key assumptions: (1) there is diversity among the members of any large population with respect to the determinants of a given health decision or behavior and also among the characteristics that affect exposure and attention to, processing of, and influence of health messages; (2) for any health-related behavior, homogeneous population subgroups can be defined based on shared patterns of these determinants and characteristics; and (3) different health communication strategies are needed to effectively reach different population subgroups. In health communication terminology, these population subgroups are called *audience segments.*

The concept of audience segmentation has its roots in marketing and advertising consumer products and services and is now widely accepted as a best practice in health communication. Historically, audience segmentation in public health has been driven by findings from disease surveillance and epidemiological studies that identified population subgroups with elevated risk or burden of disease. Because of the limited types of data typically collected in these surveillance and research activities, the resulting audience segments were often fairly unsophisticated, relying on only demographic variables (e.g., teenagers, African Americans), health status indicators (e.g., pregnancy, blood pressure), broad behavioral categories (e.g., smokers, men having sex with men), or sometimes simple combinations of the three (e.g., pregnant teenage girls who smoke). Boslaugh, Kreuter, Nicholson, and Naleid (2005) showed that simple segmentation strategies such as those relying on demographic variables alone provided little improvement over no segmentation at all in understanding physical activity behavior. Thus, while epidemiological data such as these are invaluable for identifying population subgroups with great need for risk reduction, they are of little use in helping health communication planners and developers

make critical decisions about effective message design or selection of interpersonal and media channels to reach members of those subgroups.

More sophisticated, multivariable approaches to audience analysis and segmentation consider demographic, psychographic, geographic, health status and behavioral characteristics, and the dynamic interplay among them. For example, Vladutiu, Nansel, Weaver, Jacobsen, and Kreuter (2006) observed that parents' beliefs and behaviors related to child injury prevention varied significantly based on whether or not they were first-time parents and by the age of their oldest child. In short, beliefs about the effectiveness of injury-prevention measures and perceptions of how important injury prevention was to others in their lives were related to injury prevention behaviors among parents of preschool children, but not among parents of infants and toddlers. Attitudes about injury prevention (i.e., "injuries are normal part of childhood," "injuries can't be prevented") predicted injury-prevention behaviors for first-time parents but not parents of multiple children. Thus, rather than promoting injury prevention by delivering the same information to all parents, these findings suggest that it may be more effective to segment the population of parents into multiple subgroups, for example, those with only one child versus those with two or more children. For parents with only one child, targeted communications would focus on changing specific beliefs that may undermine child injury-prevention behaviors.

Although one could conceivably identify hundreds of population subgroups or audience segments for any behavior of interest, some segments will be higher priorities than others for receiving targeted communication. When the level of need or health risk is equal, audience segments that are larger in size, easily identifiable in the population, and accessible through existing channels are generally more promising as prospects for population health improvement through targeted communication.

Tailored Health Communication

While targeted health communication seeks to reach a given population subgroup, tailored health communication is designed to reach one specific person. To customize health information in this way, data from or about an individual are gathered and processed to determine what messages will be needed to address one's unique needs, then these messages are assembled in a predesigned template for delivery to the individual. This process of assessing behaviors and key determinants of behavior change for different individuals, matching messages to those determinants, and providing tailored feedback to the individual is usually automated using a computer database application. In a typical tailoring program, a patient in a primary care setting might answer questions about his or her dietary behaviors and beliefs on a paper form or computer kiosk while in the waiting room. One's answers would be processed using a set of computerized algorithms that identify not only potential problem areas in one's diet, but also beliefs and skills that would affect one's motivation or ability to make dietary changes. Using this information, the computer program would then select from a large library of messages only those that were appropriate given one's answers on the assessment. These messages would then be printed in a magazine or newsletter for the person or shown as on-screen feedback.

Findings from individual studies and a growing number of literature reviews indicate that tailored communication is more effective than nontailored communication for capturing attention, increasing motivational readiness to change a behavior, and stimulating behavioral action. Tailoring effects have been found for a wide range of behaviors (e.g., diet, physical activity, smoking cessation, immunization, cancer screening, injury prevention), in diverse populations, and across many settings (e.g., health care, worksites, online). Tailored health communication has also taken many forms, including calendars, magazines, birthday cards, and children's storybooks. Most tailoring to date has been based on individuals' responses to questions assessing behaviors (e.g., fruits and vegetables consumed per day, injury-prevention devices used) and/or determinants of behavior change derived from theories of health behavior change (e.g., readiness to change, perceived benefits and barriers). While health information in any medium—audio, video, Internet—can be tailored to individual recipients, the evidence base for tailoring effects comes almost exclusively from tailored print communication.

Why does tailored health communication have these effects? One explanation is that recipients of tailored health information perceived it as personally relevant. The theories of information processing propose that individuals are more motivated to thoughtfully consider messages when they perceive them to be personally relevant. In a randomized trial, Kreuter, Bull,

Clark, and Oswald (1999) showed that participants exposed to tailored weight loss materials engaged in significantly more cognitive processing of the information than those exposed to two different types of non-tailored weight loss materials. Webb, Simmons, and Brandon (2005) suggest that the mere *expectation* of customized communication (i.e., telling people they will receive information made just for them) may be sufficient to stimulate tailoring-like effects. In a study of what they call "placebo tailoring," smokers were randomly assigned to receive one of three booklets that shared identical smoking-related content but varied in degrees of ostensible tailoring. As the amount of apparent tailoring increased, so too did the favorability of smokers' responses to the booklets, although no differences in actual behavior change were observed.

Because the cost and time involved in developing tailored health communication programs can exceed that required for less customized forms of communication, it should be undertaken only in situations when it has distinct advantages over other approaches. Perhaps obviously, tailoring health messages would be unnecessary and even wasteful if a single message was equally effective for all or most members of a larger group (i.e., as in *targeted* communication). However, if members of that same group vary significantly on behavioral determinants of some outcome of interest, the added effort to tailor messages for each individual might be justified. Similarly, because individual-level data are required to tailor messages, this approach should only be considered when there is a feasible way to obtain data from (i.e., through an assessment) or about (i.e., through an existing database) individual members of the intended audience. To date, no studies have directly compared effects or cost-effectiveness of targeted with tailored communication. However, for practical purposes, it is probably less important to identify one superior approach than it is to determine how the two can be used in concert to help achieve public health objectives.

—Matthew W. Kreuter and Nancy L. Weaver

See also Health Behavior; Health Communication

Further Readings

Boslaugh, S. E., Kreuter, M. W., Nicholson, R. A., & Naleid, K. (2005). Comparing demographic, health status, and psychosocial strategies of audience segmentation to promote physical activity. *Health Education Research*, *20*(4), 430–438.

Kreuter, M. W., Bull, F. C., Clark, E. M., & Oswald, D. L. (1999). Understanding how people process health information: A comparison of tailored and untailored weight loss materials. *Health Psychology, 18*(5), 487–494.

Kreuter, M. W., Strecher, V. J., & Glasman, B. (1999). One size does not fit all: The case for tailoring print materials. *Annals of Behavioral Medicine, 21*(4), 1–9.

Kreuter, M. W., & Wray, R. J. (2003). Tailored and targeted health communication: Strategies for enhancing information relevance. *American Journal of Health Behavior, 27*(Suppl. 3), S227–S232.

Slater, M. D. (1996). Theory and method in health audience segmentation. *Journal of Health Communication, 1*, 267–283.

Vladutiu, C. J., Nansel, T. R., Weaver, N. L., Jacobsen, H. A., & Kreuter, M. W. (2006). Differential strength of association of child injury prevention attitudes and beliefs on practices: A case for audience segmentation. *Injury Prevention, 12*, 35–40.

Webb, M. S., Simmons, N., & Brandon, T. H. (2005). Tailored interventions for motivating smoking cessation: Using placebo-tailoring to examine the influence of expectancies and personalization. *Health Psychology*, *24*(2), 179–188.

Target Population

The terms *statistical population, population of interest, universe*, and simply *population* are often used interchangeably when referring to a target population. What they have in common is that they define a group of people or other units that are the focus of a study and to whom the results are intended to generalize. The term *population* derives from the origins of statistics being used to describe human populations. However, a population may be people (such as the population of smokers), objects (such as hospital records), events (such as deaths or births), or measurements on or of the people, objects, or events (such as ages of smokers or occurrence of births).

A group may be defined as a population due either to an inherent characteristic of the group itself (such as residence in a particular city) or to a particular characteristic of interest to the researcher (such as having a particular health condition). A population may be very large (such as the population of the United States, estimated at more than 300 million in 2007) or very small (such as the population of

patients with progeria, of whom only 42 were known in the world as of 2006).

If a target population consists of people, objects, or events, then it is the set of sampling units about which investigators would like to draw conclusions or the set of all the members of the group under consideration. This population is the entire set of units to which findings of the survey are to be generalized. If the population consists of measurements taken on people, objects, or events, then, ideally, it is the set of all measurements that a researcher would like to have in answering a scientific question. These data are all possible or all hypothetically possible observations of those measurements.

In a statistical study, the researcher must define the population being studied. Typically, defining the target population is easy: It is all subjects possessing the common characteristics that are being studied. However, often for practical reasons the researcher must also define a study population, meaning the actual population of people from whom the sample will be drawn. For instance, if a researcher working in Ohio is interested in the effects of smoking on systolic blood pressure in adults aged 18 to 65 years, this would be the target population (sometimes called the ''theoretical population''), while for practical reasons, the study population from whom the sample would be drawn might be all people between the ages of 18 and 65 years on January 1, 2007, residing within Greene County, Ohio. Additionally, each set of measurements that might be drawn on these individuals may be considered a population, so we could speak of the population of systolic blood pressure measurements or the population of smoking status indicators (whether each individual smokes) for adults aged 18 to 65 years in Greene County, Ohio.

It is important to note that whether a data set is considered a population or a sample depends on the context in which the data are to be viewed. In the previous example, if the researcher is interested in generalizing the study to all adults within southwestern Ohio, then the set of data for all adults in Greene County would be a sample. However, if the researcher is interested only in studying the relationship of smoking and blood pressure within Greene County, then that same set of data would constitute the population. Additionally, the population ''adults aged 18 to 65 years in Greene County'' could be used to draw a sample for study: If the results are intended to generalize all adults in southwestern Ohio (or the entire United States), this would be a study population; if a sample was drawn with the intent of generalizing only the adults of Greene County, it would be the target population.

The researcher must be careful when defining the population to ensure that the appropriate sampling units are included in the sampling frame to avoid exclusion bias. In the smoking and blood pressure example, the ideal sampling frame would be a list of all adults between the ages of 18 and 65 on January 1, 2007, who reside in Greene County, Ohio; this ideal sampling frame is simply a list of all the members of the target population. If there is a difference between the target population and the study population, then the list of the people who will actually be sampled is called a ''notional sampling frame.'' The sampling units (or sampling elements) are the individual entities that are the focus of the study; in this example, these are the individual adults who meet these criteria. A sample is a subset of those adults from which observations are actually obtained and from which conclusions about the target population will be drawn. Of course, in practice such lists frequently do not exist or are incomplete, and other methods must be used to define and sample the study population.

Some other examples of studies and their populations include the following: (1) If researchers want to conduct a survey to estimate the number of people living in California who have never visited a dentist, the population consists of all people living in California, and each person is a sampling unit; (2) if researchers want to determine the number of hospital discharges during a given year that left against medical advice, then each hospital discharge is a sampling unit, and all discharges during the year are the population; and (3) if researchers want to know the number of people living in Montana who had scarlet fever in 2006, then the population is all residents of Montana in 2006, and each person is a sampling unit.

—Stacie Ezelle Taylor

See also Bias; Convenience Sample; Probability Sample; Sampling Techniques; Study Design

Further Readings

Hassard, T. H. (1991). *Understanding biostatistics.* St. Louis, MO: Mosby-Year Book.

Levy, P. S., & Lemeshow, S. (1999). *Sampling of populations: Methods and applications* (3rd ed.). New York: Wiley.

Teratogen

Teratology is the study of the effects of exposures during pregnancy on a developing fetus. These exposures, known as teratogens, can be quite varied and include agents such as medications, illicit drugs, infectious diseases, maternal metabolic states, and occupational and environmental exposures. A teratogen can cause a spontaneous loss of pregnancy or structural and/or functional disability in a child. It has been estimated that 5% to 10% of birth defects are due to an exposure during pregnancy.

The following are the five characteristics of a teratogen. The first characteristic is that the occurrence of the birth defect or pattern of birth defects is higher in the population exposed to the teratogen as compared with the general population. Since 3% to 5% of all newborns have a birth defect, the number of malformed infants must exceed that of the background risk. More specifically, the occurrence of the exact malformation or pattern of malformations must be increased. The second characteristic is that an animal model should duplicate the effect seen in humans. Animal models serve as a good system for "red flagging" an agent but can never be used to directly determine effects from human exposure or the magnitude of any potential risk. The third characteristic is that a dose-response relationship has been established; the greater the exposure, the more severe the phenotypic effect. A corresponding concept is that of a threshold effect; effects are only seen above a specific exposure level. The fourth characteristic is that there should be a plausible biologic explanation for the mechanism of action. The fifth characteristic asserts that a genetic susceptibility increases the chance of an adverse outcome from the exposure. This area of pharmacogenetics holds great promise for advancing our understanding of human teratology and the provision of individualized risk assessments.

Using the above principles, it is possible to develop a risk assessment for an individual exposed pregnancy. Several pieces of information, including the timing of the exposure, the dose of the agent, and family medical and pregnancy history information, are critical. A review of the available literature is essential. Scientific data concerning outcomes of exposed pregnancies are often conflicting, difficult to locate, and hard to interpret. Much of the data are in the form of case reports, animal studies, and retrospective reviews. To provide complete information, it may be necessary to consult various resources. It is important to appreciate the risk-benefit ratio regarding a particular agent to provide an individualized risk assessment on which pregnancy management may be based.

Despite scientific advances in clinical teratology, exposures prior to and during pregnancy still cause great anxiety and misunderstanding among both the public and health care professionals. Teratology Information Services (TIS) are comprehensive, multidisciplinary resources that provide information on prenatal exposures to health care providers and the public. The national consortium of individuals providing these services is the Organization of Teratology Information Specialists. An individual TIS has three components: service (toll-free, confidential phone consultations), education (to health care providers and the public), and research (national and international studies on specific agents).

—Dee Quinn

See also Birth Defects; Environmental and Occupational Epidemiology; Maternal and Child Health

Further Readings

Briggs, G. B., Freeman, R. K., & Yaffe, S. J. (2005). *Drugs in pregnancy and lactation: A reference guide to fetal and neonatal risk* (7th ed.). Philadelphia: Lippincott Williams & Wilkins.

Web Sites

Clinical Teratology Web: A resource guide for clinicians: http://depts.washington.edu/~terisweb.

Organization of Teratology Information Specialists, which provides further information on the individual programs, research projects, and fact sheets: http://www.otispregnancy.org.

Reprotox: An information system on environmental hazards to human reproduction and development: http://www.reprotox.org.

Terrorism

See War

Thalidomide

Thalidomide is a pharmaceutical product that was synthesized in West Germany in 1953 and sold as an antinausea drug and sleep aid under a number of different brand names beginning in 1957. Because it was believed to be nontoxic and to have no side effects, it was widely prescribed to pregnant women for relief of morning sickness and insomnia. However, thalidomide proved to be anything but nontoxic; more than 10,000 women who took the drug during pregnancy gave birth to children with severe birth defects. The best-known sign of prenatal thalidomide exposure was phocomelia (misshapen limbs), but children exposed to thalidomide before birth (commonly referred to as "thalidomide babies") suffered many other birth defects, including missing limbs, cleft palate, spinal cord defects, missing or abnormal external ears, and abnormalities of the heart, kidneys, genitals, and digestive system. Some women who took thalidomide also reported abnormal symptoms, including peripheral neuropathy. Thalidomide was removed from the market in most countries in 1961, and the events surrounding its approval and release are considered perhaps the worst case of insufficient pharmacological oversight in the modern world. In particular, thalidomide had not been tested on humans at the time of its release, and its pharmacological effects were poorly understood.

Thalidomide was never approved by the Food and Drug Administration (FDA) for sale in the United States, so the impact of the drug was much less in this country than in Europe and other countries. However, the experience of seeing thousands of "thalidomide babies" born in countries where the drug had been approved for general sale led to strengthening of several protections in the U.S. drug approval process. The major changes were incorporated in the Kefauver-Harris Amendment, passed in 1962, which required that new pharmaceutical products had to be demonstrated as both safe and effective and required that adverse reactions to prescription drugs be reported to the FDA.

Although thalidomide should not be taken by pregnant women, it has legitimate medical uses and is used in some countries to treat serious conditions such as cancer, leprosy, and AIDS. Because of thalidomide's history, any use of the drug today is controversial and some medical professionals believe that the drug should be banned entirely, while others feel that it is the best available drug to treat certain specific conditions.

One primary medical use of thalidomide today is in the treatment of leprosy, in particular to treat the symptoms of erythema nodosum leprosum. Thalidomide is also used in some cancer therapies and has become a common treatment for multiple myeloma. Thalidomide is also used to treat AIDS patients, in particular to fight mouth ulcers and wasting syndrome. Theoretically, the therapeutic use of thalidomide is carefully supervised and monitored; in practice, however, the risk of improper use remains (e.g., a thalidomide baby was born in Brazil in 1995), and this potential harm must be weighed against the benefits achieved by wider use of this drug.

—Sarah Boslaugh

See also Birth Defects; Teratogen

Further Readings

Franks, M. E., Macpherson, G. R., & Figg, W. D. (2004). Thalidomide. *Lancet*, *363*(9423), 1802–1811.

Pannikar, V. (2006). *The return of thalidomide: New uses and renewed concerns*. Retrieved August 8, 2007, from http://www.who.int/lep/research/Thalidomide.pdf.

Stephen, T. D., & Rock, B. (2001). *Dark remedy: The impact of thalidomide and its revival as a vital medicine.* New York: Perseus.

Theory of Planned Behavior

An extension of the theory of reasoned action, the theory of planned behavior (TPB) is today one of the most popular models for explaining, predicting, and changing human social behavior. It has been applied to study a number of health behaviors, including exercise, smoking, drug use, and compliance with medical regimens. According to the TPB, human behavior is guided by three kinds of considerations:

1. Beliefs about the likely outcomes of the behavior and the evaluations of these outcomes (*behavioral beliefs*); in their aggregate, these beliefs produce a positive or negative attitude toward the behavior.
2. Beliefs about the normative expectations of important others and motivation to comply with these

expectations (*normative beliefs*) that result in perceived social pressure or a subjective norm.

3. Beliefs about the presence of various internal and external factors and the perceived power of these factors to facilitate or impede performance of the behavior (*control beliefs*). Collectively, control beliefs give rise to a sense of self-efficacy or perceived behavioral control.

Attitudes toward the behavior, subjective norms, and perceived behavioral control jointly lead to the formation of a behavioral intention. The relative weight or importance of each of these determinants of intention can vary from behavior to behavior and from population to population. However, as a general rule, the more favorable the attitude and subjective norm, and the greater the perceived control, the stronger the person's intention to perform the behavior in question. Finally, given a sufficient degree of actual control over the behavior, people are expected to carry out their intentions when the opportunity arises. Intention is thus assumed to be the immediate antecedent of behavior. However, because many behaviors pose difficulties of execution, the TPB stipulates that degree of control moderates the effect of intention on behavior: Intentions are expected to result in corresponding behavior to the extent that the individual has volitional control over performance of the behavior.

Beliefs serve a crucial function in the TPB; they represent the information people have about the behavior, and it is this information that ultimately guides their behavioral decisions. According to the TPB, human social behavior is reasoned or planned in the sense that people take into account the behavior's likely consequences, the normative expectations of important social referents, and factors that may facilitate or impede performance. Although the beliefs people hold may be unfounded or biased, their attitudes, subjective norms, and perceptions of behavioral control are thought to follow reasonably from their readily accessible beliefs, to produce a corresponding behavioral intention, and finally to result in behavior that is consistent with the overall tenor of the beliefs. However, this should not be taken to imply deliberate, effortful retrieval of information and construction of intention prior to every behavior. After a person has at least minimal experience with a behavior, his or her attitude, subjective norm, and perceived behavioral control are assumed to be available automatically and to spontaneously produce a behavioral intention.

In sum, the behavioral, normative, and control beliefs that are readily accessible in memory serve as the fundamental explanatory constructs in the TPB. Examination of accessible beliefs provides substantive information about the considerations that guide people's behavior and can thus also serve as the basis for interventions designed to change behavior.

Empirical Support

The TPB has been used to predict and explain a myriad of social behaviors, including investment decisions, dropping out of high school, blood donation, driving violations, recycling, class attendance, voting in elections, extramarital affairs, antinuclear activism, playing basketball, choice of travel mode, and a host of other activities related to protection of the environment, crime, recreation, education, politics, religion, and virtually any imaginable area of human endeavor. Its most intense application, however, has been in the health domain, where it has been used to predict and explain varied behaviors such as drinking, smoking, drug use, exercising, dental care, fat consumption, breast self-examination, condom use, weight loss, infant sugar intake, getting medical checkups, using dental floss, and compliance with medical regimens.

The results of these investigations have, by and large, confirmed the theory's structure and predictive validity. Armitage and Conner (2001) found, in a meta-analytic review of 185 data sets, that the theory accounted on average for 39% of the variance in intentions, with all three predictors—attitude toward the behavior, subjective norm, and perceived behavioral control—making independent contributions to the prediction. Similarly, intentions and perceptions of behavioral control were found to explain 27% of the behavioral variance. Godin and Kok (1996), in a meta-analysis of 76 studies in the health domain related to addiction, clinical screening, driving, eating, exercising, AIDS, and oral hygiene, found that the TPB was shown to explain, on average, 41% of the variance in intentions and 34% of the variance in behavior.

Sufficiency

Investigators have suggested a number of additional variables that might be incorporated into the theory to

improve prediction of intentions and behavior. Among the proposed additions are expectation, desire, and need; affect and anticipated regret; personal or moral norm; descriptive norm; self-identity; and past behavior and habit. Some of these proposed additions can be viewed as expansions of the theory's existing components. Thus, it is possible to subsume expectation, desire, and need to perform the behavior under intention; anticipated regret and other expected affective consequences of a behavior arguably are a proper part of attitude toward the behavior; and descriptive norms perhaps contribute to perceived social pressure—that is, subjective norm.

Other proposed factors, such as self-identity and moral norms, are conceptually distinct from the original constructs in the TPB, and some studies have shown that these factors can make an independent contribution to the prediction of intentions and actions. Perhaps of greatest concern because it touches on the theory's reasoned action assumption is the suggestion that, with repeated performance, behavior habituates and is no longer controlled by intentions but, instead, by critical stimulus cues. However, evidence for the role of habit in the context of the TPB is weak; intentions are found to predict behavior well even for frequently performed behaviors that would be expected to habituate.

From Intention to Behavior

For the TPB to afford accurate prediction, intentions measured at a certain point in time must not change prior to enactment of the behavior. Empirical evidence strongly supports this expectation, showing that the intention-behavior relation declines with instability in intentions over time. Even when intentions are stable, however, people do not always act on their intentions. The concern about lack of correspondence between intentions and actions can be traced to LaPiere's classic 1934 study in which ready acceptance of a Chinese couple in hotels, motels, and restaurants contrasted sharply with stated intentions not to accept "members of the Chinese race" in these same establishments. Similar discrepancies have been revealed in investigations of health behavior where it has been found that between 25% and 50% of participants fail to carry out their stated intentions to perform behaviors such as using condoms regularly, undergoing cancer screening, or exercising. A variety of factors may be responsible for observed failures of effective self-regulation, yet a simple intervention can do much to reduce the gap between intended and actual behavior. When individuals are asked to formulate a specific plan—an implementation intention—indicating when, where, and how they will carry out the intended action, the correspondence between intended and actual behavior increases dramatically. For example, Sheeran and Orbell (2000) found that asking participants who planned to undergo a cervical cancer screening to form a specific implementation intention increased participation from 69% to 92%.

Behavioral Interventions

Given its predictive validity, the TPB can serve as a conceptual framework for interventions designed to influence intentions and behavior. Thus far, only a relatively small number of investigators have attempted to apply the theory in this fashion. The results of these attempts have been very encouraging. For example, Brubaker and Fowler (1990) developed an intervention designed to increase testicular self-examination (TSE) among high school students based on the TPB-addressed beliefs about the outcomes of TSE. This theory-based intervention was found to be considerably more successful than merely providing information about testicular cancer and TSE or a general health message. The theory-based intervention had a significantly greater impact on attitudes toward TSE, the factor directly attacked in the intervention, it was more effective in raising intentions to perform TSE, and it produced a 42% rate of compliance over a 4-week period, compared with 23% and 6% compliance rates in the other two intervention conditions, respectively. The theory has also been applied in interventions designed to promote vegetable and fruit consumption, smoking cessation, safer sex, physical exercise, and a host of other, mostly health-related behaviors; these applications are reviewed by Hardeman et al. in 2002.

—Icek Ajzen

See also Health Behavior; Health Belief Model; Intervention Studies; Self-Efficacy

Further Readings

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179–211.

Ajzen, I. (2005). *Attitudes, personality, and behavior* (2nd ed.). Maidenhead, UK: Open University Press.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behavior: A meta-analytic review. *British Journal of Social Psychology, 40*, 471–499.

Brubaker, R. G., & Fowler, C. (1990). Encouraging college males to perform testicular self-examination: Evaluation of a persuasive message based on the revised theory of reasoned action. *Journal of Applied Social Psychology, 20*, 1411–1422.

Godin, G., & Kok, G. (1996). The theory of planned behavior: A review of its applications to health-related behaviors. *American Journal of Health Promotion, 11*, 87–98.

Hardeman, W., Johnston, M., Johnston, D. W., Bonetti, D., Wareham, N. J., & Kinmonth, A. L. (2002). Application of the theory of planned behaviour in behaviour change interventions: A systematic review. *Psychology and Health, 17*, 123–158.

LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces, 13*, 230–237.

Sheeran, P., & Orbell, S. (2000). Using implementation intentions to increase attendance for cervical cancer screening. *Health Psychology, 19*, 283–289.

Time Series

Time series are time-ordered observations or measurements. A time series can consist of elements equally spaced in time—for example, annual birth counts in a city during four decades—or measurements collected at irregular periods, for example, a person's weight on consecutive visits to a doctor's office. Time plots, that is, graphical representations of time series, are very useful to provide a general view that often allows us to visualize two basic elements of time series: short-term changes and long-term or secular trends.

Time series are often used in epidemiology and public health as descriptive tools. For instance, time plots of life expectancy at birth in Armenia, Georgia, and Ukraine during the years 1970 to 2003 (see Figure 1) reveal relatively stagnant health conditions before the 1990s in these three countries of the former USSR, as well as a substantial deterioration of health

Figure 1 Life Expectancy at Birth (Years) in Three Countries Formerly Part of the USSR

Source: Created by the author using data from the European Health for All database (HFA-DB).

in the early to mid-1990s, after the breakdown of the USSR. The sharp trough in 1988 in Armenian life expectancy reflects the impact of the Spivak region earthquake that killed tens of thousands of people, including many children.

In describing a time series in mathematical terms, the observation or measurement is considered a variable, indicated for instance by y_t, where the subscript t indicates time. It is customary to set $t = 0$ for the first observation in the series, so that the entire expanded series will be represented by y_0, y_1, y_2, $y_3, \ldots, y_m$ for a series containing $m + 1$ elements. The element for time k in this series will be therefore y_k, with $k > 0$ and $k < m$.

In some cases, time series are well described by a mathematical model that can be exponential, logistic, linear (a straight line), and so on. They often reveal recurring patterns, for instance seasonal ones, associated with different seasons of the year. A series of monthly deaths due to respiratory disorders during several consecutive years will show a yearly peak in winter and a yearly trough in summer. Other patterns may be periodic but not seasonal; for instance, among Jews in Israel, deaths are more frequent on Sundays. Still other patterns are recurrent but are acyclical, that is, not periodical; for instance, in market economies, the unemployment rate reveals successive peaks (recessions) and troughs (periods of economic expansion) that make up what has come to be called the "business cycle" (Figure 2), which is not a "cycle" in the ideal sense because it occurs at irregular intervals.

In most time series, there is strong first-order autocorrelation, which means that consecutive values are highly correlated. That is, the value y_k is usually not very far from its neighbors, y_{k-1} and y_{k+1}. This is the basis for *interpolation* and *extrapolation*, the two techniques used to estimate an unknown value of a time series variable. A missing value inside a time series can be estimated by *interpolation*, which usually implies an averaging of the values in the neighborhood of the missing one. For instance, if the value for Year 8 in a series of annual values was unobserved, it can be estimated as an average of the observed values for years 6, 7, 9, and 10. In general, the error in the estimation through interpolation will

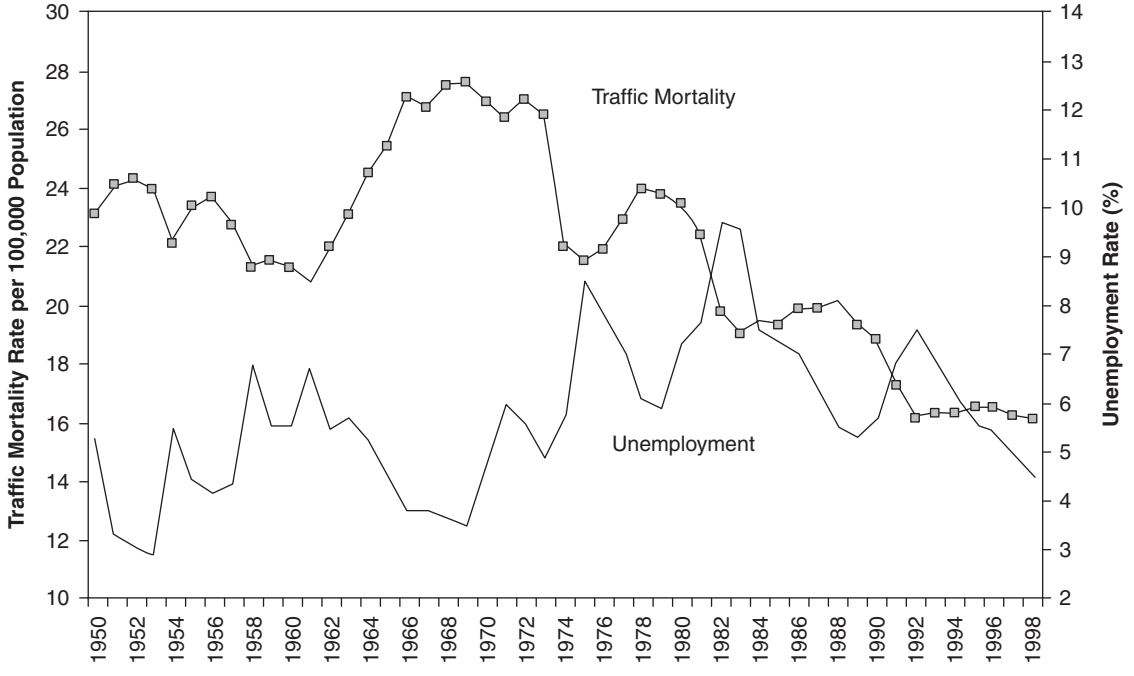


Figure 2 Traffic Mortality and Unemployment in the United States, 1950–1999

Source: Created by the author using data in the U.S. Bureau of the Census "Historical Statistics of the United States" and the National Health for Health Statistics "Vital Statistics of the United States." Available in several online and printed sources.

be smaller than the error in estimating through extrapolation, which implies estimating an unobserved value out of the time range of the time series, usually in the future. (Backward extrapolation to the past is also possible, though usually uninteresting.) *Forecasting* is a term used for predicting the value of a variable at a later time than the last one observed. When a causal model involving the major determinants of the variable to be predicted is not available, forecasting is done through more or less sophisticated techniques of extrapolation. For instance, if the suicide rates during the past 5 years in a nation were, respectively, 12.7, 13.4, 12.9, 12.2, and 13.1 per million, using a somewhat rough extrapolation we might forecast that the suicide rate next year will probably be around 12 or 13 per million. This conclusion, as any other forecast, has a large uncertainty associated with it, because time series may have sudden upturns or downturns that cannot be predicted. Obviously, the uncertainty grows exponentially as the future we try to forecast becomes more distant. The formal techniques of forecasting imply fitting a mathematical model (linear, exponential, etc.) to the observed data, then computing the expected value in the future with that mathematical model. The *a*uto*r*egressive *i*ntegrated *m*oving *a*verage models or ARIMA models, often used in forecasting (frequently referred to as the Box-Jenkins models or methodology), have been sometimes applied in biomedical sciences and epidemiology, but they constitute a quite specialized field of statistics. Like ARIMA models, spectral analysis is another specialized technique in the field of time series analysis.

Establishing causal relations between time series variables is not straightforward. While high absolute values of the correlations between variables observed in cross-sectional studies may be suggestive of causal associations, high correlations between two time series are very common and prove nothing. For instance, in many advanced countries the proportion of persons below 15 years of age in the population, the annual volume of typewriting machines sold, and the percentage of adult women not having paid work secularly dropped during the past half century. Therefore, the correlation between any two annual series of these three variables will be very high, but this does not suggest any causal relationship at all between them. To investigate causal relations between time series, the series must be *stationary*. A stationary series is one consisting of values oscillating over and above a constant mean for the whole period. Thus, a stationary series is one that has no trend. The series with trends have to be detrended or "prewhitened" if we are looking for causal associations. *Differencing* and *filtering* with a smoothing tool are common methods of detrending. For differencing a series, the differences between successive elements of the series are computed, so that the original series $y_0, y_1, y_2, \ldots, y_m$ is transformed into the series $z_0, z_1, z_2, \ldots, z_{m-1}$, where $z_0 = y_1 - y_0$, $z_1 = y_2 - y_1$, and $z_{m-1} = y_m = y_{m-1}$. The differenced series will be one unit shorter than the original one. The transformation of a variable y_t into its rate of change $r_t = (y_t - y_{t-1})/y_{t-1}$ is a variety of the procedure of differencing. Detrending a series through *filtering* involves computing a smooth trend line, and then the detrended series is computed by subtracting the values of the filtered series from the values of the original series. There are many filtering procedures or "filters," the most commonly used are moving averages (moving means), moving medians, and combinations of both, such as the T4253H smoother provided by SPSS. The Hodrick-Prescott filter is another popular tool for detrending. Since in time series analysis it is common to work with transformed values, the term *in levels* is often used to refer to the original observations (say, the annual rate of infant mortality), while "in differences" refers to their changes from one observation period to the next (the absolute change in infant mortality) and "in rate of change" refers to its period-to-period percent change (the year-to-year relative change in the infant mortality rate).

High absolute values of the cross-correlation of two stationary time series imply a strong comovement and, possibly, a causal connection, with one series causing the other or both being caused by a third variable. The graphs showing parallel movements in two series (with peaks in one coinciding with peaks in the other, and troughs in one coinciding with troughs in the other) are evidences of comovement and highly suggestive of some direct causal connection. The same is true when graphs show mirroring movements (with peaks in one series coinciding with troughs in the other, and vice versa). For example, mirroring time plots of unemployment and mortality due to traffic injuries (Figure 2) are highly suggestive of traffic deaths increasing during periods of accelerated economic activity, when unemployment diminishes, and dropping during recessions, when unemployment is high.

Panel regression methods are often useful to analyze relationships between time series variables.

—José A. Tapia Granados

See also Causation and Causal Inference; Data Transformation; Panel Data; Pearson's Correlation Coefficient; Regression

Further Readings

Diggle, P. J. (2000). *Time series: A biostatistical introduction*. Oxford, UK: Clarendon Press.

Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L., et al. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of American Medical Association, 295*(2), 1127–1134.

Pope, C. A., III, & Schwartz, J. (1996). Time series for the analysis of pulmonary health data. *American Journal of Respiratory Care Medicine, 154*, 5229–5233.

Tapia Granados, J. A. (2005). Increasing mortality during the expansions of the US economy, 1900–1996. *International Journal of Epidemiology, 34*, 1194–1202.

Warner, R. M. (1998). *Spectral analysis of time-series data*. New York: Guilford Press.

Zeger, S. L., Irizarry, R., & Peng, R. D. (2006). On time series analysis of public health and biomedical data. *Annual Review of Public Health, 27*, 57–79.

Web Sites

The European Health for All database (HFA-DB), compiled by the Regional Office for Europe of the World Health Organization: http://www.euro.who.int/hfadb.

Tobacco

The term *tobacco* refers to plants of the genus *Nicotiana*, which may be consumed in various ways. Because cigarette smoking is the predominant method of tobacco consumption in the United States, in public health the term *tobacco use* is often used as a synonym for *cigarette smoking* without consideration of the different modes of tobacco consumption and differing health risks posed by them. This entry describes the health risks associated with cigarette smoking, other tobacco smoking, and environmental tobacco smoke (ETS) and contrasts these to the effect of nicotine in itself and to the use of smokeless tobacco (ST). It also explores the importance of tobacco research in the history of epidemiology and the potential of epidemiological studies in reducing the health impacts of smoking.

Tobacco is native to the Americas, where it was cultivated by indigenous populations from about 4000 BCE and used in smoked and smokeless forms, largely for ceremonial purposes. Less than 100 years after its discovery by European explorers around AD 1500, tobacco was being used throughout the world, a testimony to the powerful psychoactive properties it delivers to human brains. Cigarette smoking started to become popular only around 1900 with the introduction of efficient mass production and wide distribution of cigarettes during the 20th century's two world wars made it the dominant form of tobacco use globally.

Ochsner and DeBakey recognized a link between smoking and lung cancer as early as 1939. Schairer and Schöniger (1943) published one of the first epidemiologic studies of this relationship, in German, during World War II (it was not widely distributed or indexed at the time, but was resurrected in English in 2001). But it was not until the studies of cigarette smoking and lung cancer by Doll and Hill (1950), Wynder and Graham (1950), and others a half century ago that the dangers of smoking were clearly established. While difficult to imagine today, the medical community that then dominated public health was sufficiently conservative that these results were not immediately accepted despite previous evidence. These early studies of the relationship between smoking and health risks also played an important role in establishing the merits of observational epidemiologic studies.

Today, it is well established that regular moderate or heavy cigarette smoking (and to a lesser extent, smoking of tobacco in other forms) causes well-known morbidity and mortality risks, with total attributable risk far exceeding that from any other voluntary exposure in wealthy countries. Cigarette smoke also creates an environmental exposure, labeled "second-hand smoke" or "environmental tobacco smoke" (ETS). Because smoking has been so prevalent for so long and causes high relative risks for many diseases, and because it offers little opportunity for experimental intervention, smoking stands as a near-perfect demonstration of what can be done with observational epidemiology (though perhaps also as a rarely attainable archetype).

In contrast, nicotine, the primary reason people smoke or otherwise use tobacco, is a relatively benign

mild stimulant, similar to caffeine. Nicotine causes transient changes in cardiovascular physiology, as do many mild stimulants that might cause a small risk of cardiovascular disease. There is general agreement that nicotine is addictive for many people (the term *addictive* is not well-defined, but nicotine consumption fits most proposed definitions, at least for a portion of the population). But nicotine by itself does not appear to cause a substantial risk of any life-threatening disease; the epidemiologic evidence on nicotine in the absence of smoking is sufficiently limited that it is impossible to distinguish small risks from zero risk. Although not extensively studied, research suggests that nicotine may have psychological and neurological health benefits, protecting against Parkinson's disease and possibly dementia, and providing acute relief from schizophrenia and other psychological morbidities.

Substantial research shows that the use of modern ST products is associated with very small health risks, similar to those from nicotine alone. There has been little research on the health effects of very light smoking or long-term pharmaceutical nicotine use, in part because it is difficult to find populations with such long-term usage patterns (not interrupted by periods of heavier smoking), and in part because most tobacco and nicotine research funding is driven by a prohibitionist agenda, and so there is limited support for quantifying these practices' presumably modest health effects.

Cigarette smoking probably remains the most researched exposure in epidemiology. However, the set of exposures related to tobacco also generate a great deal of advocacy and rhetoric, often making it a challenge to sort out the epidemiology from the politics. Epidemiology related to tobacco suffers from publication bias against studies that show no increased risk (which is particularly relevant to harm reduction and to ST), from overinterpretation of results of a few favored studies, and from a "ratchet effect," where any association found in one study is treated as established, regardless of what other evidence shows. For example, many studies of ETS have shown very small or undetectable health effects for all but extreme exposure levels, but these studies are widely ignored, or even vilified, in the popular discourse. Similarly, a few studies have found positive associations of ST use and oral or pancreatic cancer, but most studies have not; nevertheless, these positive associations are discussed as if they are indisputably established.

Perhaps these problems are no worse than in other subject matter areas, but they pose a potentially greater threat to epidemiology as an honest science because of the high stakes and high profile of tobacco issues, and are less excusable given the overwhelming amount of epidemiologic evidence that exists.

The greatest confusion comes from treating exposures to tobacco as homogeneous, despite the very different pathways and different levels of health risk. Using the term *tobacco* is particularly misleading when referring only to the health effects of smoking, since the major health impact is from inhaling smoke, which is quite unhealthy no matter what is burning; thus, emphasizing the plant rather than using the term *smoking* confuses people about the cause of the health effects.

Cigarette Smoking

Smoking prevalence peaked at about 50% in most Western countries, reaching a maximum in the 1950s and 1960s in most male populations, though often continuing to rise among women. But the health risks of smoking, highlighted in reports from the Royal College of Physicians (the United Kingdom, 1962) and the United States Surgeon General (1964), and in thousands of studies since, resulted in a steady decline over about two decades, to the prevalences in the 20-some percent range (similar for men and women) found in most Western countries today. However, despite near-universal knowledge of the health risks and aggressive antismoking advocacy and policies in many places, the rate of the decline has slowed or stopped over the past several decades. National average prevalence has dropped substantially less than 20% only in Sweden, where ST use has largely replaced smoking. Outside the West, prevalence is increasing in many countries; male prevalence remains more than 50% in many countries in Eastern Europe, the former Soviet Union, Africa, and Asia, while prevalence among women varies from negligible to quite high.

Since the lung cancer link was established, smoking has been shown to cause other cancer mortality and an even greater absolute risk of fatal cardiovascular disease. Popular claims attribute about one fifth of all current mortality in wealthy countries to smoking or in excess of 150 deaths per 100,000 person years. Extrapolations of present worldwide trends predict dramatically increasing smoking-attributable mortality in the future, predominantly in developing countries.

It should be noted that some of the most widely cited statistics about tobacco and health are produced primarily by antitobacco advocates using proprietary data and methods, and thus cannot be validated. For example, estimates of smoking-attributable deaths released by the U.S. Centers for Disease Control and Prevention (CDC) are based on relative risks derived from the American Cancer Society's (ACS) Second Cancer Prevention Study (CPS-II). Nearly everyone has heard of the CDC estimate of about 400,000 annual smoking-attributable deaths in the United States. But few realize that this and other findings from the CDC relating to health consequences of tobacco use, the basis of tobacco policies at all levels of American government, are based on data and analyses that are kept secret from investigators outside the CDC or ACS.

However, few would doubt that the true mortality from smoking is at least half of what is usually claimed, so there is no serious question that among behavioral health exposures, smoking is among the most important at the individual and social levels. In the world's healthier countries, it has a greater impact on mortality and morbidity than any other behavioral exposure. Smoking is often called the greatest or most important preventable source of disease; while such phrasing belongs to advocacy rhetoric and is scientifically meaningless (most notably, it strains the definition of "preventable" to apply it to an exposure that remains very prevalent despite massive efforts to eliminate it), the epidemiologic evidence makes clear that if we could substantially reduce the rate of smoking, it would result in greater health improvement in wealthy countries than any other change imaginable within the bounds of current technology and budgets.

By exposing the lungs, airway, and mouth to concentrated combustion products, smoking causes a still-increasing majority of the world's lung cancer. In Western men, smoking is estimated to cause as much as 90% of lung cancer and 75% of the oral, pharyngeal, esophageal, and laryngeal cancers; attributable risk for women, historically lower due to a lag in smoking uptake in the 20th century, is largely equivalent today. Smoking has also been convincingly linked to cancers of the stomach, pancreas, and urinary bladder, as well as leukemia. It is sometimes also linked to cancers of the breast and colon, but these associations are less well established. Smoking is responsible for reversing what otherwise would have been a steep decline in overall cancer mortality in Western countries during the latter half of the 20th century.

The relative risks for cardiovascular diseases are much lower than those for the sentinel cancers, but because of the greater baseline risk, the absolute total risk is higher. In the West, smoking is estimated to cause about 40% of coronary heart disease and stroke deaths in people below 65 years of age and more than 50% of deaths from aortic aneurysms. In addition, smoking is considered the proximate cause of about 20% of pneumonia and influenza deaths and about 80% of deaths related to bronchitis, emphysema, and chronic airway obstruction.

Other Tobacco Smoking

Smoking of tobacco in various types of pipes and cigars is an exposure similar to smoking cigarettes, though many (but not all) smokers of these products have lower consumption and do not draw smoke into the lungs, both of which result in lower risks. Because of the great heterogeneity of usage patterns, it is difficult to generalize about these exposures. But epidemiologic studies generally show these exposures, as practiced in the West, to cause substantially less risk than regular cigarette smoking on average, though the total risk of serious disease associated with their use is still high compared with almost every other common voluntary exposure.

Environmental Tobacco Smoke

There is fierce debate about the magnitude of the health risk to nonsmokers from ETS exposure. ETS has been linked to various acute changes in respiratory and cardiovascular physiology, but the epidemiologic evidence is only suggestive of a small risk of lung cancer and cardiovascular disease after concentrated long-term exposure such as that experienced by the nonsmoking spouses of smokers or by people who work in very smoky environments. Popular claims attribute that about 2 deaths per 100,000 person years to ETS in wealthy countries, with claimed relative risks for lung cancer and heart disease as high as about 1.3, but these numbers come from antismoking advocates and selective citation of the research, and are not widely accepted by nonadvocates. For example, the scientific literature contains competing summary analyses of studies of ETS and cardiovascular disease, with a widely cited study written by employees of an

antitobacco advocacy group finding a relative risk in excess of 1.2, while a recent study of that literature produces a summary estimate of approximately 1.05. While it stands to reason that ETS creates some of the same risks as active smoking (since it involves exposure to the same chemicals that harm smokers, via the same pathway, albeit in much lower doses), the absolute risk appears to be lower than what can be accurately measured by available epidemiologic methods.

Smokeless Tobacco

Since most of the health risk from smoking comes not from the tobacco plant, but from inhaling concentrated smoke, oral use of modern Western ST products (e.g., snuff dipping) has little in common with smoking other than nicotine absorption. This exposure has become popular in Sweden and Norway, and seems to be gaining popularity in parts of North America, due in part to the low health risks and to availability of modern products that can be used invisibly and without spitting, in contrast to traditional chewing tobacco.

The epidemiologic evidence does not definitively demonstrate an association between ST use and any life-threatening disease. There is a widespread misunderstanding, among both health professionals and the general population, that ST use creates substantial risk of oral cancer, but this is based on erroneous conclusions from early research. Extensive modern epidemiology has consistently shown that ST use causes very little or no risk of oral cancer (clearly much less than the substantial risk of oral cancer from smoking) or of any other life-threatening disease.

In contrast to studies of smoking, epidemiologic studies of ST use face considerable challenges because the prevalence of ST use in Western countries is very low (e.g., no more than 5% among adult men and well under 1% among women in the United States), the diseases putatively linked to ST use (such as oral cancer) are rare, and the relative risks, even among long-term users, are very low. Despite these challenges, there has been sufficient epidemiologic research on the subject, most usefully from the past 15 years, to conclude that Western ST use causes only a tiny fraction of the total mortality risk of smoking; calculated estimates put it at 1% to 2% and clearly less than 5%. A recent meta-analysis of epidemiologic studies of ST use and oral cancer found that the use of modern American and Swedish products (moist snuff and chewing tobacco) was associated with undetectably low risks for cancers of the oral cavity and other upper respiratory sites (relative risks ranging from 0.6 to 1.7); older studies of American dry snuff showed substantially elevated risk (relative risks ranging from 4 to 13), with the contrast due to an unknown combination of the archaic products causing measurable risk and improved study methods (e.g., better control for smoking).

ST use in South Asia and Africa may cause substantially greater disease risks. The products used are quite different from Western ST, because they use different manufacturing processes and typically include other ingredients that have their own psychoactive and health effects (indeed, sometimes these products do not even contain tobacco, but are classified in analyses as being tobacco products). The epidemiology suggests that these products are associated with a substantially increased risk of oral cancer, with relative risks for this disease similar to or higher than those from smoking. Since oral cancer is much more common outside the West, this represents a greater absolute risk than it would in the West. Little is known about other mortality risks from these products, though there is no reason to doubt that total risk is greater than that from Western ST, but still only a small fraction of that from smoking.

Epidemiology and Reducing the Health Impacts of Smoking

Beyond showing that smoking is unhealthy, epidemiologic research also contributes to identifying predictors of smoking behavior, assessing smoking cessation interventions (generally finding them to provide very little or no benefit), and measuring the effects of antismoking regulations. Important unanswered epidemiologic questions with practical implications for health policy include the health effects of very low levels of smoking (in the range of one cigarette per day), the nature of the benefits of nicotine for some users and its effect on their quality of life, and whether smokers derive important benefits from smoking apart from the nicotine.

Epidemiologic research has revealed the potential of tobacco harm reduction (the substitution of less harmful sources of nicotine for smoking) as an important public health intervention. The effectiveness of traditional antismoking efforts has plateaued in the Western world. But since other products (particularly ST, which has similar pharmacokinetics to smoking)

contain the nicotine that smokers seek, and those products have been shown to cause very little of the health risk associated with smoking, encouraging smokers to switch products is a promising intervention. Swedish men substantially replaced smoking with ST use over the past several decades, and descriptive epidemiology confirmed that the predicted reduction in disease occurred. Swedish women and Norwegians are making a similar substitution, and the approach is increasingly considered a promising option in North America and elsewhere.

—Carl V. Phillips and Brad Rodu

See also Cancer; Doll, Richard; Harm Reduction; Health Behavior; Hill, Austin Bradford; Observational Studies

Further Readings

Doll, W. R., & Hill, A. B. (1950). Smoking and carcinoma of the lung: Preliminary report. *British Medical Journal, 2*, 739–748.

Doll, R., & Hill, A. B. (1956). Lung cancer and other causes of death in relation to smoking: A second report on the mortality of British doctors. *British Medical Journal, 2*, 1071–1081.

Gately, I. (2001). *Tobacco: A cultural history of how an exotic plant seduced civilization*. New York: Grove Press.

Office of the Surgeon General. (2004). *The health consequences of smoking: A report of the Surgeon General*. Atlanta, GA: National Center for Chronic Disease Prevention and Health Promotion.

Ochsner, A., & DeBakey, M. (1939). Primary pulmonary malignancy: Treatement by total pneumonectomy. Analysis of 79 collected cases and presentation of 7 personal cases. *Surgery, Gynecology & Obstetrics, 68*, 435–451.

Phillips, C. V., Rabiu, D., & Rodu, B. (2006). Calculating the comparative mortality risk from smokeless tobacco versus smoking. *American Journal of Epidemiology, 163*, S189. Retrieved August 8, 2007, from http://www.tobaccoharmreduction.org/calculating.htm.

Phillips, C. V., Rodu, B., & the Alberta Smokeless Tobacco Education & Research Group, University of Alberta. (2006). *Frequently asked questions about tobacco harm reduction*. Retrieved August 8, 2007, from http://www.tobaccoharmreduction.org/faq/menu.htm.

Rodu, B., & Godshall, W. T. (2006). Tobacco harm reduction: An alternative cessation strategy for inveterate smokers. *Harm Reduction Journal, 3*, 37.

Royal College of Physicians of London. (1962). *Smoking and health: Summary and report of the Royal College of Physicians of London on smoking in relation to cancer of the lung and other diseases*. Toronto, Ontario, Canada: McClelland & Stewart.

Schairer, E., & Schöniger, E. (1943). Lungenkrebs und Tabakverbrauch [Lung cancer and tobacco consumption]. *Zeitschrift für Krebsforschung*, *34*, 261–269. (Republished in English: *International Journal of Epidemiology* (2001), *30*, 24–27)

Surgeon General's Advisory Committee on Smoking and Health. (1964). *Smoking and health: Report of the advisory committee to the Surgeon General of the Public Health Service*. Washington, DC: U.S. Department of Health, Education, and Welfare, Public Health Service.

Wynder, E. L., & Graham, E. A. (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma: A study of 684 proved cases. *Journal of the American Medical Association, 143*, 329–336.

Toxic Shock Syndrome

Toxic shock syndrome (TSS) is a rare but potentially fatal disease caused by toxins produced by two types of bacteria. It is most commonly associated with tampon use but has also been linked to the use of contraceptive diaphragms, wound infections, complications following surgery, and infection resulting from childbirth or abortion.

TSS is caused by the release of toxins from the strains of bacteria, *Staphylococcus aureus*, and less commonly, *Streptococcus pyogenes.* Infections caused by the latter strain are called streptococcal toxic shock-like syndrome (STSS), and although it is a similar syndrome to TSS, it is not identical. The median incubation period of infection of TSS is approximately 2 days.

Symptoms of TSS infection can develop very suddenly and typically include fever, nausea, diarrhea, vomiting, and muscle aches. A sunburn-like rash on the palms and the soles is typically present during the acute phase and peels within a few weeks. More serious complications include hypotension and sometimes even multiorgan failure. Infection is subsequently diagnosed with tests that may include blood and urine tests. On confirmation of diagnosis, treatment typically involves the administration of antibiotics, and in general, the patient recovers in approximately 7 to 10 days. In more serious cases, treatment may include hospitalization and administration of intravenous fluids.

TSS was first described in 1978 in the United States in an outbreak of seven young children. However, it became more commonly known in 1980 as a result of an epidemic associated with the prolonged use of highly absorbent tampons in menstruating, healthy,

young women. This association was due to the efficiency of superabsorbent tampons in absorbing magnesium, low levels of which are associated with increased production of TSS-associated toxin, TSS Toxin 1. After this initial epidemic, TSS became a nationally reportable disease in the United States in 1980.

Following this epidemic, the number of cases of TSS has declined significantly. Influencing factors might include changes that were made in tampon production that led to a decrease in tampon absorbency, greater knowledge of TSS among women and physicians, and the standardized labeling required by the U.S. Food and Drug Administration. Specifically, superabsorbent tampons were removed from the market after the outbreak in 1980. In 1979, before these tampons were removed from the market, menstrual TSS accounted for approximately 90% of all cases. By 1996, it accounted for approximately half of all cases. The annual incidence rate when the last surveillance was done in 1986 was approximately 1 per 100,000 women. It is fatal in about 5% of all cases.

An additional change in the epidemiology of TSS since this time is the relative increase in the proportion of nonmenstrual cases, particularly those reported following surgical procedures. This could be due to an increase in outpatient procedures and therefore increased opportunity for infection. Preventive efforts focus on patient education about early signs and symptoms and risk factors for TSS.

—Kate Bassil

See also Food and Drug Administration; Notifiable Disease; Women's Health Issues

Further Readings

Hajjeh, R. A., Reingold, A., Weil, A., Shutt, K., Schuchat, A., & Perkins, B. A. (1999). Toxic shock syndrome in the United States: Surveillance update, 1979–1996. *Emerging Infectious Diseases, 5*(6), 807–810.

Transtheoretical Model

The transtheoretical model (TTM) was developed by James Prochaska and Carlo DiClemente around 1980 to explain how people change in psychotherapy. The model was soon adapted to describe behavior change with respect to addictions, especially smoking cessation. In the past 10 years, the model has been applied across a wide range of behaviors important to public health, including diet, exercise, sun exposure, alcohol and drug abuse, mammography screening, condom use, stress management, weight control, diabetes self-management, and many more. First conceptualized primarily as a model of self-change, the model was elaborated to include how people change with professional help and has now become one of the most widely used frameworks for the development and dissemination of public health interventions.

The basic premise of the TTM is that behavior change occurs in a series of *stages of change* and that at each stage different strategies or *processes of change* are best suited to help individuals change behavior. The model is frequently referred to as the stages of change model; however, that name overlooks several important additional constructs that are integral to the change process, including several intervening or intermediate outcome variables: *decisional balance* (the pros and cons of change) and *self-efficacy* (confidence in the ability to change and temptations to engage in unhealthy behaviors across challenging situations). Together with the stages and processes of change, these constructs provide a multidimensional view of how people change.

Stages of Change

The stages of change serve as the central organizing construct of the TTM, describing change as a process instead of a singular event. Five ordered categories of readiness to change have been defined: precontemplation, contemplation, preparation, action, and maintenance.

Precontemplation

Individuals in the precontemplation stage do not intend to change in the next 6 months. People in this stage may think that their unhealthy behavior is not a real or serious problem for many reasons. They may avoid thinking, reading, or talking about their behavior, and may seem resistant, defensive, and unmotivated.

Contemplation

In the contemplation stage, individuals admit that their behavior is a problem and they are seriously considering change within the next 6 months. They

acknowledge the benefits of change but are keenly aware of the costs, resulting in ambivalence. These individuals often delay acting on their intentions and may remain in this stage for a long time ("chronic contemplation").

Preparation

Individuals in the preparation stage intend to change behavior in the next 30 days. They have a specific plan of action that includes small steps forward, such as smoking fewer cigarettes, quitting smoking for 24 hr, enrolling in an online program, or talking to a health professional. Often, they have made recent short-term attempts to change behavior.

Action

Individuals in the action stage must meet a specific and well-established behavioral criterion, such as quitting smoking (rather than cutting down), or eating five or more daily servings of fruits and vegetables (rather than eating more in one serving). Ideally, the criterion reflects expert consensus on how much change is necessary to promote health or reduce disease risk. The action stage lasts for 6 months, since this includes the period of greatest relapse risk.

Maintenance

Maintenance is defined as 6 months of successful action. The goals for this stage are to consolidate the gains achieved during action so as to continue to prevent relapse. While relapse risk diminishes during maintenance, it does not disappear. For some individuals and for some behaviors, maintenance may be a lifelong struggle.

Generally, individuals need to complete the tasks and consolidate the gains of one stage before they are ready to progress to the next. Progress through the stages is not usually linear, but more likely to be cyclical. Individuals reaching action or maintenance may lapse and recycle to an earlier stage. Once included as a distinct stage in the model, relapse is viewed as an event that initiates recycling through the stages. Most relapses do not result in regression all the way back to precontemplation since many of the gains made before the relapse remain, thus facilitating subsequent action attempts. The TTM views relapse as providing opportunities to learn from previous mistakes, to weed out unsuccessful change strategies, and to try new approaches.

Processes of Change

The processes of change are cognitive, affective, and behavioral strategies that individuals use to progress through the stages of change. Ten basic processes have been found consistently across most health behaviors (several additional processes have been identified as important for a more limited set of behaviors). These processes are organized into two higher-order processes. The experiential processes incorporate cognitive and affective aspects of change, and the behavioral processes include more observable change strategies. The experiential processes include *consciousness raising* (increasing awareness and understanding of the behavior), *dramatic relief* (experiencing feelings of personal susceptibility related to the behavior), *environmental reevaluation* (affective and cognitive understanding of how the behavior affects the psychosocial environment), *self-reevaluation* (cognitive/affective understanding of personal values and self-image with respect to behavior), and *social liberation* (awareness of social norms and support for alternative, healthier choices). The behavioral processes include *contingency (reinforcement) management* (rewarding oneself or being rewarded by others for making healthy changes), *counterconditioning* (substitution of alternative healthier behaviors for unhealthy ones), *helping relationships* (accepting and using others' support), *self-liberation* (choosing and committing to change), and *stimulus control* (removal of cues for unhealthy behaviors, addition of cues for healthy alternatives, avoiding challenging situations, and seeking supportive environments).

Decisional Balance: Pros and Cons of Change

Part of the decision to move from one stage to the next is based on the relative evaluation of the pros and cons of changing behavior. The pros are positive aspects or advantages of change, and the cons are negative aspects or disadvantages of change. The comparative weight of the pros and cons varies depending on the individual's stage of change. This relationship between the stages of change and decisional balance has been found to be remarkably consistent across a wide range of health behaviors.

Self-Efficacy: Confidence and Temptation

Adapted from cognitive-social learning theory, in the TTM self-efficacy is operationalized as how confident individuals are that they will engage in the new healthy behavior and how tempted they would be to engage in the unhealthy behavior across a range of challenging situations. Both constructs assess multidimensional situational determinants of relapse. Confidence and temptation typically show small relationships to stage of change from precontemplation to preparation, followed by strong and nearly linear increases and decreases from preparation to maintenance, respectively. Both constructs serve as good indicators of relapse risk for individuals in later stages.

Integration of TTM Constructs

TTM constructs are integrally related providing an important foundation for intervention. Transitions between stages are mediated by the use of distinct subsets of change processes and are associated with substantial changes in decision making, self-efficacy, intention, and ultimately, behavior. Individuals in the earlier stages of change tend to use experiential processes of change and report relatively low confidence and fairly high temptation, as well as overvaluing the cons of change relative to the pros. Individuals in the later stages tend to use behavioral processes, report more confidence in their ability to change and relatively less temptation to slip into unhealthy behaviors, and evaluate the pros of change more highly than the cons. These interrelationships are vital to the development of effective interventions. When treatment programs ignore or mismatch processes to stages, recruitment, retention, and behavior change efforts are likely to suffer. Stage-tailored intervention programs accelerate progress through the stages. An important advantage of this approach is that stage-tailored interventions are relevant not just to those select individuals who may be ready to change but also to the full population who may be neither prepared nor motivated to change. The TTM intervention approach, including effective treatment for individuals at all stages of readiness to change, can greatly increase the population impact of programs, by effectively increasing recruitment, retention, reach, and efficacy.

—Joseph S. Rossi and Colleen A. Redding

See also Health Belief Model; Self-Efficacy; Social-Cognitive Theory; Targeting and Tailoring; Theory of Planned Behavior

Further Readings

Prochaska, J. O. (2004). Population treatment for addictions. *Current Directions in Psychological Science*, *13*, 242–246.

Prochaska, J. O., Redding, C. A., & Evers, K. (2002). The transtheoretical model and stages of change. In K. Glanz, B. K. Rimer, & F. M. Lewis (Eds.), *Health behavior and health education: Theory, research, and practice* (3rd ed., pp. 99–120). San Francisco: Jossey-Bass.

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion, 12*, 38–48.

Prochaska, J. O., Velicer, W. F., Rossi, J. S., Goldstein, M. G., Marcus, B. H., Rakowski, W., et al. (1994). Stages of change and decisional balance for 12 problem behaviors. *Health Psychology, 13*, 39–46.

TUBERCULOSIS

Tuberculosis (TB) is a contagious disease caused by the bacilli *Mycobacterium tuberculosis*, which usually attacks the lungs but can also attack the brain, spine, and other parts of the body. TB was once the leading cause of death in the United States but is much less deadly today due to the development of drugs and combination therapies to treat it; however, the development of drug-resistant strains of TB is cause for concern. Worldwide, TB remains a major cause of morbidity and mortality, particularly in Africa and South East Asia.

TB is spread primarily through the air, when a person with active TB puts the bacilli in the air through coughing or sneezing and other people breathe in the bacilli. When a person breathes in TB, the bacilli may settle in the lungs and from there can move to other parts of the body. TB is not a highly contagious disease, and in fact only 20% to 30% of people exposed to TB bacilli become infected. Infection is most common among people who have daily or frequent contact with a person with active TB, such as a family member or coworker. The symptoms of active TB include persistent cough, coughing up blood, weakness and fatigue, weight loss, chills, fever, and night sweats.

The most common test for TB is a skin test that involves inserting a small amount of fluid under the skin of the forearm; after 2 or 3 days the skin test is "read" by a health care worker to determine if it is positive or negative. A positive skin test generally indicates exposure to TB, but does not mean that the person has active TB. In fact, most people who test positive for TB have only an inactive or latent infection, meaning that they are not currently sick but that the TB bacilli are present in their body, so they are at heightened risk of developing TB later in their lives. Persons with latent TB have no symptoms and cannot spread the disease to others. Risk factors for developing active TB include age (babies and young children are at greater risk), gender (males are more at risk during infancy and after 45 years of age, women in adolescence and early adulthood), occupational exposure to silicosis, and stress. Any condition that weakens the immune system also places a person with latent TB at risk: Today a common cause of diminished immunity is infection with HIV, and the combination of the two diseases has worsened the global TB burden. Persons with latent TB infection are often advised to take medication to prevent the latent infection from becoming active, and persons known to have weakened immune systems are sometimes treated prophylactially if they have frequent contact with someone known to have active TB.

Active TB is usually treated with a combination of drugs, the most common of which include streptomycin, isoniazid, rifampin, ethambutol, thiacetazone, and pyrazinamide. In most cases, a course of treatment must be continued for at least 6 months to kill all the TB bacilli in a person's body. However, because the person often feels better with only a few weeks of treatment, he or she may cease to take medications on schedule, therefore risking the chance of becoming ill again and also of breeding drug-resistant strains of TB. Directly observed therapy, in which a TB patient takes medications in the presence of a health care worker, has become common for at least initial TB treatment and is recommended by both the Centers for Disease Control and Prevention (CDC) and the WHO.

History

TB is an ancient disease: It was known to the ancient Greeks as *phthisis* and to the Romans as *tabes*, and evidence of TB has been detected in Egyptian mummies and remains of Neolithic man in Germany, France, Italy, and Denmark. It was established in Western Europe and the Mediterranean states by AD 100, but became a major health concern with the mass population migrations to cities beginning in the 17th century: The crowded city environment created excellent conditions for spreading the disease. In the United States, TB arrived with *the Mayflower* and was well established in the colonies by the 1700s. TB was largely unknown in Africa until the early 1900s, when it was spread by European colonists. Robert Koch identified the *Mycobacterium tuberculosis* in 1882 and received the Nobel Prize in 1905 in recognition of this discovery.

Early treatment of TB involved rest, exercise, dietary changes, bloodletting, and sometimes a change of climate (such as moving to the mountains or the seaside), none of which may have had any effect on the disease. In the late 1850s, sanatorium treatment became popular, and TB patients were often sent to live in institutions built in mountainous or rural areas solely for that purpose, a practice that may not have helped the patients (beyond what could have been gained by normal bed rest in any climate) but did decrease the probability of their spreading the disease to others. The first effective treatment developed for TB was streptomycin, introduced in 1946. However, streptomycin-resistant strains appeared within months of its introduction. Other early drugs demonstrated to be effective against TB were sulfanilamide, isoniazid, and para-aminosalicylic acid. The success of pharmacological treatment of TB led many in the medical community to believe that the disease was a thing of the past, at least in the industrialized world.

Neglect of TB control programs, coupled with emergence of HIV, led to resurgence in TB rates in the 1980s, both in the industrialized world and in developing countries. The WHO in 1993 declared TB to be a global health emergency, which led to increased efforts toward TB control. Particularly in developing countries, the high prevalence of latent TB infection, high prevalence of HIV infection, and the emergence of drug-resistant strains of TB make the disease particularly difficult to control.

A vaccine for TB was developed in 1921 by Albert Calmette and Camille Guerin; their vaccine, known as BCG (Bacille Calmette Guerin), was first put into common use in France in 1924. Vaccination became common in Europe, until the "Lübeck Disaster" of 1930 in which a number of children were accidentally

vaccinated with virulent tubercle bacilli and many died. After World War II, use of the BCG vaccine was reinstated in Europe and today is a standard vaccine in the WHO Expanded Programme on Immunization and is used in most countries in the world but is not recommended by the CDC for use in the United States except under very limited conditions. The BCG vaccine has variable effectiveness in different populations and on average probably prevents about half of infections. A BCG-vaccinated individual will be positive for a skin test while the vaccine is still effective. Therefore, its use complicates the identification of individuals with latent or newly acquired disease in low-risk areas such as the United States.

Incidence, Prevalence, and Mortality

The WHO collects and reports data on global TB annually: Reporting is voluntary but nearly all countries in the world participate. The WHO estimates that one third of the world's population, approximately 1.9 billion people, is infected with TB. It is the 8th leading cause of death in the world and caused approximately 1.8 million deaths in 2003, more than any infectious disease other than HIV. Most TB cases occur in the developing world, where it causes 25% of adult preventable deaths.

The WHO annual TB reports are presented by geographic region, which somewhat confuses the picture because some regions include countries with both high and low incidence. Africa has the highest annual incidence (new cases) rate, with 345 cases per 100,000 people, followed by South East Asia (including India, Pakistan, Bangladesh, Nepal, the Maldives, Thailand, Indonesia, Timor-Leste, Myanmar, and North Korea) with 190 cases per 100,000. The WHO estimates that 60% to 70% of adults in the African and South East Asian regions are infected with latent TB. The incidence of TB is lowest in Europe and the Americas, with 50/100,000 and 43/100,000 cases, respectively, although there is wide variation by country within these regions.

In the United States, data on TB have been collected by the CDC, in cooperation with state and local health departments, since 1953. In 2005, 14,097 cases of TB were reported, for a case rate of 4.8/100,000. Asians (25.7/100,000) had the highest case rate among ethnic groups, and the rate was much higher for foreign-born (21.9/100,000) than for U.S.-born (2.5/100,000) persons. Approximately, 1.0% of the U.S. cases were of primary multidrug-resistant TB.

—Sarah Boslaugh

See also Epidemiology in Developing Countries; Public Health Surveillance; World Health Organization

Further Readings

Centers for Disease Control and Prevention. (2005). *Questions and answers about TB.* Atlanta, GA: Author. Retrieved December 13, 2006, from http://www.cdc.gov/nchstp/tb/default.htm.

Coberly, J. S., & Chaisson, R. E. (2007). Tuberculosis. In K. E. Nelson & C. F. M. Williams (Eds.), *Infectious disease epidemiology: Theory and practice* (2nd ed., pp. 653–697). Sudbury, MA: Jones & Bartlett.

National Center for HIV, STD and TB Prevention: Division of Tuberculosis Elimination. (2006). *Reported tuberculosis in the United States, 2005.* Atlanta, GA: Centers for Disease Control and Prevention. Retrieved December 13, 2006, from http://www.cdc.gov/nchstp/tb/surv/surv2005/default.htm.

World Health Organization. (2006). *Global tuberculosis control: Surveillance, planning, financing.* Geneva, Switzerland: Author. Retrieved December 13, 2006, from http://www.who.int/tb/publications/global%5Freport/2006/pdf/full%5Freport.pdf.

World Health Organization. (2006). *Tuberculosis* (Fact Sheet No. 104). Geneva, Switzerland: Author. Retrieved December 13, 2006, from http://www.who.int/mediacentre/factsheets/fs104/en.

Tukey, John
(1915–2000)

John W. Tukey was a mathematician and statistician responsible for many innovations in data analysis. He was born in New Bedford, Massachusetts, and educated at home until he entered Brown University in 1933. After earning degrees in chemistry at Brown, Tukey entered Princeton University in 1937 to continue his study of chemistry. He began attending lectures in the Department of Mathematics, and, in 1939, received a Ph.D. in mathematics at Princeton. He remained there for the rest of his career as Professor of Mathematics, and later, he served as the founding chairman of the Department of Statistics. For most of his career, Tukey also held positions at AT&T's Bell Laboratories, where he worked on projects such as

the Nike missile system, the methods for estimating the depth of earthquakes, and the development of an index for the literature on statistics. He retired from both Princeton and Bell Laboratories in 1985.

Tukey served as a consultant for many clients, including the U.S. government. During World War II, he joined Princeton's Fire Control Research Office, where he worked on issues related to artillery fire control. Later, he applied his expertise in solving time-series problems to the issue of distinguishing nuclear explosions from earthquakes. As a consultant for Merck, he worked on statistical methods for clinical trials, drug safety, and health economics. His education-related consulting included work for the Educational Testing Service and on the development of the National Assessment of Educational Progress.

In 1950, Tukey was a member of an American Statistical Association committee that criticized, in a balanced report, the methodology used in Alfred C. Kinsey's research on sexuality. From 1960 until 1980, he worked for NBC on the development of methods for rapidly analyzing incoming election-night data. Later, he argued in favor of using statistical procedures to adjust U.S. Census enumerations.

Tukey was a proponent of exploratory data analysis (EDA), a data-driven approach that he thought provided a much-needed complement to inferential, model-driven, confirmatory data analysis methods. EDA emphasizes visual examination of data and the use of simple paper-and-pencil tools such as box-and-whisker and stem-and-leaf plots, both of which were developed by Tukey for quickly describing the shape, central tendency, and variability of a data set. These methods remain in use today and have been incorporated into most statistical software programs.

Tukey's work on the problem of controlling error rates when performing multiple comparisons using a single data set resulted in the development of his "honestly significant difference" test. His creation of the "jackknife" procedure for estimating uncertainty in a statistic whose distribution violates parametric assumptions is one widely recognized product of his work on the issue of statistical robustness. With this method, the variability of a statistic is estimated by successively excluding different subsets of data from computations. Because he viewed this procedure as a tool that is suitable for many tasks but ideal for none, Tukey coined the term *jackknife*. Other widely known terms first used by Tukey include *bit* (for binary digit), *software*, *data analysis*, and the acronym "ANOVA" (to refer to analysis of variance).

—Scott M. Bilder

See also Box-and-Whisker Plot; Robust Statistics; Stem-and-Leaf Plot

Further Readings

Brillinger, D. R. (2002). John W. Tukey: His life and professional contributions. *Annals of Statistics, 30(6)*, 1535–1575.

Tuskegee Study

The Tuskegee Study of untreated syphilis in the African American male was conducted between 1932 and 1972. It was the longest nontherapeutic study conducted on humans in the history of medicine. When the numerous breaches of ethical behavior by researchers conducting the Tuskegee Study became known, public outcry was so great that the protection of the rights of participants in medical research were made a priority through legislation and administration.

The Tuskegee Study, conducted by the U.S. Public Health Service, included 616 participants (412 infected with syphilis and 204 controls). The study participants were low-income African American males in Macon County, Alabama, a poor community with a high prevalence of the disease. The purpose of the study was to assess the course of untreated syphilis in African American males and to compare it with that noted in the Oslo, Norway, Study (1929), a retrospective study of untreated primary and secondary syphilis in whites, which was conducted at a time when minimal treatment and no cure was available for syphilis. Other purposes of the Tuskegee Study included raising the public's consciousness of the problem of syphilis, maintaining the momentum of public health work in the area by sustaining cooperative arrangements among state and local governments and the Tuskegee Institute medical personnel, and standardization and developing invention of serologic tests for syphilis.

The researchers involved in the Tuskegee Study believed it represented high-quality research and published various articles on its findings; the idea that the study was unethical on any level was not considered.

Although the researchers may have had good intentions, multiple ethical violations occurred, including (1) there was no informed consent of participants, even though in 1914 the U.S. Supreme Court ruled that every adult human being of sound mind has the right to determine what is to be done with his or her own body; (2) participants were denied treatment of their disease (arsenic and bismuth were available as a treatment for syphilis at the initiation of the study, and penicillin became available as a cure for syphilis during the 1940s); (3) participants were not informed of their illness; (4) participants were not educated as to how their illness was transmitted; and (5) participants were not informed of the risks of participating in the study.

The Tuskegee Study has also been criticized as being the first to address potential biological and genetic differences as rationale for the differences in syphilis among blacks and whites rather than addressing differences in social class, environment, education level, cultural differences, and access to health care. The ramifications of the study are still apparent today with mistrust of the medical and research fields by minorities in America.

The study has been credited by some for its attempt to be culturally sensitive in its approaches to recruitment and retention of research subjects in the midst of unethical practices. Eunice Rivers, an African American nurse, was the liaison for the Public Health Service physicians and the subjects. Eunice Rivers also provided transportation to subjects along with organizing and tracking them for physical examinations.

As a result of the Tuskegee Study, the National Research Act of 1974 was passed by the U.S. Congress. The act led to the creation of the National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research. In 1978, the Commission released *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects.* The report recommended three principles that should guide research on human subjects: beneficence, personal respect, and justice. Beneficence is the performance of actions or behaviors that actively do good or that actively protect from harm. The principle of beneficence requires protection of research participants and mandates specific safeguards to minimize risks that might occur to subjects. It requires that benefits to participants, research investigators, scientific disciplines, and society at large be maximized. The second principle, respect for persons, requires acknowledgment of the research subject's right to autonomy and informing the subject of his or her rights and protection of those with diminished autonomy. The requirements of personal respect involve specific policies to ensure that research subjects are protected from the following: (1) involvement in a study without knowledge or consent, (2) coercion of subjects to participate in studies, (3) invasion of privacy, (4) unfairness and lack of consideration, (5) nondisclosure of the true nature of the study, and (6) deception. Finally, the principle of justice requires fairness in the distribution of the burdens and the benefits of research. Researchers should make every attempt to involve subjects who are most likely to benefit from the research findings in any application. The Belmont principles were also developed to prevent exploitation of vulnerable populations because of race, ethnicity, gender, disability, age, or social class.

The National Research Act also mandated the installation of an institutional review board (IRB) at all research institutions receiving federal funding. The IRB was initially introduced in the 1960s to ensure that adequate measures are taken to secure informed consent in experimental studies. The role of the IRB is to determine if the proposed selection of patients is equitable and to protect the rights and welfare of human subjects.

—Keneshia Bryant

See also Ethics in Health Care; Ethics in Human Subjects Research; Health Disparities; Institutional Review Board; Sexually Transmitted Diseases

Further Readings

Cassell, E. (2000). The principles of the Belmont Report revisited: How have respect for persons, beneficence and justice been applied to clinical medicine? *Hastings Center Report*, *30*(4), 12–21.

Corbie-Smith, G. (1999). The continuing legacy of the Tuskegee syphilis study: Considerations for clinical investigation. *American Journal of Medical Science*, *317*(1), 5–8.

Jones, J. (1993). *Bad blood: The Tuskegee syphilis experiment*. New York: Free Press.

Sales, B., & Folkman, S. (2000). *Ethics in research with human participants.* Washington, DC: American Psychological Association.

White, R. (2000). Unraveling the Tuskegee study of untreated syphilis. *Archives of Internal Medicine*, *160*(5), 585–598.

Twin Studies

Twin studies have been instrumental in building our knowledge of the etiology of common disorders because they provide a mechanism for estimating the relative magnitude of genetic and environmental contributions to disease. Monozygotic (identical) twins share 100% of their genes, and dizygotic (fraternal) twins share on average 50% of their genetic material—the same as in any pair of full siblings. By combining information from monozygotic (MZ) and dizygotic (DZ) twin pairs, an index of heritability can be calculated in the form of a ratio of MZ to DZ twin correlations for a given disorder. Since DZ twins are the same age (unlike other sibling pairs), differences observed between members of DZ twin pairs cannot be attributed to age or cohort differences. A higher correlation among MZ versus DZ twins therefore suggests a genetic contribution to the disorder.

Origins and Assumptions of the Twin Model

The identification of familial clustering of a disorder is only the starting point for investigating its heritability. Because parents provide both genes and environment to offspring, attributions of genetic versus environmental causes of disease cannot be properly made in a traditional parent-offspring study. Studies of twins reared apart allow for a clear distinction to be drawn between genetic and environmental factors in the etiology of disorders. Variability in outcomes between individuals with the same genes in two different environments are attributed to factors that distinguish the two environments. However, the relatively unusual circumstances under which twins are separated at birth need to be considered in interpreting findings (including their possible relationship to parenting behaviors and/or heritable traits). Even more important, the rarity of twins being reared apart creates significant obstacles in the acquisition of samples that are sufficiently large to detect differences between affected and unaffected individuals, especially in the study of relatively uncommon diseases. Studies of twins reared together do not suffer from the above limitations and, although distinguished from single births by shorter gestational periods, disadvantages associated with twin status do not appear to persist beyond 5 years of age. The prevalence or risk factors for numerous adult health conditions, including psychiatric disorders, do not differ between twins and singletons, making findings from twin studies generalizable to the larger population.

The twin method is based on the premise that the environments of etiologic relevance to the trait being studied do not differ significantly between DZ and MZ twins. The equal environments assumption (EEA) is critical in the interpretation of findings from twin studies because it is the basis on which distinctions between MZ and DZ similarity in a given trait are attributed to genetic rather than environmental sources. It has been argued that MZ twins experience environmental conditions that differ from those of DZ twins in that MZ twins are treated more similarly by parents and other individuals in their social environments and that they typically spend more time together and enjoy a closer emotional bond. Whereas critics of the twin method have argued that these apparent violations of the EEA invalidate findings from twin studies, twin researchers have argued for testing the assumption in a number of ways, including measuring the relationship between environmental similarity and outcomes under study and, in cases in which parents are misinformed about zygosity of twins, comparing the impact of actual versus perceived zygosity on twin resemblance. Little evidence of violations of the EEA has been produced, but twin researchers continue to promote rigorous testing and adjustments for violations when they are found.

A Brief History of Twin Studies

Francis Galton is credited with being the first to recognize the utility of twin methodology in establishing heritability of a trait or disorder. In his 1875 paper "The History of Twins," he described twins as affording a means for evaluating the effects of nature versus nurture and acknowledged that there are two kinds of twins, one in which both twins are derived from a single ovum (MZ) and a second in which twins develop from two separate ova (DZ). It was not until the 1920s, however, that the idea of comparing MZ and DZ concordance rates was proposed as a method for assessing heritability. In 1924, both dermatologist

Herman Siemens's study examining melanocytic naevi (moles) and psychologist Curtis Merriman's study on cognitive abilities described comparisons between identical and nonidentical twin similarity to determine the heritability of the traits under investigation, marking these as the first true twin studies. An article published by John Jinks and David Fulker in 1970 signified another major development in the history of twin studies, as it argued for the application of a biometrical genetic approach to the study of human behavior and proposed a framework for testing hypotheses in genetically informed designs.

Partitioning Variance in Twin Models

The goal in using twin methodology is to estimate the proportion of variance in a given phenotype (the detectable outward manifestation of a genotype) attributable to each of three sources: genetic factors, shared environment, and unique environment. Genetic variance, represented by "a" in the twin path model (see Figure 1), refers to the combined effect of all additive genetic factors that contribute to variability in the phenotype, which in the case of complex traits generally means multiple genes. Covariance between a1 and a2 is 1.0 for MZ twins and 0.5 for DZ twins, as MZ twins are genetically identical and DZ twins share on average half of their genes. Shared environment is denoted as "c" in the model and by definition has a covariance of 1.0 in both MZ and DZ twins, as it represents environmental factors common to both members of the twin pair. Unique variance ("e") refers to variance that is not attributable either to genetic factors or to environmental factors common to both twins and is by definition unshared between twin pairs of either zygosity. Using the above nomenclature, variance of a given phenotype is denoted as a2 + c2 + e2. The covariance between MZ twins is represented as a2 + c2 and the covariance between DZ twins as (0.5)a2 + c2.

Computer programs such as Mx and LISREL have been created to build more complex models that take into account additional factors that affect variance estimations, but simple comparisons between correlations of MZ versus DZ twins provide broad indicators of the proportion of variance in a phenotype attributable to genetic and environmental sources. Genetic influence is suggested by higher correlations between MZ than DZ twins, as it is the greater genetic similarity between MZ twins that distinguish them from DZ twins. Shared environmental influence is indicated by DZ twin correlations exceeding half of MZ twin correlations. That is, if DZ twins are more alike than would be expected if similarities were based entirely on genetic factors, the implication is that influences in the shared environment are playing a role in the development of the phenotype. Unique environmental influences are approximated by subtracting the MZ twin correlation for the phenotype (which encompasses both genetic and shared environmental influences on MZ twins) from 1, the total variance.

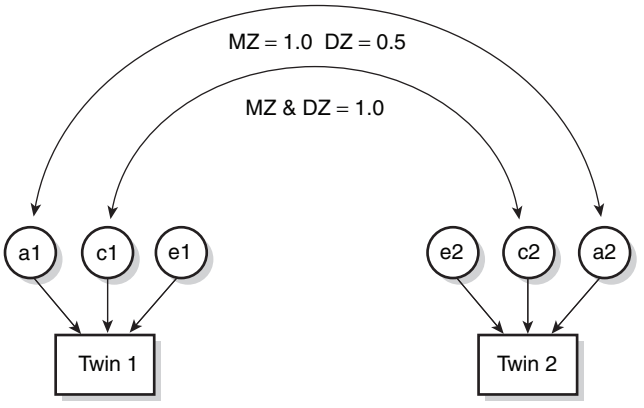


Figure 1 Partitioning Variance in Twin Models

Extensions of the Twin Model and Future Directions

One major limitation of traditional twin methodology is that genetic effects are confounded with gene-environment interactions and genotype-environment correlations. Inflated estimates of genetic contributions to disorders can result from failure to control for variability in genetically controlled sensitivity to environmental factors associated with the phenotype (e.g., responsiveness to stress and depression). Similarly, as individuals with certain genotypes are more likely to seek out particular environments or to evoke differential responses from the environment, environmental exposures cannot be assumed to be randomly distributed in the population. Individuals with (heritable) antisocial traits, for example, are more likely to associate with deviant peers—a known environmental risk factor for developing substance use disorders. Assortative mating or nonrandom choice of sexual

partners (and coparents of offspring) based on similarities that have a genetic basis can also influence patterns of family resemblance.

Data from spouses and additional family members, collected in twin-family studies, can be informative in assigning genetic and environmental sources of variance to the disorder of interest. Another approach designed to address gene-environment correlations and interactions is the offspring of twins model, in which offspring of twins are characterized as high versus low environmental risk and high versus low genetic risk based on the twin parent's status in combination with the parent's cotwin's status on the phenotype. For example, an individual whose father is not depressed but whose father's cotwin meets depression criteria would be considered at high genetic and low environmental risk for depression. Gene-environment effects may also be addressed in part by determining whether estimates of heritability vary according to a specified environmental exposure.

Finally, in addition to its continued utility in evaluating genetic and environmental sources of variance in disease, the twin method has the potential to contribute significantly to the identification of specific genes that impact the development of various disorders. Using DZ twins in linkage studies can increase power because they provide built-in controls for family environment and age. Association studies similarly benefit from the use of twins through their provision of ethnically matched controls as well as the means for estimating genetic variance attributable to a given polymorphism.

—Carolyn E. Sartor

See also Gene-Environment Interaction; Genetic Epidemiology; Genotype; Phenotype

Further Readings

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics, 3*, 872–882.

Kendler, K. S. (1993). Twin studies of psychiatric illness. Current status and future directions. *Archives of General Psychiatry, 50*, 905–915.

MacGregor, A. J., Snieder, H., Schork, N. J., & Spector, T. D. (2000). Twins: Novel uses to study complex traits and genetic diseases. *Trends in Genetics, 16*, 131–134.

Martin, N., Boomsma, D., & Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genetics, 17*, 387–392.

Neale, M. C., & Maes, H. H. M. (1992). *Methodology for genetic studies of twins and families.* Dordrecht, The Netherlands: Kluwer Academic.

Type I and Type II Errors

Type I and Type II errors are two types of errors that may result when making inferences from results calculated on a study sample to the population from which the sample was drawn. They refer to discrepancies between the acceptance or rejection of a null hypothesis, based on sample data, as compared with the acceptance or rejection that reflects the true nature of the population data. Both types of error are inherent in inferential statistics, but they can be minimized through study design and other techniques.

Type I Error

The probability of a Type I error, also known as alpha (α), is the probability of concluding that a difference exists between groups when in truth it does not. Another way to state this is that alpha represents the probability of rejecting the null hypothesis when it should have been accepted. Alpha is commonly referred to as the significance level of a test or, in other words, the level at or below which the null hypothesis is rejected. It is often set at 0.05 which, although arbitrary, has a long history that originated with R. A. Fisher in the 1920s. The alpha level is used as a guideline to make decisions about the p value that is calculated from the data during statistical analysis: Most typically, if the p value is at or below the alpha level, the results of the analysis are considered significantly different from what would have been expected by chance. The p value is also commonly referred to as the significance level and is often considered analogous to the alpha level, but this is a misuse of the terms. There is an important difference between alpha and p value: Alpha is set by the researcher at a certain level before data are collected or analyzed, while the p value is specific to the results of a particular data analysis. For instance, a researcher might state that he or she

would use an alpha level of 0.05 for a particular analysis. This means two things: First, that he or she accepts the fact that if his or her analysis was repeated an infinite number of times with samples of equal size drawn from the same population, 5% of the time the analysis will return significant results when it should not (a Type I error) and that results with p values of 0.05 or less will be considered significant—that is, not due to chance. The p value calculated for a particular experiment can be any number between 0 and 1: In this example, a p value of 0.02 would be considered significant while a p value of 0.8 would not be.

As an example of a Type I error, consider the case of two normally distributed populations whose true means are equal. If an infinite number of samples are drawn from those populations, the means of the samples will not always be equal, and sometimes will be quite discrepant. Because in most cases we do not know the true population means, we use statistics to estimate how likely the differences in the means found in our samples are, if the population means were truly the same. In doing this, we accept that in some percentage of the cases, we will make the wrong decision, and conclude that the population means are different when they are truly the same: The probability of making this incorrect decision is Type I error or alpha.

Type II Error

The probability of a Type II error is known as beta (β). Beta is the probability of concluding that no difference exists between groups when in truth it does exist. As with alpha, we accept that there is some probability of drawing incorrect conclusions merely by chance: Often, the probability is set at 20%.

The complement of beta (i.e., $1 - \beta$) is known as statistical power, and describes the probability of detecting a difference between sample groups when a difference of specified size truly exists in the population. The commonly accepted power level is 80%, corresponding to a beta of 20%, meaning that if a true difference at least as large as we specify truly exists in the population from which our samples are drawn, over an infinite number of trials, we will detect that difference 80% of the time. If the power of a study is low, it may not be able to detect important differences that may truly exist, thereby missing potentially important associations.

Importance of Type I and Type II Errors

Type I and Type II errors are generally thought of in the context of hypothesis testing. In hypothesis testing, the null hypothesis (H_0) is often that there is no difference between groups while the alternative hypothesis (H_a) is that there is a difference between groups. Type I and Type II errors are the two types of errors that may occur when making a decision based on the study sample as to whether the null hypothesis or the alternative hypothesis is true. The 2×2 table shown below (Table 1) illustrates when a Type I or Type II error occurs in the context of hypothesis testing. These errors are important concepts in epidemiology because they allow for the conceptualization of how likely study results are to reflect the truth. From them, guidelines can be set as to what is an acceptable amount of uncertainty to tolerate in the sample to make an inference to the truth in the population and gives an idea of how likely the data are to be able to detect a true difference.

The probability of a Type I error, alpha, and the complement of the probability of a Type II error, power, are used in the calculation of sample size. Prior to beginning a study, it is necessary to decide on the levels of error that are acceptable and from this, determine the sample size that corresponds to the chosen levels of error. As stated previously, although the common alpha and power level are 0.05 and 0.80, respectively, sometimes researchers choose different levels. Their choice depends in part on the relative importance of making a Type I or Type II error, because there is a trade-off between the alpha and power levels: When the alpha level is set lower, the beta necessarily becomes higher and vice versa. Figure 1 demonstrates why this trade-off occurs. Figure 1a shows a scenario where, using a one-sided test and specifying the alternative hypothesis as the average amount by which males are taller than females, or delta (Δ), the alpha is set at 0.05, and the beta is 0.20. When the alpha level is changed to 0.10, keeping all

Table 1 Hypothesis Testing and Errors

	H_0 Is True	*H_0 Is Not True*
Accept H_0	Correct	Type II error
Reject H_0	Type I error	Correct

other factors (i.e., sample size) the same, the beta necessarily lowers to 0.12 as seen in Figure 1b. This happens because the amount of overlap between the two curves is predetermined by the values given to the null and alternative hypotheses. Increasing the alpha level shifts the cutoff to the left, thereby decreasing the size of beta (and increasing power).

Generally, beta is set much higher than alpha because some consider it to be a less serious error to make, though there is controversy in this statement. Deducing that no difference exists between groups when it does seems a less harmful mistake because it may lead to lack of action on the part of scientists (i.e., not implementing an efficacious intervention or drug treatment). On the other hand, deducing that there is a difference when there really is not may lead to inappropriate action and could lead to harmful side effects that do not bring with it the expected benefits. The debate, however, comes about with the realization that lack of action is not always less harmful and, therefore, the levels at which the alpha and beta are set depend on the potential costs or benefits that may result from a Type I or Type II error.

The concepts of Type I and Type II errors also pertain to instances where the outcome measure is a categorical variable, not continuous. The principle is the same although statistics appropriate to categorical data are used to estimate effect size, such as chi-square or odds ratio (*OR*), rather than a statistic such as mean difference between groups. An example using the *OR* is demonstrated in Figure 2, which also demonstrates the influence of sample size on Type I and Type II errors. Let us assume the real effect is $OR = 1.5$ in the population and the alpha level is set at 0.05. The larger study sample has a smaller confidence interval (CI) range and is able to detect the difference between the experiment and the control groups at the set alpha level (0.05); whereas a smaller study sample fails to do so because the CI includes $OR = 1$ (H_0). In other words, the analysis based on the smaller study sample resulted in a Type II error, failing to detect a true difference, which could also be stated as failing to reject the null hypothesis when it should have been rejected.

In general, increasing the study sample size decreases the probability of making a Type II error without having to increase the alpha level. This is, in part, because increasing the sample size decreases sample variance and increases statistical power. Along the same line, when the sample size is very large, there is a high likelihood of finding statistically significant differences between study groups; however, the differences are not necessarily clinically significant.

In epidemiology, there has been some discussion as to the ethicality of conducting a study that has a high probability of Type II error, even if the likelihood of a Type I error is low. The issue arises because study participants are asked to take on risks by being in a study that they wouldn't take on otherwise, such as the use of experimental drugs or potential breaches to confidentiality, and many researchers consider it unethical to expose them to those risks unless the study has a high probability of finding differences if they truly exist. In a well-designed study with proper methodology, human subjects' protections, and ample power, the risks taken on by

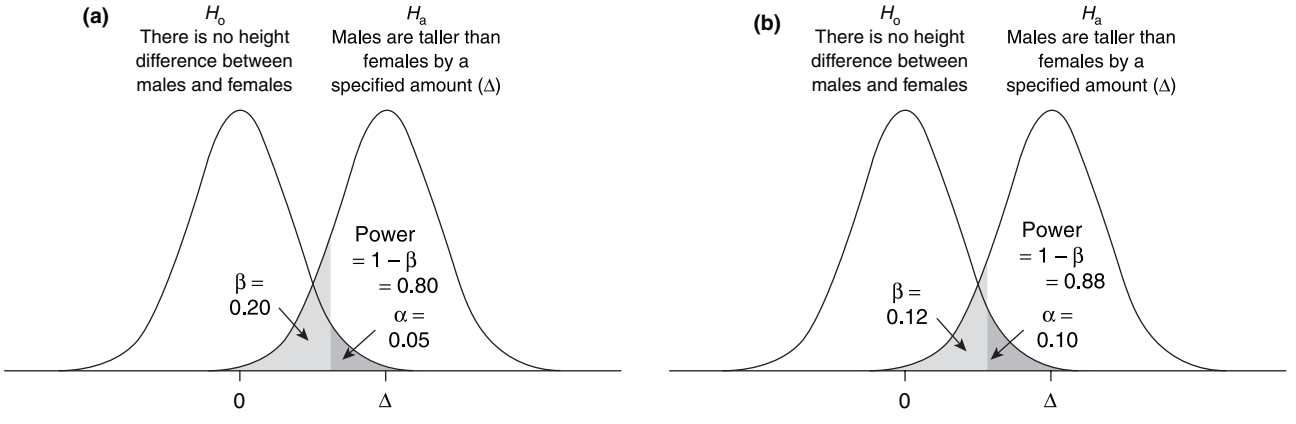


Figure 1 Type I and Type II Error Trade-Off

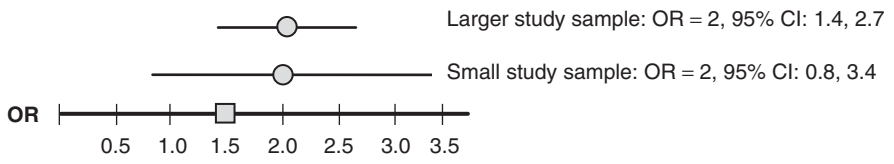


Figure 2 Type I and Type II Errors for Categorical Variables

participants are outweighed by the potential benefits to society that come with scientific findings. However, in a study with low power, the risks may not be outweighed by the potential benefits to society because of the lesser probability that a true difference will be detected. Grant applications and institutional review board proposals often require a power analysis for this reason, and further require that researchers demonstrate that they will be able to attract sufficient study subjects to give them adequate power.

—Rebecca Harrington and Li-Ching Lee

See also Hypothesis Testing; Multiple Comparison Procedures; *p* Value; Sample Size Calculations and Statistical Power; Significance Testing

Further Readings

Gordis, L. (2004). Randomized trials: Some further issues. In *Epidemiology* (3rd ed., pp. 130–146). Philadelphia: W. B. Saunders.

Rosner, B. (2000). Hypothesis testing: One-sample inference. In *Fundamentals of biostatistics* (5th ed., pp. 211–271). Pacific Grove, CA: Duxbury Press.

Rothman, J. R., & Greenland, S. (1998). Approaches to statistical analysis. In J. R. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 181–199). Philadelphia: Lippincott-Raven.

Clemson University. (2001). Type I and type II errors: Making mistakes in the justice system. *Amazing applications of probability and statistics*. Retrieved August 12, 2006, from http://intuitor.com/statistics/T1T2Errors.html.

U

United Nations Children's Fund

The United Nations Children's Fund (UNICEF) is a United Nations program focused on the rights of children. UNICEF is headquartered in New York City and is primarily funded by governments and charitable donations.

History

Originally dubbed the United Nations International Children's Emergency Fund, UNICEF was founded in December 1945 to provide food, clothing, and health care to impoverished children in Europe after World War II. In 1950, UNICEF's mandate was expanded to address the needs of children and women in all developing countries, becoming a permanent program of the United Nations in 1953. UNICEF was awarded the Nobel Peace Prize in 1965. In 1989, the General Assembly of the United Nations adopted the Convention on the Rights of the Child, a set of standards ensuring human rights of children aged 18 years and younger. This framework serves as the basis for UNICEF's work.

Priorities

Working in 191 countries, UNICEF currently has five focus areas: child survival and development; provision of basic, compulsory education for all boys and girls; HIV/AIDS prevention for children; protection of children from violence and exploitation; and public policy. The child survival program centers on using evidence-based, high-impact interventions to lower the number of preventable maternal, newborn, and child deaths. The education program is based on the principles of human rights and gender equality, with the philosophy that education is a means to ending poverty and disease. In fighting HIV/AIDS, UNICEF has set out to reduce the number of new infections in children, particularly among infants and young adults. In addition, the program focuses on providing support to orphans and families affected by HIV/AIDS. The child protection focus area advocates for the development of a protective environment for children, free from the threats of violence, abuse, and exploitation. Finally, the public policy focus area uses data analysis to clarify the pathways by which policy affects the well-being of women and children in developing countries.

The Millennium Development Goals

In 2000, the world's leaders met at a summit to address the eradication of poverty, resulting in eight Millennium Development Goals with a target date of 2015. Six of the eight goals have an intrinsic link to children: eradicate poverty and hunger; achieve universal education; promote gender equality and empower women; reduce child mortality; improve maternal health; and combat HIV/AIDS, malaria, and other diseases. UNICEF's work in the five focus areas directly relates to these goals. UNICEF indicators are used as measures of progress toward a number of goals.

—Anu Manchikanti

See also Child and Adolescent Health; Health, Definitions of; Maternal and Child Health Epidemiology

Web Sites

UNICEF: http://www.unicef.org.

Unit of Analysis

In a statistical model, the unit of analysis is the entity about which inference is being made. For example, in a clinical study, an investigator must decide if inference is to be made with regard to individual patient outcomes or with regard to the physicians treating the patients. If the former, then the unit of analysis is the patient, and the resulting odds ratios or relative risks (or other statistics) would be interpreted as reflecting changes in patient risk or differences in patient characteristics. Similarly, it may be desirable to make inference regarding the treating physicians, each of whom may have treated multiple patients. In this case, statistical methods should be chosen so as to address questions related to the physician.

In most instances, the selection of the unit of analysis is straight-forward. In a cross-sectional survey of patients in the emergency department waiting room, the unit of analysis would be the survey respondent, the patient. In a randomized clinical trial of the effectiveness of a new medication treatment, the unit of analysis would again be the individual patient.

Proper identification of the unit of analysis is critical. Failure to do so may result in biased or invalid results. In a clinical study of 100 patients treated by 5 different doctors, if the unit of analysis is the patient, we end up ignoring the fact that patients treated by one doctor will have certain characteristics in common compared with patients treated by another doctor. That is, a doctor is likely to approach different patients in a roughly similar way. Ignoring this ''clustering'' by physician, or selecting analytic techniques that do not take this into account will yield incorrect estimates of variance, leading to erroneous confidence intervals or p values.

If we are interested in making inference with regard to the treating physicians, patient outcome measures can be summarized with means or proportions within treating physician. The analyses of these types of data require different statistical tests and have a different interpretation than if the patient was the unit of analysis. Analyses that take into consideration effects at these different levels (e.g., patient and physician) are often referred to as ''multilevel'' or ''hierarchical'' models.

—Annette L. Adams

See also Confidence Interval; Inferential and Descriptive Statistics; Multilevel Modeling; Point Estimate; Predictive and Associative Models

Further Readings

Divine, G. W., Brown, J. T., & Frazier, L. M. (1992). The unit of analysis error in studies about physicians' patient care behavior. *Journal of General Internal Medicine*, *7*(6), 623–629.

Pollack, B. N. (1998). Hierarchical linear modeling and the ''unit of analysis'' problem: A solution for analyzing responses of intact group members. *Group Dynamics: Theory, Research, and Practice*, *2*(4), 299–312.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Urban Health Issues

Demographic trends suggest that there is an urgent need to consider the health of urban populations. Cities are becoming the predominant mode of living for the world's population. According to the United Nations, approximately 29% of the world's population lived in urban areas in 1950. By 2000, 47% lived in urban areas, and the United Nations projects that approximately 61% of the world's population will live in cities by 2030. Overall, the world's urban population is expected to grow from 2.86 billion in 2000 to 4.94 billion in 2030. As the world's urban population grows, so does the number of urban centers. The number of cities with populations of 500,000 or more grew from 447 in 1975 to 804 in 2000. In 1975, there were four megacities with populations of 10 million or more worldwide; by 2000, there were 18, and 22 are projected by 2015. Most cities are in middle- to low-income countries; in 2000, middle- to low-income countries contained 72% of the world's cities.

Epidemiology can play a central role in studying both health and disease in the urban context and how

urban characteristics may influence the health of populations. Characteristics of the urban environment that may shape population health include features of the social and physical environment and features of the urban resource infrastructure. Features of the social and physical environment and the urban resource infrastructure in turn are shaped by municipal, national, and global forces and trends.

Defining Urban Areas

One of the key challenges that faces epidemiologic inquiry about health in cities and how city characteristics influence health is that there is little consensus about the definition of *urban* and what constitutes a city. The U.S. Bureau of the Census defines an urbanized area by specifying a minimum population (50,000 people) and a particular minimum population density (1,000 people per square mile). The Census Bureau thus provides a dichotomy whereby territory, population, and housing units within specific size and density parameters are designated as urban and those that are outside those parameters are nonurban. However, there are inherent limitations to these definitions; urban areas exist in contrast to rural or simply in contrast to nonurban areas. In the 21st century, only a few cities, such as Las Vegas, exist in extreme isolation where what is not defined as city is rural. Most cities (e.g., New York City, London, Bangkok) are actually far-reaching densely populated areas, containing periurban and suburban areas, which continue relatively uninterrupted for miles beyond the municipal city boundaries and the city center. To accommodate varying conceptions of what constitutes an urban area, alternative definitions have been developed. They vary in how they define rates of disease, risk, and protective behaviors.

The definition of *urban* also varies widely between countries. Among 228 countries for which the United Nations had data in 2000, almost half (100) include size and density as criteria, 96 include administrative definitions of *urban* (e.g., living in the capital city), 33 include functional characteristics (e.g., economic activity, available services), 24 have no definition of urban, and 12 define all (e.g., Anguilla, Bermuda, the Cayman Islands, Gibraltar, the Holy See, Hong Kong, Monaco, Nauru, Singapore) or none (e.g., Pitcairn Island, Tokelau, and Wallis and Futuna Islands) of their population as urban. Official statistics (e.g., United Nations statistics detailed above) rely on country-specific designations and, as such, vary widely. In specific instances, definitions of *urban* in adjacent countries vary tremendously (e.g., Cambodia vs. Vietnam). Furthermore, definitions of *urban* have evolved in different ways in different countries. Therefore, global statistics are subject to country-level differences in the definition of *urban* that may be based on population density or specific urban features (e.g., proportion of agricultural workers, municipal services).

Urban "Exposure" As a Determinant of Health

It may be heuristically and methodologically useful to conceptualize urban exposure in two main ways: urbanization and the urban environment. Epidemiologic inquiry can be guided by an understanding of how these different facets of urban exposure may influence population health.

Urbanization refers to the change in size, density, and heterogeneity of cities and provides a perspective for public health planning. Factors such as population mobility, segregation, and industrialization frequently accompany urbanization. More simply stated, urbanization is the process that involves the emergence and growth of cities. Thus, the process of urbanization does not depend on definition of *urban* per se but rather on the dynamics of agglomeration of individuals. Although the pace of urbanization is independent of the base size of the population, the population size and density of surrounding areas may shape the pace of urbanization. For example, urbanization may include the establishment (or destruction) of new buildings or neighborhoods, development (or removal) of transportation routes and the in-migration and out-migration of people, and changing racial/ethnic composition of cities.

How the dynamics of urbanization affect health can be considered with examples. An influx of impoverished peoples to a city (e.g., immigration driven by food or work shortages in nonurban or other urban areas) in search of jobs and services may tax available infrastructure, including transportation, housing, food, water, sewage, jobs, and health care. Overtaxed sanitary systems may directly lead to rapid spread of disease, as has been the case many times in North America during the past century and as continues to be the case in the developing world today. Also, the population strain on available jobs may result in

devaluation of hourly wage rates, higher unemployment, and changing socioeconomic status for persons previously living in a given city. This lowering of socioeconomic status can result in more limited access to health care and may lead to poorer health. Therefore, characteristics of urbanization—including the intensity, rate, and duration of such changes as well as the response to these changes—may have health effects on urban residents. Common mechanisms may exist through which urbanization affects health independent of the size of the city in question.

The *urban context or environment* can be defined as the specific characteristics or features of cities that influence health within a particular city. It is helpful to think of the urban environment as involving three distinct concepts: the social environment, the physical environment, and the urban resource infrastructure. The social urban environment comprises contextual factors that include social norms and attitudes, disadvantage (e.g., neighborhood socioeconomic status), and social capital (e.g., social trust, social institutions). The urban physical environment refers to the built environment, pollution, access to green space, transportation systems, and the geological and climatic conditions of the area that the city occupies. Features of the urban resource infrastructure that influence health may include factors such as the availability of health and social services and municipal institutions (e.g., law enforcement). Features of the social and physical environment and infrastructural resources are all, in turn, shaped by municipal, national, and global forces and trends.

Studies of Health in Urban Populations

Three study designs—urban versus rural studies, interurban studies, and intra-urban studies—have been principally employed to consider both the health of urban populations and how characteristics of cities may influence the health of urban residents. Each has strengths and weaknesses, and these methods may lend themselves to addressing different questions. A multiplicity of methods, including qualitative and quantitative methods, may be employed within each of these designs.

Urban Versus Rural Studies

Urban versus rural studies typically contrast urban areas with rural areas in the same country or consider morbidity and mortality in urban versus nonurban areas. Essentially, these studies seek to determine whether morbidity and mortality due to a specific health outcome is different in specific urban areas as compared with specific nonurban areas.

Urban versus rural (or nonurban) comparisons are useful in drawing attention to particular health outcomes that may be more or less prevalent in urban areas and merit further investigation to examine the specific features of the urban (or rural) environment that are associated with that outcome. Recognizing that urban-rural comparisons are too blunt, more recent work has refined distinctions such as urban core, urban adjacent, urban nonadjacent, and rural. However, such studies are limited in their ability to identify what those factors may be and the pathways through which they affect the health of urban dwellers. Features of cities change over time, and some factors may not be conserved between cities (e.g., racial/ethnic distribution). Thus, it is not surprising that different urban-rural comparisons have provided conflicting evidence about the relative burden of disease in urban and nonurban areas. At best, these studies reveal gross estimates of the magnitude and scope of health measures in broad areas by geographical areas typically defined by size and population density.

Interurban Studies

Interurban studies typically compare health outcomes between two or more urban areas between or within countries. Such studies can simply identify differences between cities or can examine specific features of cities that influence health. Examples of the former are numerous. For example, Vermeiren, Schwab-Stone, Deboutte, Leckman, and Ruchkin (2003) have compared mental health outcomes among adolescents in New Haven (United States), Arkhangelsk (Russia), and Antwerp (Belgium), providing insights into the cross-cultural, cross-urban similarities and differences in antisocial behavior, depression, substance use, and suicide. A study of Puerto Rican injection drug users in New York City (United States) and Bayamóa (Puerto Rico) revealed several differences between the two ethnically similar populations; injection drug users in Puerto Rico injected more frequently and had higher rates of needle sharing as compared with their New York counterparts. The authors pointed to similarities in drug purity and

differences in the onset of the crack epidemic as city-level factors that influenced injector risk behaviors. When using the city as the unit of analytic interest, one implicitly assumes that city-level exposures are equally important for all residents. Studying differences in drug use risk behaviors among two cities does not permit analysis of differences in behaviors within cities because of location of residence, intra-urban variability in barriers to safer behaviors, or variations in access to key services (e.g., drug treatment, needle exchange) provided to different urban residents. However, interurban studies such as the examples mentioned here can help guide municipal and state policymakers when making decisions on service provision throughout a city.

Intra-Urban Studies

Intra-urban studies typically compare health outcomes within cities and are being widely used to investigate specific features of the urban environment. These studies often focus on neighborhoods, specific geographic areas within a city that are generally administrative groupings (e.g., census tracts in Canada, subareas or suburbs in South Africa). However, it is important to note that administrative groupings may not represent residents' perceptions of their neighborhoods.

Intra-urban studies may contribute important insights into the relations between specific urban features and health outcomes. However, it may be difficult to generalize from one city to another. For instance, the relation between collective efficacy and violence may be modified by different levels of policing or differential access to illicit substances within a given city. Furthermore, it is important to consider that neighborhood residence is a function of geographical location and other types of social ties that are facilitated or necessitated by the urban environment.

Defining and Quantifying Urban Exposures

When considering a complex and broad exposure such as urbanization or the urban environment, epidemiologic inquiry may fruitfully be guided by considering the elements of urban areas that mechanistically may shape the health of urban populations. It may be useful to consider how the social environment, the physical environment, and the urban resource infrastructure may influence population health.

Social Environment

The urban social environment includes features such as social norms and attitudes, social capital, and income distribution. This list is by no means exhaustive; the further readings provide a more comprehensive look at the urban social environment.

Social norms are patterns of behaviors that are considered accepted and expected by a given society. From the perspective of urban health, the multiple levels of societal and cultural norms are important considerations when thinking about the behavior of urban dwellers. Persons within the urban environment may be influenced by the social norms of their local, geographically defined community, with its unique physical and social structures and cultural characteristics. However, communities may not be limited to one geographic location. Persons in urban areas may also be influenced by the norms operating within the broader urban community.

Social cohesion is typically defined as the connectedness among groups and includes both the presence of strong intra-group bonds and the absence of intra-group conflict. Social capital, a related construct, is thought to provide resources for collective action. Both may be particularly important in densely populated urban areas, where social interaction shapes daily living. There is evidence that the absence of social capital is associated with negative health outcomes such as increases in mortality, poor self-rated perception of health, higher crime rates, and violence.

Income inequality is the relative distribution of income within a city or neighborhood and is typically operationalized with the Gini coefficient. Income inequality is thought to operate through material and psychosocial pathways to shape population health independently of absolute income. Income inequality has been associated with several health outcomes, including self-rated health, cardiovascular mortality, and consequences of illicit drug use. Additionally, emerging work suggests that intra-urban neighborhood income inequality is associated with adverse health outcomes.

Physical Environment

The urban physical environment refers to the built environment (e.g., green space, housing stock,

transportation networks), pollution, noise, traffic congestion, and the geological and climate conditions of the area the city occupies. The built environment includes all human-made aspects of cities, including housing, transportation networks, and public amenities. Recent studies have suggested that poor quality of built environments is associated with depression, drug overdose, and physical activity. Green space (e.g., parks, esplanades, community gardens) has the potential to significantly contribute to the health of urban dwellers. Living in areas with walkable green spaces has been associated with increased longevity among elderly urban residents in Japan, independent of their age, sex, marital status, baseline functional status, and socioeconomic status.

Urban transportation systems include mass transit systems (i.e., subways, light rail, buses) as well as streets and roads. Urban transportation systems are key in the economic livelihoods of city residents as well as cities as a whole. On the other hand, there are significant health considerations for mass transit and roadways, including security and violence, noise, and exposure to pollutants. These exposures are relevant not only for transit workers but also for transit riders.

Pollution is one of the well-studied aspects of the urban physical environment. Urban dwellers are exposed to both outdoor and indoor pollutants that include heavy metals, asbestos, and a variety of volatile hydrocarbons. For example, one study conducted by Ruchirawat et al. (2005) in Bangkok (Thailand) reported high levels of benzene and polycyclic aromatic hydrocarbons among street vendors and school children sampled from traffic-congested areas as compared with monks and nuns sampled from nearby temples.

Urban Resource Infrastructure

The urban resource infrastructure can have both positive and negative effects on health. The urban infrastructure may include more explicit health-related resources such as health and social services as well as municipal structures (e.g., law enforcement), which are shaped by national and international policies (e.g., legislation and cross-border agreements).

The relation between availability of health and social services and urban living is complicated and varies between and within cities and countries. In wealthy countries, cities are often characterized by a catalog of health and social services. Even the poorest urban neighborhood often has dozens of social agencies, both government and nongovernmental, each with a distinct mission and providing different services. Many of the health successes in urban areas in the past two decades, including reductions in HIV transmission, teen pregnancy rates, tuberculosis, and childhood lead poisoning, have depended in part on the efforts of these groups. For example, social and health services may be more available in cities than in nonurban areas, contributing to better health and well-being among urban residents. Despite wider availability of social and health services in cities, many cities experience remarkable disparities in wealth between relatively proximate neighborhoods. This variance is often associated with disparities in the availability and quality of care. Low-income urban residents face significant obstacles in finding health care both in wealthy and less wealthy countries.

Municipal, National, and Global Forces and Trends

Municipal, national, and global forces and trends can shape the more proximal determinants of the health of urban populations. For example, legislation and governmental policies can have substantial influence on the health of urban dwellers. Historically, municipal regulations regarding sanitation in the 19th and 20th centuries facilitated vast improvements in population health and led to the formation of national groups dedicated to improving population health such as the American Public Health Association. A contemporary example of the power of legislation to influence health has been ongoing in New York State since the early 1970s. In 1973, the New York State Legislature, with the encouragement of then Governor Nelson Rockefeller, enacted some of the most stringent state drug laws in the United States. Characterized by mandatory minimum sentences, the Rockefeller Drug Laws have led to the incarceration of more than 100,000 drug users since their implementation. Those incarcerated under the Rockefeller Drug Laws overwhelmingly are New York City residents (78%) and Black or Hispanic (94%). Ernest Drucker (2002) estimated the potential years of life lost as a result of the Rockefeller Drug Laws to be equivalent to 8,667 deaths.

Regional and global trends can affect not only urban living but also the rate and process of urbanization or deurbanization. Changes in immigration policies or policy enforcement can affect urban dwellers in a variety of ways, including, but not limited to, changes in access to key health and social services for some subpopulations, changes in community policing practices, and changes in social cohesion and levels of discrimination. Terrorist attacks in urban centers (e.g., Baghdad, Jerusalem, London, Madrid, and New York City) are associated not only with morbidity and mortality among those directly affected by the event but also with significant psychological distress for other residents of the cities. Armed conflicts have resulted in mass displacement of individuals, some of whom have fled cities for other cities, regions or countries, or camps for displaced individuals (e.g., Darfur).

Future Research

Global demographic trends suggest that urban living has become normative, and there is an urgent need to consider how urban living may influence the health of populations. Epidemiologists may fruitfully be engaged in studying how urban characteristics—including features of the social and physical environment and features of the urban resource infrastructure—can influence health and disease in the urban context. The study of urban health requires a multidisciplinary perspective that can consider different types of studies, including inter- and intra-urban studies and urban-rural comparisons. Epidemiologists' work in the area can complement the work of public health practitioners, urban planners, as well as social, behavioral, clinical, and environmental health scientists in conjunction with the active participation of community residents and civic, business, and political leaders.

—Sandro Galea, Danielle C. Ompad, and David Vlahov

See also Epidemiology in Developing Countries; Governmental Role in Public Health; Multilevel Modeling; Urban Sprawl; Violence as a Public Health Issue

Further Readings

Berkman, L. F., & Kawachi, I. (2000). *Social epidemiology.* New York: Oxford University Press.

Bourgois, P. I. (2003). *In search of respect: Selling crack in el barrio* (2nd ed.). New York: Cambridge University Press.

Colon, H. M., Robles, R. R., Deren, S., Sahai, H., Finlinson, H. A., Andia, J., et al. (2001). Between-city variation in frequency of injection among Puerto Rican injection drug users: East Harlem, New York, and Bayamon, Puerto Rico. *Journal of Acquired Immune Deficiency Syndromes*, *27*(4), 405–413.

Drucker, E. (2002) Population impact of mass incarceration under New York's Rockefeller drug laws: An analysis of years of life lost. *Journal of Urban Health: Bulletin of the New York Academy of Medicine, 79*(3), 434–435.

Friedman, S. R., Cooper, H. L., Tempalski, B., Keem, M., Friedman, R., Flom, P. L., et al. (2006). Relationships of deterrence and law enforcement to drug-related harms among drug injectors in US metropolitan areas. *AIDS (London, England)*, *20*(1), 93–99.

Galea, S., Ahern, J., Rudenstine, S., Wallace, Z., & Vlahov, D. (2005). Urban built environment and depression: A multilevel analysis. *Journal of Epidemiology and Community Health*, *59*(10), 822–827.

Galea, S., Ahern, J., Vlahov, D., Coffin, P. O., Fuller, C., Leon, A. C., et al. (2003). Income distribution and risk of fatal drug overdose in New York City neighborhoods. *Drug and Alcohol Dependence*, *70*(2), 139–148.

Galea, S., & Schulz, A. (2006). Methodological considerations in the study of urban health: How do we best assess how cities affect health? In N. Freudenberg, S. Galea, & D. Vlahov (Eds.), *Cities and the health of the public.* Nashville, TN: Vanderbilt University Press.

Galea, S., & Vlahov, D. (2005). Urban health: Evidence, challenges, and directions. *Annual Review of Public Health*, *26*, 341–365.

Ruchirawat, M., Navasumrit, P., Settachan, D., Tuntaviroon, J., Buthbumrung, N., & Sharma, S. (2005). Measurement of genotoxic air pollutant exposures in street vendors and school children in and near Bangkok. *Toxicology and Applied Pharmacology*, *206*(2), 207–214.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *277*(5328), 918–924.

Semenza, J. C. (2005). Building health cities: A focus on interventions. In S. Galea & D. Vlahov (Eds.), *Handbook of urban health: Populations, methods, and practice.* New York: Springer.

Subramanian, S. V., Delgado, I., Jadue, L., Vega, J., & Kawachi, I. (2003). Income inequality and health: Multilevel analysis of Chilean communities. *Journal of Epidemiology and Community Health*, *57*(11), 844–848.

Takano, T., Nakamura, K., & Watanabe, M. (2002). Urban residential environments and senior citizens' longevity in megacity areas: The importance of walkable green spaces. *Journal of Epidemiology and Community Health*, *56*(12), 913–918.

Vermeiren, R., Schwab-Stone, M., Deboutte, D., Leckman, P. E., & Ruchkin, V. (2003). Violence exposure and

substance use in adolescents: Findings from three countries. *Pediatrics*, *111*(3), 535–540.

Whyte, W. F. (1943). *Streetcorner society: The social structure of an Italian slum.* Chicago: University of Chicago Press.

Urban Sprawl

Sprawl is a single-use, low-density, disconnected approach to community design. By separating places where people live, work, and play and limiting direct connections between these activities, sprawl most often renders driving as the only rational travel option. Distances are often too vast, and walking is most often difficult, if not dangerous, in sprawl. Urban sprawl is associated with several adverse health outcomes, including less walking and overall physical activity, increased sedentary time, exposure to air pollution from automobiles, and increased rates of obesity.

Sprawl is one extreme of a continuum of approaches to land development and transportation investment that collectively determine the urban form of an area, which in turn influences the behavior of the residents—for instance, by making it easier or more difficult to include walking in their daily routines. Urban forms range from sprawl that is auto-dependent all the way to smart growth or new urbanist design, which is arguably pedestrian and transit supportive at the expense of reduced auto access. Therefore, sprawl is one of many typologies of urban form along a continuum of auto to pedestrian and transit orientation.

There are several ways in which sprawl has been measured. Typical metrics of urban form includes measures of both the proximity between complementary land uses (residential, shopping, work, entertainment) and the connectivity or directness of travel between locations dedicated to these uses. Proximity is based on the compactness or density and the intermixing of land uses. Another element used to describe urban forms is the design of street networks, which may range from a connected grid to a disconnected cul-de-sac sprawl-type environment. Other measures include the presence of a continuous pedestrian or bike network, crosswalks that are safe and well demarcated, and the placement or setback of development from the edge of a street. These microscale or site-level measures have been less studied but create the character of a place. For instance, an extremely different environment emerges based on whether shops are next to the street or set behind a *parking lagoon* (large parking lot), a term coined by Howard Kunstler who authored *The Geography of Nowhere* (1993).

Sprawl is a highly regulated and metered approach to developing land and to investing in transportation, but at a scale that is too vast for the pedestrian: A sprawl environment is designed for movement at 40 miles per hour and is therefore boring to the walker, since walking is relatively static at 3 miles per hour. Sprawl may be better understood in contrast with its opposite, walkability. Where sprawl describes an auto-dependent environment, walkability defines those elements of an environment that support active forms of travel (walking and biking) and public transit and reduce car dependence. A voluminous body of literature has emerged in recent years on the health and environmental benefits of walkability. Research has extended the relationship between urban form (sprawl vs. walkable) and associated travel patterns to vehicle-based air pollution, physical activity, and prevalence of obesity.

There is general agreement at this point that as one moves away from sprawl and toward the walkability end of the urban form continuum, per capita vehicle use, greenhouse gas and air pollutants, and obesity prevalence decline, while walking and physical activity levels increase. More recently, studies are showing significant associations between sprawl, climate change, and per capita energy consumption due to increased auto dependence. It is arguable that this line of reasoning looks at only a part of the relationship between sprawl and climate change. Studies should also evaluate differences in per capita home-based energy consumption due to larger spaces and lack of shared energy sources for heating and cooling in sprawling single family environments—in contrast to the sharing of energy sources that is inherent in multifamily housing.

Taken collectively, sprawl is a resource and energy consumptive development pattern and requires more energy; more land; more roads, sewer, water, and other services; and more time to move about from place to place. One study in Atlanta, Georgia, showed that residents of sprawling environments drove 30% more during the week and 40% more on the weekend than those in more walkable areas of that region. Another report from this same study known as SMARTRAQ showed that households in the most sprawling areas of the Atlanta region

consume an average of 1,048 gallons of gas and spend $2,600 per year (assuming two cars per household and $2.50/gallon). Those living in the most walkable areas of the region consumed 262 fewer gallons of gas a year and spend $640 less per year on average. These estimates are conservative. Going from a two- to a one-car household is more feasible in more walkable areas; moreover, inevitable spikes in energy costs can rapidly increase the gap between transportation costs and the expense of heating and cooling a typical home in urban sprawl versus one in a more walkable setting.

Increased awareness of looming natural resource limitations relative to increased population, aging baby boomers, and changing household demographics, and a renewed interest in urban living renders sprawl's future uncertain. Studies are beginning to show that a significant proportion of those who live in sprawl would in fact prefer to be in more walkable environments but that there is an undersupply of homes in walkable neighborhoods, given the patterns of construction and development in the past half century, which produced most new homes in sprawling environments.

Many argue that while there is an association, there is only limited evidence of a causal connection between sprawl and travel patterns, obesity, and the environment (see Special Report 282 from the Transportation Research Board and Institute of Medicine, 2005). This argument rests on the premise that people's behavior is a function of their preferences and predisposition for specific types of travel choices and neighborhood environments. Research to date has most often not disentangled the effect of these preferences versus that of community design on travel, health, and environmental outcomes. More recently, a new body of research is emerging that compares people with similar preferences who are located in different types of urban environments. At least two studies now show that regardless of preferences, exposure to different types of environments (sprawling or walkable) results in different amounts of driving, and that those preferring walkable environments also walk significantly more and have a lower prevalence of obesity when they are located in a walkable environment.

Why do people continue to build and dwell in sprawling urban environments? Over the past several hundred years, people have wanted to have more living space. Also, urban environments created an environment that was ideal for the transmission of disease, leading to people's desire to live in less densely populated areas. Psychological factors may also be at work; some suggest that sprawl offers *defensible space*, a term coined by Oscar Newman as a form of development that does not require dealing with unknown people and where one knows who does and does not belong in their domain. Therefore, living in a sprawling environment has been a rational choice for many. However, the costs of sprawl are now becoming clear, including the costs in terms of health due to increased pollution from automobiles, as well as the barriers to physical activity presented by this form of development.

—Lawrence Frank

See also Geographical and Social Influences on Health; Obesity; Physical Activity and Health; Pollution

Further Readings

Frank, L. D., Saelens, B., Powell, K. E., & Chapman, J. (in press). Stepping towards causation: Do built environments or individual preferences explain walking, driving, and obesity? *Social Science and Medicine*.

Frank, L. D., Sallis, J. F., Conway, T., Chapman, J., Saelens, B., & Bachman, W. (2006). Many pathways from land use to health: Walkability associations with active transportation, body mass index, and air quality. *Journal of the American Planning Association*, *72*(1), 75–87.

Frumkin, H., Frank, L. D., & Jackson, R. (2004). *The public health impacts of sprawl.* Washington, DC: Island Press.

Kunstler, H. (1993). *The geography of nowhere: The rise and decline of America's man-made landscape.* New York: Simon & Schuster.

Newman, O. (1996). *Creating defensible space.* Washington, DC: U.S. Department of Housing and Urban Development.

Rees, W., & Wackernagel, M. (1996). *Our ecological footprint: Reducing human impact on the earth.* Philadelphia: New Society.

SMARTRAQ final summary report: New data for new era. Goldberg, D., McCann, B., Frank, L., Chapman, J., & Kavage, S. (2007, January). The Active Transportation Collaboratory, University of British Columbia. Retrieved March 22, 2007, from http://www.act-trans.ubc.ca.

SMARTRAQ full final technical report: Integrating travel behavior and urban form data to address transportation and air quality problems in Atlanta. Chapman, J., & Frank, L. (April 2004). Active Transportation Collaboratory, University of British Columbia. Retrieved March 22, 2007, from http://www.act-trans.ubc.ca.

Transportation Research Board & Institute of Medicine. (2005). *Does the built environment influence physical*

activity? Examining the evidence (TRB Special Report 282). Washington, DC: Transportation Research Board.

U.S. Public Health Service

As part of the U.S. Department of Health and Human Services, the U.S. Public Health Service (PHS) works both to investigate the causes of disease and to combat epidemics. It traces its origins to an Act signed by President John Adams in 1798. This Act created the Marine Hospital Service, a network of hospitals intended to serve the nation's merchant marines. In 1873, a "Supervising Surgeon General" was named to oversee the Service, and in 1889, the Commissioned Corps, a uniformed and mobile division of medical officers, was created.

During the late 19th century, the growth of trade, travel, and immigration networks led the Service to expand its mission to include protecting the health of *all* Americans. To reflect this change, the Marine Hospital Service was renamed the Public Health and Marine Hospital Service in 1902. Ten years later, in 1912, the name was shortened to the Public Health Service.

Under this new name, the PHS was given clear legislative authority "to investigate the diseases of man and [the] conditions influencing the propagation and spread" of these diseases in 1912. All types of illness, regardless of their cause, now fell under the purview of the PHS.

However, even before this name change, commissioned officers had advocated the use of aggressive and innovative means to investigate and combat diseases as they emerged. As part of this initiative, the Marine Hospital Service launched the *Bulletin of the Public Health* (later renamed *The Public Health Reports*) in 1878. This weekly report tracked epidemics both within and outside the United States. Throughout the past 125 years, this publication, along with *Mortality and Morbidity Weekly Report*, has provided PHS officers with the latest information on disease outbreaks, enabling them to chart an epidemic as it develops.

With the advent of the 20th century, the PHS began to use increasingly sophisticated techniques to track and fight diseases. In 1906, for example, the PHS initiated and implemented an epidemiological investigation into a typhoid epidemic centered in Washington, D.C. Covering a four-state area, this was one of the most comprehensive epidemiological investigations implemented by a public health agency. The investigation—and the corresponding eradication of the epidemic—provided the nation's legislators with a dramatic and very local demonstration of the power of epidemiological investigations both to prevent and arrest epidemics. In the wake of this success, PHS officer Wade Hampton Frost developed and implemented an innovative epidemiological investigation of the 1918 to 1919 influenza pandemic.

Building on these successes, the PHS launched a program to eradicate malaria within the United States during World War II. Based in Atlanta, the Malaria Control in War Areas program (MCWA) gradually expanded to include the control of other communicable diseases such as yellow fever, typhus, and dengue. In 1946, the MCWA adopted a new name, the Communicable Disease Center (CDC) and became a permanent component of the PHS. The CDC received its current designation as the Centers for Disease Control and Prevention in 1992.

Although the CDC has not been the only PHS agency to investigate the "conditions influencing the propagation and spread" of disease both within and outside the United States, it has consistently been at the forefront of this aspect of the PHS mission. In 1951, the creation of the Epidemic Intelligence Service (EIS) at the CDC provided a blueprint for training medical officers in epidemiology. EIS officers, then and now, received a brief but intensive training in epidemiology and statistics followed by an assignment with a public health unit associated with the PHS or a state or local public health department. These EIS officers are on permanent call and can be dispatched quickly to investigate a disease outbreak.

Today, the PHS continues its work through its eight operating divisions: the Centers for Disease Control (CDC), the National Institutes of Health (NIH), the Indian Health Service (IHS), the Food and Drug Administration (FDA), Substance Abuse and Mental Health Services Administration (SAMHSA), the Health Resources and Services Administration (HRSA), Agency for Healthcare Research and Quality, and the Agency for Toxic Substances and Disease Registry.

—Alexandra M. Lord

See also Centers for Disease Control and Prevention; Frost, Wade Hampton; Governmental Role in Public Health; Surgeon General, U.S.

Further Readings

Mullan, F. (1989). *Plagues and politics: The story of the United States public health service*. New York: Basic Books.

Web Sites

Office of the Public Health Service Historian: http://lhncbc.nlm.nih.gov/apdb/phsHistory.

V

Vaccination

Vaccination is the process of producing immunity against a disease by exposing individuals to weakened, dead, or closely related (but relatively harmless) versions of the pathogen that causes this disease. With the advent of widespread vaccination within populations, rates of vaccine-preventable diseases have dropped dramatically, leading to significant decreases in both morbidity and mortality. Through vaccination, for example, smallpox has been completely eradicated and other diseases, such as polio and measles, are in the process of becoming eliminated. In spite of these successes, however, vaccination is not without controversy. Concerns over possible adverse effects have caused individuals in many areas of the world to question the benefit of vaccines.

Credit for the development of vaccination is given to Edward Jenner, an 18th century English physician. In 1796, Jenner successfully vaccinated an 8-year-old boy against smallpox by exposing him to the related, but much less virulent, cowpox virus. Since Jenner's initial success, many more vaccines have been developed to combat a variety of diseases, including measles, polio, diphtheria, rabies, pertussis, and the flu. Vaccines continue to be developed today to combat diseases, such as HIV and malaria, as well as infections, such as human papillomavirus (HPV) that can cause certain types of cancer.

The central idea behind vaccination is that exposure to weakened or dead microbes, parts of these microbes, inactivated toxins, or, like Jenner's smallpox vaccine, closely related but relatively harmless pathogens, can cause an immune response within individuals that can prevent subsequent infection. More specifically, when a person is vaccinated, he or she is exposed to a version of a pathogen that has been altered so that it does not produce disease, but so that it still contains antigens, or the parts of the pathogen that stimulate the immune system to respond. The B lymphocytes in an individual's blood then detect these antigens in the vaccine and react as if the real infectious organism was invading the body. During this process, the B lymphocytes clone themselves producing two types of cells: plasma cells and memory B cells.

The plasma cells produce antibodies that attach to and inactivate the pathogen. This response is known as the primary immune response; it can take up to 14 days for this process to reach maximum efficiency. Over time, the antibodies gradually disappear, but the memory B cells remain. If an individual is exposed to the disease-causing pathogen again, these dormant memory cells are able to trigger a secondary immune response. This occurs as memory B cells multiply quickly and develop into plasma cells, producing antibodies that in turn attach to and inactivate the invading pathogen. Unlike the primary response, this secondary response usually takes only hours to reach maximum efficiency. It is through this process that vaccination is able to protect individuals from disease.

Vaccination is also beneficial at the population level. When a sufficient number of individuals in a population are immune to a disease, as would occur if a large proportion of a population was vaccinated, herd immunity is achieved. This means that if there is

random mixing of individuals within the population then the pathogen cannot be spread through the population. Herd immunity acts by breaking the transmission of infection or lessening the chances of susceptible individuals coming in contact with a person who is infectious. Herd immunity is important because it provides a measure of protection to individuals who are not personally immune to the disease—for instance, individuals who could not receive vaccines due to age or underlying medical conditions or individuals who received vaccines but remain susceptible. It is herd immunity that made the smallpox eradication campaign possible, and it is herd immunity that prevents the spread of diseases such as polio and measles today.

In spite of these benefits to individuals and populations, vaccination itself is not without risk. Common reactions to vaccines include redness and soreness around the vaccination site. More severe adverse reactions are also possible for some vaccinates; these include vomiting, high fevers, seizures, brain damage, and even death, although such reactions are fairly rare. The most serious adverse reactions, for example, occur in less than one case out of a million for most vaccines.

In addition to these known adverse effects of vaccination, claims have also been made that vaccination is responsible for adverse health conditions, such as autism, speech disorders, and heart conditions. While none of these claims is well accepted in the scientific community, they have had a significant impact on individuals' perceptions about the safety of vaccines. Combined with the fact that most individuals have never personally experienced, or seen someone experience, many of the vaccine-preventable diseases in their lifetime, the focus of concern for many people has shifted from the negative effects of the diseases to the possible negative effects of the vaccines themselves. This complacency about vaccine-preventable diseases, combined with concerns over the effects of vaccination, has led to decreasing levels of vaccination coverage in many areas of the world.

Not vaccinating has two general results. Individually, people who are not vaccinated against a disease are at greater risk of contracting this disease than individuals who are vaccinated. At a population level, if vaccination rates are allowed to drop far enough, a loss of herd immunity will result. When herd immunity is not maintained, disease outbreaks occur, and the costs to society in terms of loss of work, medical care, disability, and even loss of life can be high. Such situations have already occurred in Japan, England, and the Russian Federation, where, after concerns about the pertussis vaccine led to significant decreases in the number of vaccinated children, outbreaks involving thousands of children and resulting in hundreds of deaths occurred.

—Emily K. Brunson

See also Disease Eradication; Herd Immunity; Jenner, Edward; Pasteur, Louis; Polio; Smallpox

Further Readings

Feikin, D. R., Lezotte, D. C., Hamman, R. F., Salmon, D. A., Chen, R. T., & Hoffman, R. E. (2000). Individual and community risks of measles and pertussis associated with personal exemptions to immunization. *Journal of the American Medical Association*, *284*, 3145–3150.

Gangarosa, E. J., Galazka, A. M., Wolfe, C. R., Phillips, L. M., Gangarosa, R. E., Miller, E., et al. (1998). Impact of anti-vaccine movements on pertussis control: The untold story. *Lancet*, *351*, 356–361.

May, T., & Silverman, R. D. (2003). "Clustering of exemptions" as a collective action threat to herd immunity. *Vaccine, 21*, 1048–1051.

Salmon, D. A., Teret, S. P., MacIntyre, C. R., Salisbury, D., & Halsey, N. A. (2006). Compulsory vaccination and conscientious or philosophical exemptions: Past, present and future. *Lancet*, *367*, 436–442.

Validity

Validity refers to the extent to which something does what it is intended to do. From this broad perspective, validity equally applies to an object designed to perform certain tasks, to a program targeted to certain goals, or to an instrument intended to measure a given concept or construct. Therefore, it is not limited to the problem of measurement, although it is in this context that we will use it throughout this entry. The concept of validity applies to all measurement situations, but is particularly crucial to social epidemiology, where most research work has to deal with constructs that have to be operationalized. This entry contains a brief review of the concept with emphasis on its empirical and theoretical implications.

The constructs are variables that cannot be directly observed or measured. Quality of life, motivation, satisfaction, and socioeconomic status are typical

examples of constructs. Whenever a construct has to be operationalized, that is, defined in terms of concrete data that can be gathered or behaviors that can be observed, one is inevitably confronted with the problem of validity. Measuring a construct by means of an instrument, such as a questionnaire, always poses a problem of validity; for instance, we may try to measure someone's quality of life by asking him or her a series of questions about limitations, pain, and so on, but we must bear in mind that the answers to these questions are not a direct measure of their quality of life but at best an approximation of it. Another frequent example of a validity problem is statistical inference, where we have to estimate some parameters of a finite or infinite population by examining a sample of it. A sample is valid if it adequately reproduces the characteristics of the population that we want to study. Valid samples are usually said to be representative.

An instrument used to measure a construct can be a single indicator (e.g., income as a measurement of socioeconomic status) or can consist of a set of items (e.g., questionnaires designed to measure quality of life). Validity can be assessed on two grounds: theoretical and empirical.

The theoretical validation of an instrument implies a thorough examination of its contents with the purpose of verifying whether it reflects the meaning we have attached to the construct it is intended to measure. On the other hand, the empirical validation entails a careful testing of the properties that should correspond in practice to that meaning. Accordingly, two main dimensions are involved in the process of validation: the ontological and the methodological dimensions.

To theoretically verify that an instrument reflects the meaning of the construct, we must explicitly state what the construct is or what it means for us. We thereby assume an ontological position that is in practice equivalent to making a contextual definition. For instance, an instrument designed to measure quality of life can be validated neither on theoretical nor on practical grounds if we have not previously defined what quality of life is, or what it means for us. Obviously, the definition may vary from one context to another. For example, it may not be the same when applied to patients with cancer as when applied to healthy people, and it may also differ from one cultural setting to another. People with different views of quality of life will surely not agree on the pertinence of an instrument designed to measure it.

Normally, the processes of creating and validating an instrument are parallel. We do not wait for an instrument to be in its final version to validate it. Rather, we build an instrument by validating portions of it, that is, by discarding some of its items, modifying others, and adding new items to it.

Types of Validity

Validity can be subclassified in two conceptually distinct aspects, content validity and criterion validity. Content validity is assessed on theoretical grounds. Criterion validity has to be tested empirically. Other types of validity include convergent validity, discriminant validity, and predictive validity.

Content Validity

Content validity refers to the capacity of adequately reflecting the essence of the construct in the indicators, items, or observable variables we have chosen as components of our instrument. When we make a choice regarding the number and identity of the dimensions underlying satisfaction, academic performance, professional aptitudes, or quality of life, the pertinence of our choice is closely connected with content validity. The questions or other measures of the construct must address those dimensions, while someone with a different theoretical conception of the same construct could object to the inclusion of an indicator or could suggest the addition of some others.

An important component of content validity is called face validity and can be conceived of as the first step in the assessment of content validity. An instrument has face validity or is valid prima facie if it looks valid at a first nonexpert inspection, meaning that a nonexpert in the field would agree that the instrument measures the construct it is intended to measure.

Criterion Validity

An instrument can be theoretically valid if it contains all and only the relevant components of the construct, and yet it may fail to meet one or more of the practical requirements that are required to create a valid measurement of the construct it is intended to measure. For this reason, instruments have to be tested for criterion validity. For example, if a metric for quality of life is sensitive to different stages in the evolution of disease and correlates highly with

external factors associated with health and well-being, then it fulfills two criteria of validity.

Let us take, for instance, the construct satisfaction that is crucial in health systems assessment. Choosing the indicators that fully represent and exhaust our notion of satisfaction is typically a problem of content validity. Showing that those indicators are associated with outcomes or processes related to satisfaction, and that they are sensitive to positively effective interventions, is part of what we have to prove for criterion validity.

Convergent and Discriminant Validity

These two criteria work closely together. To establish convergent validity, we need to show that measures of similar constructs are in fact related to each other. On the other hand, if measures of constructs that are theoretically not related to each other are proved to be in fact not related to each other, that is evidence for discriminant validity. A common way to estimate the degree to which any two measures are related to each other is to use the correlation coefficient. Correlations between theoretically similar measures should be high, while correlations between theoretically dissimilar measures should be low. There are no fixed rules to say when a correlation is high and when it is low, but in construct validation we always want our convergent correlations to be higher than the discriminant correlations.

Often, convergent and discriminant validity are tested as part of a single inferential act, for instance, through the administration of a questionnaire containing items to be used to assess both types of validity, so that we can judge the relative values of discriminant and convergent correlations. For example, a group of items that are intended to measure the functional capacity component of the quality of life in people 65 years or more should have higher correlations with each other than with items intended to measure the psychological component of quality of life, and these, in turn, should be more highly correlated among themselves than with items that measure other dimensions.

Predictive Validity

Assessing predictive validity requires examining the relationship of events occurring at different points in time. There are two possible interpretations of predictive validity. One is related to the property of being sensitive to changes over time. For instance, an instrument intended to measure a person's quality of life should reflect changes in his or her health condition over time. For instance, if they are known to have suffered a major illness or injury, this should be reflected by changes in their quality of life score as measured by that instrument. To take another example, a set of indicators designed to assess the quality of a health delivery system should reflect the improvement experienced by the system after interventions that are known to be effective and the deterioration of the system after the ocurrence of events that are known to affect its quality. The same principle is in use when a measure is validated by administering it to groups that are expected to score differently: The evidence for validity is established if the instrument distinguishes between groups that can be theoretically expected to be different. For instance, populations that do not benefit from certain quality health services should show lower levels of satisfaction than populations that do benefit from those services, and an instrument that purports to measure satisfaction is expected to show different levels when applied to those subpopulations. For similar reasons, if the type of surgery is actually relevant with respect to self-perception of quality of life in breast cancer, women who were operated on with conservative surgery should perform better when the instrument used to measure quality of life is applied to them, than those who were operated on with radical surgery.

Another meaning of predictive validity is that a measure is correlated with some event that occurs at a later date. Such an instrument, measure, or model has predictive validity if there is good coincidence between prediction and outcome. For instance, scores on an examination may be used to select students for medical training. If the scores correlate highly with their grades or success in training, this is evidence of the predictive validity of the examination.

—Jorge Bacallao Gallestey

See also EuroQoL EQ-5D Questionnaire; Measurement; Quality of Life, Quantification of; Quality of Well-Being Scale (QWB); SF-36® Health Survey

Further Readings

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment*, *7*, 248–260.

Guion, R. M., & Cranny, C. J. (1982). A note on concurrent and predictive ability design: A critical re-analysis. *Journal of Applied Psychology*, *74*, 1337–1349.

Kenny, D. A. (1995). The multitrait-multimethod matrix: Design, analysis and conceptual issues. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods and theory: A festschrift honoring Donald W. Fiske* (pp. 111–124). Hillsdale, NJ: Erlbaum.

Schmitt, N., & Landy, F. J. (1993). The concept of validity. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 275–309). San Francisco: Jossey-Bass.

Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, *17*, 396–408.

Smith, G. T., Fischer, S., & Fister, S. M. (2003). Incremental validity principles in test construction. *Psychological Assessment*, *15*, 467–477.

Vector-Borne Disease

Vector-borne diseases are caused by infectious agents such as bacteria, viruses, or parasites that are transmitted to humans by vectors. In most instances, vectors are bloodsucking invertebrates, usually arthropods such as ticks, mosquitoes, or flies, although vertebrates, including rodents, raccoons, and dogs, can also be vectors of human disease. Infectious agents are most often transmitted by the bite, sting, or touch of a vector, although ingesting or handling the feces of an infected animal can also result in disease transmission. Vector-borne diseases are most common in tropical and subtropical regions where optimal temperatures and moisture levels promote the reproduction of arthropods, especially mosquitoes. Diseases such as malaria, dengue fever, sleeping sickness, and encephalitis have occurred and, in some instances, are still present at endemic or epidemic levels. Reemergence of vector-borne disease is a constant concern due to the rapid rate at which they are capable of spreading. These diseases have played a large role in integrating public health agencies, research, and relief and assistance to areas that are troubled by vector-borne pathogens.

Vector-Borne Disease Transmission

There are two main types of pathogen transmission by vectors, known as internal transmission (sometimes called mechanical transmission) and external transmission. Internal transmission means that a pathogen is carried inside a vector. This can occur as biological transmission, in which the pathogen passes through a necessary stage in its life cycle inside the vector host, which it could not do inside a different host organism. An example of a pathogen that experiences biological transmission is *Plasmodium,* the infectious pathogen that causes malaria. Internal transmission may also occur as harborage transmission, in which the pathogen remains in the same form and life stage inside the vector as when initially entering the vector. The plague bacterium, *Yersinia pestis*, is a harborage transmission pathogen because it does not change morphologically when inside fleas, the common vectors that transmit plague. External transmission occurs when a pathogen is carried on the body surface of the vector. When it lands on a human, the vector passively transmits the pathogen to its new host. An example of a pathogen that is transmitted by external transmission is *Shigella*, which is carried on the legs of flies.

Vector-Borne Pathogens

The list of vector-borne pathogens is long and despite improvements in insect control, understanding of disease, and sanitary conditions of large populations of humans, many of these pathogens are still endemic or epidemic in some parts of the world today. Malaria has beleaguered humans for centuries if not millennia; however, the causative agent *Plasmodium* wasn't identified until 1880 by French army surgeon Charles Louis Alphonse Laveran. There are multiple species of *Plasmodium* that cause malaria, some more infectious than others, with *Plasmodium falciparum* being the most common in occurrence. When a female mosquito bites a human, *Plasmodium* sporozoites are transmitted into the bloodstream, introducing the pathogen to its new, human host. The sporozoites travel through the blood to the liver where they enter cells, mature into schizonts, undergo asexual reproduction, and cause the liver cell to rupture, releasing hundreds of merozoites. Some species of *Plasmodium* will lay dormant in the liver for long periods of time (sometimes years) before maturing and rupturing cells, whereas others will cause liver cell rupture within 2 weeks of initial infection. Merozoites released into the blood enter red blood cells where they rapidly reproduce, causing cell rupture and release of *Plasmodium* that may

be in different life-cycle stages. Some may be immature trophozoites that will enter new red blood cells, mature into schizonts, and lead to the release of merozoites, resulting in a continuous proliferation of *Plasmodium* within the human host, sometimes resulting in a chronic case of malaria. Other *Plasmodium* that are released from rupturing blood cells are gametocytes in the sexual stage that cause sporogonic development in mosquitoes. It is these gametocytes that the female mosquito ingests in her bloodmeal from humans, allowing the life cycle of *Plasmodium* to continue. This period of development typically occurs in mosquitoes of the genus *Anopheles*. During this time, gametocytes (male and female) begin the sporogonic stage in the mosquito's stomach, leading to the production of sporozoites that migrate to the mosquito's salivary glands, facilitating their transfer to a human host when the mosquito ingests another bloodmeal. This stage of the *Plasmodium* life cycle highly depends on temperature in that it requires a minimum temperature to be initiated and will stop if temperature becomes too high. Moisture levels also greatly influence the success of pathogen replication. This temperature and moisture dependence is reflected in the greater occurrence of malaria in tropical and subtropical climates that encourage optimal mosquito body temperature and life span and provide sufficient moisture for mosquito breeding, allowing the pathogen to flourish.

Viruses known as flaviviruses that are typically spread by ticks and mosquitoes cause several vector-borne diseases. Among these are West Nile virus, dengue fever, yellow fever, Japanese encephalitis, and Saint Louis encephalitis. In the case of West Nile, the virus is transmitted to a mosquito when it bites an infected animal, usually a bird, which serves as the pathogen reservoir because it develops immunity to the virus. Multiple species of mosquitoes can transmit the disease to humans and horses, although the primary life cycle of the virus only requires interaction between reservoir hosts and mosquitoes; humans and other mammals are considered incidental hosts. Yellow fever virus is another flavivirus of which there exist two types: Jungle yellow fever, which is transmitted to monkeys by infected mosquitoes and rarely infects humans, and urban yellow fever, which is transmitted to humans by infected *Aedes aegypti* mosquitoes. Yellow fever is endemic in areas of Africa and South America where *A. aegypti* mosquitoes thrive in the warm, moist environments. Often, the virus will lay dormant during the dry season inside $< 1\%$ of the population of female mosquitoes. With the onset of the rainy season, the virus travels to the salivary glands of the mosquito and is transmitted to humans when the mosquito bites and feeds on a bloodmeal. Yellow fever virus has experienced a significant reemergence since the 1980s likely due to reduced mosquito control and lack of vaccination in large populations in susceptible areas.

Tick-borne pathogens are of great concern in North America where they cause Lyme disease, Rocky Mountain spotted fever, tularemia, and others. The tick *Ixodes scapularis* is the vector of Lyme disease, which is endemic in certain areas even though only an estimated 2% to 3% of people bitten by *Ixodes* ticks develop the disease. The bacterium *Borrelia burgdorferi* infects tick larvae when they feed on infected animals, such as mice, and establishes itself in the tick as the tick grows and matures. The ticks then transmit the bacteria to animals or humans when they feed. Two species of ticks, *Dermacentor variabilis* and *D. andersoni*, are responsible for transmission of *Rickettsia rickettsii*, the bacterial pathogen that causes Rocky Mountain spotted fever. This disease occurs in North, Central, and South America and is named for the characteristic rash that develops in the late stages of infection. The ticks are both the natural host and reservoir host of *R. rickettsii*, which survives by living inside the cells of the host organism. There are several ways that the bacteria can be passed on through generations of ticks, including infection of eggs by females and through spermatozoa passed by males to females. The bacteria are transmitted to humans by infected tick saliva passed into the bloodstream when the tick bites and feeds on a human. It often takes a few hours before *R. rickettsii* successfully reaches the human bloodstream.

An example of a vector-borne disease transmitted by a vertebrate animal is hantavirus pulmonary syndrome. In 1993, multiple incidences of acute respiratory syndrome occurred in the southwestern United States (specifically, ''the four corners'' area where Arizona, New Mexico, Colorado, and Utah meet). Researchers quickly identified the cause as a particular type of hantavirus, which they named Sin Nombre virus (SNV). The virus was traced to the deer mouse, *Peromyscus maniculatus*, that was present in unusually high numbers that year due to

heavy rainfall that increased rodent food supplies and, consequently, breeding. The SNV is present in the urine, feces, and saliva of infected mice and is transmitted to humans when the virus particles that linger in the air near where the mice have been are inhaled. Humans may also become infected when handling contaminated mice or items that have been within proximity of the infected mice. Rodents are the only animals that can transmit hantaviruses to humans, which cannot be transmitted from human to human. While hantaviruses are not new, the outbreak served as evidence of their destructive capabilities as nearly half of those who were infected lost their lives.

Prevention and Treatment

Most vector-borne diseases can be prevented by following careful sanitation measures and controlling insect populations with insecticides or biological control methods. Some vaccines are available, such as the vaccines against yellow fever and plague; however, many vector-borne diseases are not treated until severe symptoms have developed. Reasons for this may include ambiguity of symptoms, such as the early stages of Rocky Mountain spotted fever or malaria, which are generally vague, or lack of resources (including insecticides and drugs), such as in poorly developed countries where many of these diseases are prevalent. Some drugs are effective in minimizing disease severity or eliminating disease symptoms. However, most treatment can only stabilize persons with severe disease, such as in the case with hantavirus pulmonary syndrome in which antiviral drugs tend to be ineffective, and the use of a respirator is the best option for providing some relief for affected people. There appears to be greater promise in finding ways to control or eliminate vectors of disease, but unfortunately in countries with few resources even this is difficult because of the unpredictability of vector status from year to year. Informing people about vectors and vector-borne diseases in the areas where they live is of great importance and seems to be the most successful way of preventing disease outbreaks available today.

—Kara E. Rogers

See also Epidemiology in Developing Countries; Insect-Borne Disease; Malaria; Yellow Fever

Further Readings

Marquardt, W. H. (Ed.). (2004). *Biology of disease vectors* (2nd ed.). New York: Academic Press.

Service, M. W. (Ed.). (2002). *The encyclopedia of arthropod-transmitted infections*. Cambridge, MA: CABI.

Thomas, J. C., & Weber, D. J. (Eds.). (2001). *Epidemiologic methods for the study of infectious diseases*. New York: Oxford University Press.

World Health Organization. (1997). *Vector control*. Geneva, Switzerland: Author.

Vehicle-Related Injuries

Injuries from motor vehicle crashes represent a significant public health issue throughout the world. Recent figures compiled by the World Health Organization estimate that 1.2 million people are killed in road traffic crashes each year. Presently, injuries from road traffic crashes rank as the 11th leading cause of death worldwide. This figure is expected to rise exponentially, by up to 83%, over the next two decades as more vehicles are bought and used in the developing world, with major increases forecast for India and China.

The epidemiology of injuries from motor vehicle crashes were eloquently framed by William Haddon more than 50 years ago in the context of the Haddon Matrix. The Haddon Matrix is a model for understanding the dynamics of events related to crash injury. It is drawn by outlining the issues that lead to, occur during, and follow a crash. Important issues that describe injuries in these three crash phases include human, vehicle, and environmental factors. The Haddon Matrix has been expanded in recent years to also incorporate a sociobehavioral component as an additional issue.

The advent of the Haddon Matrix brought about an enhanced understanding of the multiple factors that underlie crashes and their outcomes. The findings from research over the past 50 years have identified a number of risk factors related to vehicle injuries. These include issues related to gender, age, behavior, education, vehicular safety and safety devices, vehicle and road user relationships, road type, climate, emergency response time, and trauma care availability, among others. Today, varied research and prevention programs are underway to reduce the burden related to motor vehicle crashes, including programs from multiple disciplines, such as epidemiology, engineering,

medicine, behavior science, transportation planning, and government.

High-Risk Groups for Motor Vehicle Injury

Several common risk markers have been observed for road traffic crash injury around the world. The major risk groups include younger and older drivers and passengers, persons under the influence of alcohol, aggressive and distracted drivers, motorcycle operators and passengers, and vulnerable road users, primarily pedestrians. While the magnitude of difference in risk of injury may differ for these groups between countries, the general observation that individuals within these groups have a higher injury risk is remarkably consistent. Thus, most present-day research and prevention agendas focus on these risk groups in their road traffic injury prevention efforts.

Young Drivers

Young drivers are often the risk group with the highest reported crash and injury rates. In the United States, the crash rates of young drivers (15 to 20 years of age) are 2 to 3 times higher than drivers of all other ages, except for the elderly. Many issues in young drivers contribute to the higher risks observed. A common factor cited is inexperience with the performance tasks related to driving. The increased crash rate is most pronounced among younger drivers in situations requiring higher performance such as driving in limited visibility and driving at higher speeds. Other factors cited as contributors to higher risks in young drivers are immaturity and higher levels of risk-taking behaviors, particularly alcohol intake. Risks are also borne by the passengers in vehicles driven by younger operators. Most passengers are also between 15 and 20 years of age. Both factors contribute to the fact that injuries from motor vehicle crashes are generally the leading cause of death for persons below 35 years of age in developed regions and a significant cause of death for this age group in the developing regions of the world as well.

Older Drivers and Pedestrians

Older road users are particularly vulnerable to motor vehicle crash injury. Older drivers in developed countries have higher crash rates than all but the youngest drivers. Pedestrian injury is also much higher in persons above 70 years of age. The limitations in physical functioning and performance are the most common factors thought to contribute to this risk. Several reports, for example, note declining visual acuity and peripheral perception in older drivers and link these factors to higher crash frequency. Older drivers also appear to be overly involved in crashes involving turns against traffic, suggesting that lower levels of perception and action and reaction times may contribute to crashes in this group. Older persons are also more vulnerable because their threshold for injury is lower due to declines in muscle mass and bone strength. Injuries in crashes that might have been survivable by younger persons often lead to fatalities in the elderly.

Alcohol Impairment

It is well-known that alcohol is a major contributor to vehicle-related injury globally. Alcohol impairs judgment and delays reaction times in individuals, in general, and is more pronounced with increasing levels of blood alcohol concentration. A motor vehicle crash is considered to be alcohol related in most countries if a driver or pedestrian involved in the crash has a blood alcohol level above the legal limit. In most countries, this limit is between 0.05 and 0.08 g/dl or a blood alcohol level reached by most people consuming one to two drinks in a short period of time. While the frequency of drinking alcohol and driving varies by country, alcohol is generally involved in more than 40% of all fatal motor vehicle crashes. Alcohol involvement in crashes is more pronounced among younger drivers (15 to 34 years of age) and males. Crashes involving alcohol are higher at night (compared with day) and more frequent on the weekend. In the United States, more than one-half of impaired drivers involved in fatal crashes had blood alcohol levels at twice the legal limit or higher. Repeat offending with respect to driving while drinking is common. In the United States, impaired drivers in fatal crashes were 9 times more likely to have a prior conviction for driving under the influence of alcohol compared with drivers with a zero blood alcohol concentration level.

Aggressive Driving

Aggressive driving is emerging as a significant issue in motor vehicle safety. The definition of aggressive driving varies by the context and road

culture of an area. In some cultures, for example, aggressive driving may be perceived as the norm. Many believe that they know aggressive driving when they see it, but the classification of aggressive driving behavior differs between individuals. In many locations, authorities consider aggressive driving to be driving that endangers other persons or property. It typically involves multiple violations or moving vehicle traffic offenses. Common traffic offenses under this scenario include speeding, red light running, failing to allow proper distance between vehicles (tailgating), and failure to yield to other vehicles.

Speeding, alone, does not constitute aggressive driving, but it is a major contributor to motor vehicle crashes and injuries. As injuries are characterized by energy transfer and the body's ability to withstand this transfer, the general rule is that the faster you drive, the greater the likelihood of injury, and of greater severity of injuries. According to the World Health Organization's 2004 report on road traffic injury prevention, an increase in speed of 1 km/hr can result in a 3% higher risk of a crash involving injury and a 4% to 5% increase in risk of a fatal crash. Thus, controlling vehicle speed can help reduce the occurrence of a crash and also the development of injuries when a crash occurs. Speeding may be defined as either exceeding the posted speed limit or driving too fast for road conditions. The data compiled by the World Health Organization indicate that speeding is a factor in about 30% of fatal crashes in developed countries. Alcohol impairment is correlated with speeding, with speeding being more frequently involved in fatal crashes where the driver was impaired. Fatal crashes in which speeding was a factor are also more common among males, younger drivers (15 to 34 years of age), at night, and on rural roads.

Distractions While Driving

Recent research indicates that distractions while driving are frequent contributors to motor vehicle crashes and injuries. The 100-car study recorded the driving experiences of 100 cars over a 1-year period of time and assessed driver behavior with computer and video recordings. Eighty-two crashes were observed in this time, although most were minor. Driver inattention to the driving task was found to be a major issue in crashes. Inattention was noted in 78% of all crashes. Driver inattention in these events was due to the use of a wireless device (cell phone), internal distractions, driver conversations with passengers, and personal hygiene undertaken by the driver. This work is important because many in the highway safety discipline believe that the majority of crashes are related to driver behavior. Interventions to affect driver behavior, then, may provide the greatest success in the future for the prevention of vehicle-related injury. This work, and others, begins to highlight potential areas where interventions might best be placed.

Motorcycles

Lower in cost than automobiles, motorcycles are the primary forms of road vehicle in low-income countries, and they are common in high-income countries as well. Motorcycle riders and operators, though, are vulnerable road users. Limited protection places motorcycle users at greater risk of injury when they are involved in collisions with other vehicles and objects. According to the National Highway Traffic Safety Administration in the United States, "per vehicle mile, motorcyclists are about 34 times more likely than passenger car occupants to die in a traffic crash" (National Center for Statistics and Analysis, 2006, p. 3). One factor for the high vehicle-related injury burden in developing countries is the mix of vehicles on roads, because motorcyclists are vulnerable in crashes with larger vehicles. Motorcycle involvement in crashes varies globally in proportion to the number of motorcycles on the road. In high-income countries, motorcyclists crash fatally with other vehicles about 50% of the time. One aspect behind this frequency is likely to be the limited visibility or lack of awareness of motorcyclists by larger vehicle operators. An age relationship to fatal motorcycle crashes exists, with younger riders at a heightened risk. However, several recent reports also note a high frequency of fatalities in riders above 40 years of age and some link among these riders to higher engine size cycles. In developed countries, most fatalities also involve males and riders operating under the influence of alcohol.

Pedestrians

Pedestrians are the most vulnerable of all road users. A pedestrian is a person on the road who is not in or on a vehicle (motorized or nonmotorized). Injuries to pedestrians occur primarily from mishaps

with other motorized vehicles on the road. The World Health Organization data indicate that pedestrians account for a large proportion of road traffic deaths in low- and middle-income countries, particularly in East and South Asia, and at much higher frequencies than found in high-income countries. Older individuals (above 70 years of age) have the highest rates of pedestrian injury of any age group. About one third of pedestrians involved in fatal events are impaired with alcohol.

Prevention Strategies

The high frequency of road traffic injury has drawn a great deal of attention with respect to preventive interventions. Prevention has been a part of the highway safety discipline for several decades now, and success in reducing motor vehicle fatalities has been recognized as one of the top 10 public health achievements in the 20th century. Unique strategies have been employed to reduce the burden of vehicle-related injury. Most center on passive approaches to design safer vehicles and legislation of safety through traffic codes and laws. A few approaches target high-risk individuals and their behavior in an active attempt to change behavior that is detrimental to transportation safety.

Seat Belts and Car Seats

Seat belts are one of the most effective means to reduce motor vehicle injury and are currently estimated to save more lives than any other preventive strategy. The reports find that three-point safety belts (a lap and shoulder belt attached to the vehicle at the hip) reduce the risk of death in front seat occupants of passenger vehicles by 45% to 61% compared with unrestrained individuals. The largest benefit of seat belts appears for frontal impact, rollover, and rear impact crashes. A seat belt, though, is only effective if it is worn. Large differences exist in the use of restraints around the world, with young male drivers in most locations being the least likely group to wear seat belts. The majority of individuals involved in fatal crashes are also not wearing restraints. Current efforts are underway to increase occupant restraint use. Legislation or laws to require the use of seat belts by occupants in motor vehicles represent the most widely applied approach at this time. Studies indicate that primary seat belt laws (where an individual can be cited by the police simply for not wearing a belt) are more effective in increasing restraint use compared with secondary seat belt laws (where an individual can only be cited for not wearing a belt if they have committed another traffic offense) or no law.

Additional efforts are underway to increase the use of child safety seats, although these efforts are largely focused in high-income countries. Child seats for infants and toddlers can reduce fatality risk by 71% and 54%, respectively. In some areas, booster seats for young children are also required. Booster seats work by placing the child higher in the seat and in a position where the seat belt in the vehicle can be properly deployed. Laws mandating child restraints have been shown to be effective in increasing their use.

Air Bags

Air bags are another safety feature that are a part of most new vehicles today. Air bags work by placing a barrier between the vehicle occupant and the vehicle. The barrier helps dissipate the energy transfer involved in the crash and thereby reduces injury frequency. Over the past 20 years, an estimated 20,000 lives have been saved by air bags in the United States. Air bags combined with three-point seat belts offer the best protection to vehicle occupants. Air bags, though, carry an injury risk of children seated in child safety seats, and children in these seats should not be placed in front seats with air bags. Deployment of the air bag with a rear facing safety seat has led to injury to the child.

Graduated Driver's License

A recent prevention measure focused on young drivers is the graduated driver's license (GDL). A GDL has the intended purpose of controlling the exposure of young drivers to difficult driving situations and increasing the experience level of drivers. There is no standard GDL program, but common features include a permit stage where teenage drivers must be accompanied by an adult for the purpose of getting practice in the driving process. This is usually followed by a provisional license stage, where the driver cannot operate a vehicle late at night and there are restrictions on the number of teenage passengers that can be in the vehicle. A large study has recently found that the most restrictive GDL programs are associated with a 38% reduction in fatal crashes and a 40% reduction in injury crashes. Any type of program

also was beneficial, with 11% and 19% reductions in fatal and injury crashes, respectively.

Helmets

Helmets for motorcycle and pedal cycle users are another effective injury prevention measure. Helmets are worn on the head to protect the individual from serious head injury. They are estimated to prevent fatal injuries to motorcyclists by 37%. Helmets, though, are only effective if they are worn. Helmet use rates, unfortunately, are low in many parts of the world. Factors cited in the low use of helmets include few or no laws requiring helmet use and limited enforcement efforts for existing laws. Areas with a law requiring helmet use have markedly higher rate of use than areas without such laws. Recently, arguments have also been against helmet laws with claims that helmet laws hinder civil rights or that helmets hinder hearing or are too hot for some climates.

Law Enforcement

Law enforcement is an important discipline in highway safety. Strenuous and high visibility enforcement of existing traffic safety laws increases compliance and saves lives. Common programs of law enforcement in highway safety focus on efforts to reduce speeding and driving under the influence of alcohol and increase the use of seat belts and helmets.

Conclusion

Our knowledge of vehicle-related injuries and the factors underlying them has expanded tremendously over the past 50 years. Today, it is possible to predict the frequency of crash and injury with precision and, as such, it is possible to view the majority of crashes as largely preventable. Reductions in most high-income countries in motor vehicle fatalities and injuries have been brought about through a multidisciplinary approach to safety. Engineering, law enforcement, public health, medical, and planning strategies can work to significantly reduce injury. Low- and middle-income countries, though, continue to face a large burden related to motor vehicle injuries. This burden is expected to increase over the next 20 years. Strategies exist to lower the burden of injury from motor vehicles in all parts of the world. The implementation of these strategies, raising awareness of the issue, and enforcing existing laws remain the challenge ahead. These changes and changes in our cultural way of thinking regarding the acceptability of motor vehicle injury will be necessary to improve highway safety.

—Thomas Songer

See also Alcohol Use; Governmental Role in Public Health; Injury Epidemiology; Prevention: Primary, Secondary, and Tertiary

Further Readings

Baker, S. P., Chen, L. H., & Li, G. (2007). *Nationwide review of graduated driver licensing*. Washington, DC: AAA Foundation for Traffic Safety. Retrieved April 9, 2007, from http://www.aaafoundation.org/pdf/NationwideReviewOfGDL.pdf.

National Center for Statistics and Analysis. (2006). *Motorcycles, traffic safety facts, 2005 data* (DOT Publication No. DOT-HS-810–620). Washington, DC: National Highway Traffic Safety Administration. Retrieved July 26, 2007, from http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSF2005/2005TSF/810_620/810620.htm.

Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J., & Goodman, M. (2005). *An overview of the 100-car naturalistic study and findings* (Paper No. 05–0400). Washington, DC: National Highway Traffic Safety Administration. Retrieved April 9, 2007, from http://www-nrd.nhtsa.dot.gov/pdf/nrd-12/100Car_ESV05summary.pdf.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., et al. (Eds.). (2004). World report on road traffic injury prevention. Geneva, Switzerland: World Health Organization.

Traffic Safety Facts 2005. (2005). Various reports on high risk groups and prevention strategies. Washington, DC: National Highway Traffic Safety Administration. Retrieved July 26, 2007, from http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/TSF2005.pdf.

Veterinary Epidemiology

Veterinary epidemiology is a specialized area within veterinary medicine that was historically termed epizootiology until the mid-1990s. Like human epidemiology, it involves identifying risk factors for diseases, characterizing outbreaks, quantifying incidence and prevalence, describing the natural history of disease, developing disease control and prevention programs, and assessing the effectiveness of these

programs. Veterinary epidemiologists participate in these activities in both human and animal populations when disease agents are zoonotic (infectious and capable of spreading between animals and people), although the potential impacts of environmental agents (such as pesticides) on animal and human health and the challenges of cancer and of chronic diseases are also topics for investigation. Veterinarians are trained in medicines of all species, including primates, and so are often involved in identifying disease risks to humans after being alerted to health issues in animals. This can be used in health surveillance, with animals acting as sentinels of human health concerns. One classic example was the use of canaries to detect toxic gases in coal mines.

The concept of ''one medicine'' was described and expanded by visionary veterinary epidemiologist Calvin Schwabe in the 1980s and refers to the common basis of veterinary and human medical knowledge that can be applied to diseases affecting all species. The value of veterinarians and veterinary epidemiologists in active participation in global health research activities has been recognized fairly recently. As in human epidemiology, a primary goal of veterinary epidemiology is prevention of disease rather than treatment.

From ancient times, it has been important to identify patterns of disease in herds and groups of animals used for human consumption (milk, meat, fiber, and eggs) and activities (transportation and farming), and veterinary medicine had its foundations in the treatment of large animal diseases that have financial and survival consequences. As urban centers increased and smaller animals joined human households as companions, veterinarians have expanded their services to include cats, dogs, mice, rats, ferrets, rabbits, birds, and other creatures small enough to coexist in these smaller spaces. Because veterinarians have been trained to identify and treat diseases in groups or herds, they are well suited for and often ''automatically'' engaged in epidemiology and, by extension, public health. While veterinary epidemiology has focused on herd health, that is, on disease patterns in large groups of cattle and other farmed animals, the same principles of recognition and control of infectious diseases hold true in large groups of small animals, such as in catteries, breeding kennels, and animal shelters, as well as in veterinary hospitals, where nosocomial (hospital-based) infections are also of concern.

Veterinary epidemiology uses the same tools as human epidemiology, including observational studies, cross-sectional and longitudinal studies, case-control studies, prospective studies, and experimental and field trials of vaccines, diagnostic procedures, medicines, and treatment protocols. Case reports and case series are often reported by veterinarians engaged in clinical practice. Veterinary epidemiologic research often involves methodologic issues, such as sampling techniques for herds and wildlife populations, and appropriate statistical applications to analyze complex data sets such as capture/recapture data. In survey-based studies, veterinary epidemiologists rely nearly exclusively on proxy respondents, such as owners or farmers, for observations and accurate histories of the animals in their care. Observational studies of animal diseases often depend on recruitment of producers (farmers) and owners of small animals contacted through advertisements in publications of trade associations and breed clubs, or through veterinarians to their clients. An area of considerable study is the determination of test characteristics (sensitivity and specificity) for rapid, portable diagnostics used to screen animal populations for common diseases for which ''gold standard'' testing is too expensive to be used on individual animals. Geographical information system software has been employed to track distributions of herds or disease vectors, the appearance of new disease cases over time (such as of avian influenza), and changes in vegetation (food and shelter habitat for desirable and parasite species) due to weather patterns. Modeling of disease reservoirs and agent transmission has been used to predict outbreaks; other models have been used to show how population sizes may change through implementation of oral contraceptive baiting schemes. As in human epidemiology, a concern in disease reporting is correct identification of denominators, which poses a greater challenge than in human epidemiology because no systematic census exists for wildlife or companion animals.

Financial resources are considerably less for research in veterinary epidemiology compared with epidemiologic studies of humans, especially for the study of small animal diseases, because there has not been the same investment in animal health infrastructure as there is in human medicine. As a consequence, few veterinary epidemiologists have been able to study risk factors for commonly diagnosed animal diseases, particularly of small animals. However, some of these diseases share similarities with human

diseases, and thus, these small animals may provide insight into human health. Examples are prostate cancer and feline immunodeficiency virus infection. Dogs can spontaneously develop prostate cancer and are therefore a model for understanding this common human health problem. Cows, monkeys, and cats have species-specific retroviruses that result in immunodeficiency. The feline immunodeficiency virus is a model for the human immunodeficiency virus, and cats have been studied in an effort to develop vaccines effective in stopping the spread of this worldwide human health problem. A very common disease of older cats is hyperthyroidism, which is similar to one form of human hyperthyroidism. In this case, the human disease has served as a model for the identification of risk factors for disease in cats.

Data Sources for Veterinary Epidemiologic Research

While epidemiologic studies of humans have numerous sources of data for investigation and quantifying disease impacts in the human population, including cancer registries, vital statistics records, occupational registries, and hospital records, similar data sources either do not exist or are not readily available for veterinary epidemiologic research. Internationally, the Office International des Epizooties (OIE), also called the World Organization for Animal Health, maintains reports of notifiable diseases that cross species (including zoonotic diseases), such as anthrax, and species-specific diseases, such as African swine fever. A subset of the OIE-listed "multiple species diseases" are on the Centers for Disease Control and Prevention (CDC) Category A or Category B lists of bioterrorism agents or diseases. Because numerous pathogens on these CDC lists are zoonotic, including anthrax and plague, veterinarians work with public health agencies to prepare for and respond to disease outbreaks that affect several species. The Food and Agriculture Organization of the United Nations has been engaged in numerous programs to control livestock diseases, to protect food safety, and to reduce poverty in developing countries, and by extension, all countries.

Veterinary epidemiologists have been somewhat limited in the availability of active surveillance tools for the study of diseases beyond those of consequence to large animal production (diseases that affect meat, milk, and egg production as well as reproduction, food safety issues, including bacterial diseases such as *Salmonella* spp. and *Listeria* spp., and patterns of antibiotic resistance). The U.S. Department of Agriculture (USDA), Animal and Plant Health Inspection Service (APHIS), and Centers for Epidemiology and Animal Health (CEAH) provide information about animal health issues, emerging diseases, and market conditions, and coordinate animal disease information for international agencies, including the OIE. Within CEAH, the National Center for Animal Health Surveillance includes programs to conduct studies of animal health, and to monitor, integrate, and analyze large animal health data from state and federal agencies to safeguard the food supply and to communicate disease status to agribusiness industries and backyard farmers.

One database tool for veterinary epidemiologists outside government agencies is the Veterinary Medical Database (VMDB), established in 1964 through a grant from the National Cancer Institute. The VMDB receives veterinary medical records for cows, horses, sheep, goats, pigs, birds, dogs, and cats from participating veterinary school hospitals in the United States and Canada. Not all veterinary schools have provided records continuously to the VMDB since its creation, and limitations include the potential for wide disparity in reporting of diagnoses, because standards for animal disease reporting are relatively recent. In spite of limitations, the VMDB has until recently provided the only generally available database of animal diseases, particularly of small animals.

In 2002, the National Companion Animal Surveillance System at Purdue University was established for near-real time syndromic surveillance of signs and symptoms of disease in small animals that could provide alerts of potential outbreaks of zoonotic disease of suspicious origin. This system analyzes records of a nationwide chain of small animal veterinary clinics that are uploaded daily to centralized hubs to evaluate practice methods. The database has been used to identify geographic patterns in the distribution of serovars (strains) of *Leptospira* spp. infections in dogs and potentially in humans (since this is a zoonotic disease) and to examine patterns of vaccine reactions in dogs. These and future studies will benefit from the large number of records made possible by this nationwide primary veterinary care reporting to find patterns in otherwise rare events. The VMDB has also been used in this way, but depends on records from the few veterinary school

hospitals (third-tier referrals), and with time lapses and potentially inconsistent reporting, many more common or rare diseases may not be identified or reported for study.

Training and Employment as a Veterinary Epidemiologist

Most veterinarians engaged in epidemiology have obtained advanced training beyond veterinary school in specialized graduate programs, either at veterinary schools or schools of public health. Program participants are often international, and the focus of research is generally on large animal diseases of economic consequence, such as bluetongue (a vector-borne viral disease of sheep, goats, cattle, and other species), foot-and-mouth disease (a viral disease impacting milk production in cattle, sheep, goats, and pigs), and Newcastle disease (a viral disease in chickens and other bird species). Research results include descriptions of geographic ranges of vectors of disease agents (such as flies, mosquitoes, fleas, and ticks), weather-related disease risks, and management and demographic risk factors.

Veterinary epidemiologists work in government agencies (county, state, national, and international), including the CDC and its Epidemic Intelligence Service, the USDA's Food Safety and Inspection Service and APHIS, and the Center for Veterinary Medicine within the Food and Drug Administration of the U.S. Department of Health and Human Services, in academia at veterinary and medical schools and at schools of public health, and in private industry and consulting. Veterinary epidemiologists at the CDC have studied waterborne disease outbreaks as well as injuries from second-hand smoking. Veterinarians in state government have been instrumental in identifying and tracing the sources of monkeypox in Indiana and the Midwest and rabies in California. Veterinarians with epidemiologic training make up a small cadre within the uniformed Public Health Service.

While many veterinary epidemiologists concentrate on large animal disease recognition and prevention, some veterinary epidemiologists have focused on small animal issues, including diseases of dogs such as bladder cancer, prostate cancer, and breed-specific diseases, and of cats such as hyperthyroidism and *Bartonella* spp. (the zoonotic agents responsible for cat scratch disease), as well as upper respiratory diseases of cats and dogs in animal shelters and their prevention with vaccines. Others have looked at the interface between human society and animals: hoarding of companion animals, injuries to owners of companion animals as well as the social and exercise benefits of dog ownership, reasons for shelter relinquishment of dogs and cats, infectious disease prevalence among feral cats, and risk factors for human failures to evacuate in disasters such as fires or floods. Veterinary epidemiologists also work at the interface between wildlife, domestic animal, and human populations to study disease transmission patterns and potential zoonotic risks. Examples include chronic wasting disease in wild and potentially farmed deer and elk (a disease similar to bovine spongiform encephalopathy (mad cow disease), severe acute respiratory syndrome in humans and wildlife sold in markets, raccoon roundworms and larval migrans disease, rabies in bats and companion animals, tuberculosis in cattle and humans, and avian influenza in migratory and farmed bird populations.

Organizations of veterinary epidemiologists include the Association for Veterinary Epidemiology and Preventive Medicine (formerly the Association of Teachers of Veterinary Public Health and Preventive Medicine), the International Symposium for Veterinary Epidemiology and Economics, the Canadian Association of Veterinary Epidemiology and Preventive Medicine, and the European-based Society for Veterinary Epidemiology and Preventive Medicine. Other groups include the National Association of State Public Health Veterinarians and the Veterinary Public Health special interest group within the American Public Health Association. Board certification in veterinary preventive medicine is available through the American College of Veterinary Preventive Medicine, with an additional subspecialization available in Epidemiology, and may be valuable for veterinarians in government agencies or large human medical institutions.

—Charlotte H. Edinboro

See also Foodborne Diseases; Public Health Surveillance; Zoonotic Disease

Further Readings

Dohoo, I., Martin, W., & Stryhn, H. (2004). *Veterinary epidemiologic research*. Charlottetown, Prince Edward Island, Canada: AVC.

Hugh-Jones, M. E., Hubbert, W. T., & Hagstad, H. V. (2000). *Zooneses: Recognition, control, and prevention*. Ames: Iowa State University Press.
Martin, W. S., Meek, A. H., & Willeberg, P. (1987). *Veterinary epidemiology: Principles and methods.* Ames: Iowa State University Press.
Salman, M. D. (Ed.). (2003). *Animal disease surveillance and survey systems: Methods and applications*. Ames: Iowa State University Press.
Schwabe, C. W. (1984). *Veterinary medicine and human health* (3rd ed.). Baltimore: Williams & Wilkins.
Slater, M. R. (2002). *Veterinary epidemiology (practical veterinarian)*. Oxford, UK: Butterworth-Heinemann.
Smith, R. D. (2005). *Veterinary clinical epidemiology* (3rd ed.). Boca Raton, FL: CRC Press.
Thursfield, M. (2005). *Veterinary epidemiology* (3rd ed.). Oxford, UK: Blackwell Science.

Violence as a Public Health Issue

In 1996, the 49th World Health Assembly declared violence to be a leading worldwide public health problem. The World Health Organization (WHO) defines violence as

> the intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation. (WHO, 2002, p. 5)

This entry describes the types of violence, its consequences, and its prevalence worldwide. It also examines characteristics of interpersonal violence as experienced by women, the elderly, and children and adolescents. Although violence as a criminal act remains primarily within the purview of judicial system, public health interventions can reduce the incidence of violence and its impact. A scientific approach to understanding the underlying causes of violence and risk factors is necessary to devise effective prevention programs and protecting health. Such an understanding can provide a basis for public health interventions that can reduce the incidence of violence and its impact.

Types of Violence

Violent acts include physical attacks, sexual abuse, psychological threat, and deprivation or neglect. Violence is categorized into three broad categories according to characteristics of the perpetrators: self-directed violence, interpersonal violence, and collective violence. Suicidal behavior and self-abuse are self-directed violence. Intimate partner violence, child abuse within a family, and violence between unrelated individuals are examples of interpersonal violence. Collective violence refers to that committed by larger groups of individuals (terrorist activities, insurgency, gang and mob violence) or by states (war).

Consequences of Violence

There are serious consequences of violence for individuals, families, communities, health care systems, and countries. Globally, violence is a major cause of death for people aged 15 to 44 years, and the economic and social costs of violence are substantial. More than 1.6 million people worldwide died as a result of self-inflicted, interpersonal, or collective violence in 2000, for an overall age-adjusted rate of 28.8 per 100,000 population. Nearly, half of these 1.6 million violence-related deaths were suicides, almost one third were homicides, and about one fifth were war-related. Among low- and middle-income countries, self-inflicted injuries and violence accounted for 8.9% of deaths. Suicide was the third leading cause of death for women in the age group 15 to 44 years.

Interpersonal Violence

Women, children, and elderly people are the major victims of interpersonal violence.

Violence Toward Women

Physical abuse by an intimate partner is the most common form of violence that women experience. It is estimated that between 10% and 52% of women experience some form of violence at the hands of their husband or male partner. Intimate partner violence, often referred to as domestic violence, includes the range of sexually, psychologically, and physically coercive acts used against adult and adolescent women by current or former male intimate partners without their consent.

A WHO multicountry study conducted in 2001 found that the lifetime prevalence of physical or sexual domestic violence varied widely, from 6% to as high as 71% across the countries studied. There is

growing recognition of the pronounced burden that violence against women exacts on their reproductive and overall health. Using a disability-adjusted life years approach, a World Bank study estimated that physical and sexual violence jointly account for 5% to 16% of healthy years of life lost by women of reproductive age in developing countries, ranking with obstructed labor, HIV, and cancer as causes of disability among women. Abused women are more likely to report gynecological morbidity, sexual problems, pelvic inflammatory disease, HIV, STD, urinary tract infections, and substance abuse. Studies found strong association between domestic violence and depressive disorders.

A substantial proportion of women experiencing physical violence also experience sexual abuse. In Mexico and the United States, studies estimate that 40% to 52% of women experiencing physical violence by an intimate partner have also been sexually coerced by that partner. Both physical and sexual domestic violence have been shown to be significantly associated with an increased risk of unintended pregnancy, short interpregnancy intervals, and lower contraceptive use, including condom use.

Effects of Violence During Pregnancy

Studies have shown that domestic violence does not abate during pregnancy and may possibly even be aggravated during this period. Violence is a major cause of death for pregnant and postpartum women in the United States, but still remains underreported. The percentage of women reporting domestic violence at the hands of their husband or intimate partner during pregnancy has varied between 0.9% and 20.1% in different studies. Deleterious maternal health correlates of abuse during pregnancy include depression, premature labor, injury, kidney infections, and antepartum hemorrhage. Homicide by partners, the most extreme form of abuse during pregnancy, has been identified as an important cause of maternal mortality in several countries.

Violence during pregnancy not only adversely affects the health of women but also affects birth outcomes. The adverse consequences of violence during pregnancy on birth outcomes such as low birthweight and prematurity have been extensively studied. Research has also shown that violence during pregnancy increases perinatal and infant mortality. The pathways through which domestic violence may lead to elevated risks of early childhood mortality are not fully understood. One possible pathway is through the direct effects of blunt physical trauma and resultant fetal death or subsequent adverse pregnancy outcome. A second potential pathway is through elevated maternal stress levels and poor nutrition, both associated with low birthweight or preterm delivery, well-known risk factors for adverse early childhood mortality outcomes. A third mechanism through which domestic violence may contribute to elevated risks of childhood mortality is through its deterrent effect on women's use of preventive or curative health services during pregnancy or delivery, or postnatally.

Risk Factors

Alcohol abuse is an important risk factor for interpersonal violence. Community-level cultural and contextual variables are also important determinants of intimate partner violence across cultures. Women's status, including personal autonomy, economic opportunity, political power, and the ability to participate in women's group activities, affects the risk of violence. Intergenerational exposure to domestic violence—witnessing family violence as a child—has shown to be one of the few consistent predictors with the risk of being a perpetrator (men) or victim (women) of domestic violence.

Elder Abuse

It is estimated that 4% to 6% of older people (above 65 years of age) experience violence, either at home or at institutional facilities (nursing homes, residential care, hospitals, and day care facilities). It is generally agreed that abuse of older people is either an act of direct abuse or of neglect and that it may be either intentional or unintentional. The abuse may be of a physical nature, it may be psychological (involving emotional or verbal aggression), or it may involve financial or other material maltreatment.

Children and Adolescents as Victims of Violence

Children are subjected to both physical and sexual abuse. The term battered child syndrome drew attention to the problems of physical abuse in young children in the late 1980s. In 1999, the WHO Consultation on Child Abuse Prevention defined violence to children as follows:

> Child abuse or maltreatment constitutes all forms of physical and/or emotional ill-treatment, sexual abuse, neglect or negligent treatment or commercial or other exploitation, resulting in actual or potential harm to the child's health, survival, development or dignity in the context of a relationship of responsibility, trust or power. (WHO, 2002, p. 59)

According to the World Health Organization, there were an estimated 57,000 deaths attributed to homicide among children below 15 years of age in 2000. Global estimates of child homicide suggest that infants and very young children are at greatest risk, with rates for the 0- to 4-year-old age group more than double those of 5- to 14-year-olds. The highest homicide rates for children below 5 years of age are found in the WHO African Region—17.9 per 100,000 for boys and 12.7 per 100,000 for girls.

Often, children are neglected by the parents and caregiver. There are many forms of child neglect, including failure to seek appropriate health care, noncompliance with health care recommendations, and deprivation of food. Abandonment, inadequate parental supervision, exposure to poor living conditions and hygiene, and lack of schooling may also be forms of child neglect. Parent's negligence to the exposure of children to drugs and alcohol remains a major public health concern.

There are serious adverse health consequences for child abuse. Abused children not only experience severe form of physical injuries, but also develop reproductive health problems, psychological and behavioral problems, including alcohol and drug habits, risk taking, antisocial and suicidal behaviors. They are more likely to experience unwanted pregnancy, STDs, poor academic performance, and drop out from the school.

Homicide and nonfatal injuries are the major causes of premature deaths and disability among youths. In 2000, an estimated 199,000 youth homicides (9.2 per 100,000 population) occurred globally. In other words, an average of 565 children, adolescents, and young adults between 10 and 29 years of age die each day as a result of interpersonal violence. Among all the homicide cases, three fourths are male victims and youth violence is the underlying factor.

—Saifuddin Ahmed

See also Child Abuse; Intimate Partner Violence; War

Further Readings

Robertson, L. S. (1992). *Injury epidemiology*. New York: Oxford University Press.

Rosenberg, M. L., Butchart, A., Narasimhan, V., Waters, H., & Marshall, M. S. (2006). *Interpersonal violence*. In D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, et al. (Eds.), *Disease control priorities in developing countries* (2nd ed., pp. 755–770). New York: Oxford University Press.

Winett, L. B. (1998). Constructing violence as a public health problem. *Public Health Reports*, *113*(6), 498–507.

World Health Organization, WHO Consultation. (1996). *Violence against women*. Geneva, Switzerland: Author.

World Health Organization. (2001). *Putting women first: Ethical and safety recommendations for research on domestic violence against women*. Geneva, Switzerland: Author.

World Health Organization. (2002). *World report on violence and health.* Geneva, Switzerland: Author.

World Health Organization. (2002). *Guide to United Nations resources and activities for the prevention of interpersonal violence.* Geneva, Switzerland: Author.

Vitamin Deficiency Diseases

Vitamins are required in the diet for human consumption. Compounds that are known collectively as vitamins are either insufficiently produced in the body or are not synthesized at all and yet are essential to normal body functions. In general, the concentrations of vitamins stored within the body vary. Some vitamins (such as A and B12) remain in the body in sufficient quantity such that a person may not develop a deficiency for months or years despite low dietary intake. However, other vitamin deficiencies may develop within a matter of weeks. Deficiencies of vitamins (and minerals) may be caused by, and may result in, a variety of diseases. This entry discusses the principal vitamin deficiencies and highlights important contributing factors and treatment regimens. Although the discussion of trace mineral deficiencies is beyond the scope of this entry, additional information and further readings on the topic are presented at the end of this section.

Water-Soluble Vitamins

These vitamins are soluble in water and generally cannot be stored within the body for an extended period of time. Vitamins that are naturally water

soluble must be constantly replenished through diet intake; otherwise, a deficiency of one or more of these vitamins can result.

Thiamine (B1)

Thiamine converts to a coenzyme in its active form, catalyzing the conversion of amino acids and the metabolism of carbohydrates. Primary food sources of thiamine include pork, beans, and other legumes, nuts, beef, and whole grains. Deficiency is primarily a result of poor dietary intake; however, in developed countries, thiamine deficiency more often results from chronic illness (e.g., cancer) or alcoholism.

A person in the early stages of thiamine deficiency will show symptoms of irritability and poor food intake. The full manifestation of thiamine deficiency is known as beriberi, and is characterized by muscle weakness and wasting, an enlarged heart (cardiomyopathy), pain in the legs and hands (peripheral neuropathy), weakness of one or more eye muscles (opthalmoplegia) and possible swelling of the extremities (edema). If the deficiency arises as a result of chronic alcoholism, the person may also experience central nervous dysfunction that may include loss of balance and psychosis.

The diagnosis of thiamine deficiency is confirmed by functional enzymatic assay. Treatment consists of intravenous or oral thiamine supplementation.

Riboflavin (B2)

Riboflavin is essential as a contributing factor in carbohydrate, fat, and protein metabolism. The most important sources of this vitamin are dairy products and whole grains. Other foods containing riboflavin include broccoli, legumes, eggs, fish, and other meats. Deficiency usually results from lack of dietary intake, and those who follow particularly strict diets (e.g., vegetarians and vegans) are at particular risk if they do not ensure adequate intake of vegetable sources of riboflavin.

The clinical manifestations of riboflavin deficiency include red or purple coloration of the tongue, cracking of the skin around the corners of the mouth, and dandruff. Additional indications of deficiency may include anemia, irritability, or other personality changes.

The diagnosis of riboflavin deficiency can be confirmed with laboratory testing of the blood or urine. Lab diagnostic tools are commonly used, because the clinical symptoms are nonspecific and similar to other vitamin deficiencies. Riboflavin deficiency is treated with supplementation of riboflavin.

Niacin (B3)

This vitamin catalyzes DNA repair and calcium transport reactions. The most common sources of niacin include protein-rich foods: dairy, meat, eggs, and beans. Another food source includes enriched flour, and daily intake of niacin in the United States usually exceeds FDA recommended guidelines. Deficiency results from poor dietary intake and is typically found among people who subsist on corn-based diets (as in parts of China, India, and Africa). Similar to thiamine deficiency, niacin deficiency in developed countries often is seen in chronic alcoholics. Niacin deficiency also occurs in people with congenital malabsorption or a chronic disease such as carcinoid syndrome, where niacin is insufficiently produced from its amino acid derivative.

Deficiency of niacin results in the clinical disease known as pellagra. This constellation of symptoms is manifested by a characteristic rash, a bright red tongue, diarrhea, disorientation, and possible memory loss. Severe niacin deficiency can result in death. The factors that may contribute to a niacin deficiency include alcoholism, pyridoxine (vitamin B6) deficiency, or riboflavin deficiency.

Diagnosis of this deficiency is usually based on clinical assessment, and treatment consists of oral niacin supplements.

Pyridoxine (B6)

Vitamin B6 is an essential cofactor for amino acid metabolism. It is also involved in the metabolism of certain other vitamins, including niacin. This vitamin is available in all food groups, but it is found in highest concentration in meats, whole grains, nuts, and legumes. Deficiency is usually due to alcoholism or use of specific medications for treatment of a chronic condition. The medications that can cause B6 deficiency include isoniazid (used for treating tuberculosis), L-dopa (used for the treatment of Parkinson's disease), and penicillamine (for patients with rheumatoid arthritis or Wilson's disease). It is rare that a person is born with a congenital disorder that would require B6 supplementation, but examples of such conditions include sideroblastic anemia and cystathionine beta-synthase deficiency.

Vitamin B6 deficiency results in symptoms similar to those seen in other B vitamin deficiencies, in

particular, skin changes such as dandruff and cracking of the skin. Additionally, severe deficiency can affect the nervous system, resulting in pain, seizures, and confusion. Anemia also may be associated with this vitamin deficiency.

Diagnosis of B6 deficiency is confirmed by measuring low levels of pyridoxal phosphate in the blood. Treatment consists of B6 supplementation.

Folate

Folate, also known as folic acid, is the coenzyme for metabolism of amino and nucleic acids and is essential in cell DNA synthesis. Folate is also involved in embryogenesis, and recent studies have shown that increased folate intake to be associated with a reduced risk of neural tube defects in the newborn. The primary sources of folate include raw vegetables and fruits. In addition, grain products sold in the United States are now enriched with folate. Deficiency commonly results from malnutrition but tends to manifest itself clinically only after several months of poor dietary intake. Those with folate deficiency are often severely undernourished and include persons suffering from chronic alcoholism, narcotic addiction, chronic hemolytic anemia, or intestinal malabsorption. Additionally, certain prescription drugs (sulfasalazine, pyrimethamine) can cause folate deficiency.

Clinical findings in folate deficiency include megaloblastic anemia (red blood cells are larger than normal), inflammation of the tongue, depression, diarrhea, and cracking at the edges of the mouth. In contrast to cobalamin deficiency, no neurological symptoms occur as a result of a deficiency in folate. Children born to women with folate deficiency have an increased risk of spinal cord malformations, including spina bifida.

Diagnosis of folate deficiency is based on the finding of sufficiently large blood cells (cell volume > 100 fl) on blood smear. Treatment usually consists of oral folic acid supplements. Women in their first 6 weeks of pregnancy, as well as women of child-bearing age, are recommended to supplement their diet with a multivitamin containing folate to reduce the risk of neural tube defects in the newborn.

Cobalamin (B12)

Vitamin B12, also known as cobalamin, catalyzes the reaction that forms methionine, which is a key factor in the metabolism of folate. Cobalamin can only be found in animal products: either meat or dairy foods. Deficiency is often the result of malabsorption due to chronic illness, such as pernicious anemia or disease of the small intestine. Deficiency can also be the result of poor dietary intake and can be seen in people taking certain prescription drugs or who are strict vegetarians.

Features of cobalamin deficiency include megaloblastic anemia (similar to folate deficiency), inflamed tongue, weight loss, and diarrhea. In addition, deficiency of cobalamin can eventually cause peripheral nerve degeneration resulting in numbness, pain, muscle weakness and imbalance. Some of these findings may be permanent if the deficiency is not treated immediately.

Diagnosis of cobalamin deficiency is based on characteristic red blood cell changes similar to those seen in folate deficiency. These laboratory findings are used in combination with clinical symptomatology. Cobalamin deficiency can also occur without the characteristic anemia and is quite common in elderly persons. Treatment of the deficiency includes cobalamin supplementation and treatment of the underlying disorder.

Vitamin C

Vitamin C, also known as ascorbic acid, functions as an antioxidant and facilitates several biochemical reactions, including iron absorption and norepinephrine synthesis. It also plays a role in maintaining connective tissue and is important in several enzyme systems, including the mixed-function oxidase system. Vitamin C is found in leafy vegetables, citrus fruits, and tomatoes. Deficiency usually results from poor dietary intake, although in developed countries, deficiency is usually seen in alcoholics and in those who consume less than 10 mg of vitamin C per day (the elderly or those with very low incomes).

The constellation of symptoms that result from vitamin C deficiency is known as scurvy. Symptoms include poor wound healing, bleeding gums, extensive bruising, and additional internal bleeding. Deficiency in children is often associated with reduced bone growth.

Diagnosis of vitamin C deficiency is based primarily on clinical assessment and is confirmed by low levels of white blood cells in the blood. Oral vitamin C supplementation is the usual treatment for deficiency.

Fat-Soluble Vitamins

These vitamins are soluble in lipids (fat) and are stored within the tissues of the body. Generally, once these vitamins are stored, they tend to remain in the body. However, if a person has too little fat intake, or they are unable to absorb fat adequately, those fat-soluble vitamins will also be poorly absorbed, leading to a vitamin deficiency.

Vitamin A

Vitamin A and its active metabolites are essential for normal vision, growth, and cell specialization. Food sources of vitamin A include fish, liver, brightly colored fruits, and green leafy vegetables. Children in particular are susceptible to deficiency, because sufficient levels of vitamin A are not supplied through either cow's or breast milk. The particular areas that have high levels of vitamin A deficiency include Southern Asia, South America, and parts of Africa. In developed nations, factors that contribute to deficiency include alcoholism, malnutrition, malabsorption syndromes, and infection. Additionally, patients taking mineral oil, neomycin, or cholestyramine are at risk of deficiency, because these medications interfere with vitamin A absorption.

Clinical manifestations of vitamin A deficiency include night blindness, impaired development, increased susceptibility to infection (as a result of immune dysfunction), and skin lesions. Children with vitamin A deficiency are at particular risk of death due to measles, diarrhea, and respiratory infection.

Diagnosis of vitamin A deficiency is confirmed with measurement of retinol levels in the blood. Treatment consists of vitamin A supplementation, and for those persons with night blindness, this can be given in the form of an intramuscular injection or in the form of oral supplements.

Vitamin D

Vitamin D is a combination of two active metabolites that act to maintain adequate levels of calcium and phosphorus in the blood. The principal sources of vitamin D are actually nondietary, as it can be produced within the skin during sun exposure. In response to ultraviolet radiation, precursor chemicals within the skin are cleaved into vitamin D. As a result, vitamin D is classified as a hormone rather than a vitamin. There are, however, dietary sources of vitamin D, including fish oil and fortified dairy and cereal products. Deficiency may be caused by a number of things: poor dietary intake, malabsorption, impaired production within the skin, the use of specific drugs (barbiturates, phenytoin, isoniazid, or rifampin), liver disease, kidney disease, or congenital disorders.

Vitamin D deficiency has several clinical manifestations. In young children, deficiency causes a characteristic retardation of bone growth known as rickets that results in lack of bone mineralization and is distinguished by bowed legs, scoliosis, and deformities of the chest wall. In developed countries, rickets due to vitamin D deficiency is quite rare, although rickets can occur as a result of several other diseases and disorders. In adults, deficiency of vitamin D is known as osteomalacia. This condition is often the result of poor dietary intake and is the result of impaired mineralization of bone. Small fractures in the scapula, pelvis, and femur are common in adults with this condition. In both children and adults, weakness of the upper arm and thigh muscles are characteristic.

Diagnosis of vitamin D deficiency can be done by assessing patient symptoms in combination with radiology and laboratory findings. Supplementation with vitamin D, in combination with calcium, is recommended to prevent vitamin D deficiency. Treatment of clinical deficiency varies based on the underlying disorder.

Vitamin E

Vitamin E is important as an antioxidant and also serves to regulate several enzyme pathways that inhibit blood clotting. This vitamin is actually a family of eight different vitamins; alpha-tocopherol is the most important type in humans. The important food sources of vitamin E include olive oil, sunflower and safflower oils, and wheat germ. Additional sources include meats, grains, nuts, leafy vegetables, and some fruits. Deficiency of vitamin E as a result of poor dietary intake does not exist; however, severe and prolonged malabsorption can result in deficiency. Additionally, those with cystic fibrosis or a congenital disorder are also susceptible to vitamin E deficiency.

This deficiency is characterized by a variety of signs and symptoms. Hemolytic anemia, muscle wasting, and retinal disease are some of the manifestations.

Severe deficiency can also cause degeneration of specific neural pathways, resulting in uncoordinated walking and loss of vibration and position sense in the lower limbs.

Diagnosis is based primarily on knowledge of the condition underlying the deficiency, since poor dietary intake does not cause deficiency. Treatment consists of oral supplementation with alpha-tocopherol.

Vitamin K

This vitamin is essential in the formation of blood clots. Primary food sources include green leafy vegetables, but vitamin K can also be found in meats and dairy products. Deficiency is often the result of an underlying disease; however, newborns are also susceptible because they lack fat stores at birth and receive only a minimal amount of vitamin K through breast milk. In adults, deficiency is seen in patients with poor fat absorption (i.e., diseases of the small intestine), liver disease, alcoholism, or using particular antibiotics.

Symptoms of deficiency are manifested by widespread bleeding. Severe deficiency can result in intracranial hemorrhage, intestinal bleeding, and poor wound healing.

Diagnosis of this vitamin deficiency is confirmed with an elevated prothrombin time, one of two laboratory tests that measure blood clot formation. For adults, treatment consists of oral vitamin K supplementation. Newborns are given an injection of vitamin K at birth as standard prophylaxis against deficiency.

—Ashby Wolfe

See also Alcohol Use; Birth Defects; Cancer; Child and Adolescent Health; Eating Disorders; Malnutrition, Measurement of; Pellagra

Further Readings

Babior, B. M., & Bunn, H. F. (2005). Megaloblastic anemias. In D. L. Kasper, E. Braunwald, A. S. Fauci, S. L. Hauser, D. L. Longo, & J. L. Jameson (Eds.), *Harrison's principles of internal medicine* (pp. 601–607). New York: McGraw-Hill.

Fairfield, K. M., & Fletcher, R. H. (2002). Vitamins for chronic disease prevention in adults: Scientific review. *Journal of the American Medical Association*, *287*(23), 3116–3126.

Institute of Medicine. (1999). *Dietary reference intakes for calcium, phosphorous, magnesium, vitamin D and flouride*. Washington, DC: National Academies Press.

Institute of Medicine. (2000). *Dietary reference intakes for thiamin, riboflavin, niacin, vitamin B6, folate, vitamin B12, pantothenic acid, biotin and choline*. Washington, DC: National Academies Press.

Institute of Medicine. (2000). *Dietary reference intakes for vitamin C, vitamin E, selenium and carotenoids*. Washington, DC: National Academies Press.

Institute of Medicine. (2001). *Dietary reference intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium and zinc*. Washington, DC: National Academies Press.

Office of Dietary Supplements, National Institutes of Health. (2005). *Vitamin and mineral supplement fact sheets*. Bethesda, MD: Author. Retrieved July 30, 2006, from http://ods.od.nih.gov/Health_Information/Vitamin_and_Mineral_Supplement_Fact_sheets.aspx.

Russell, R. M. (2005). *Vitamin and trace mineral deficiency and excess*. In D. L. Kasper, E. Braunwald, A. S. Fauci, S. L. Hauser, D. L. Longo, & J. L. Jameson (Eds.), *Harrison's principles of internal medicine* (pp. 403–411). New York: McGraw-Hill.

Volunteer Effect

The volunteer effect is a form of selection bias, sometimes referred to as self-selection or volunteer bias. This phenomenon is based on the idea that individuals who volunteer to participate in epidemiological studies are different in some way from the target population that they have originated from. The result of this effect is that the resulting measure between exposure and disease is distorted.

The selection of the study population is a critical part of any epidemiological study. One of the methods of selecting individuals is to recruit volunteer participants from the target population. However, the main risk in this method is that this subset of the population differs in some way from the general population. It has been suggested that individuals who volunteer to participate in epidemiological studies are often healthier than the general population, and it is their interest in health that motivates them to participate in such studies. As a result, healthier individuals who may be more knowledgeable about health in general will be overrepresented in the study population. A consequence of this is that the measure of effect could be distorted, and the results cannot be generalized to the larger population. An example of this was seen in the Iowa Women's Health Study that investigated the relationship between mortality and cancer in women. When participants and

nonparticipants were compared, it was found that there was a higher proportion of smokers in the nonparticipant group. This group also had a greater occurrence of smoking-related cancers. A similar situation is seen in studies examining treatment interventions and alcoholism. Often, individuals who volunteer for these studies have a different baseline level of alcoholism severity than those who do not volunteer.

Participants who volunteer may not actually be healthier than the target population, but it could be that they are interested in a study because they have a family history of the disease in question and are actually more at risk of the disease than the general population. This is another form of the volunteer effect. For example, a study investigating a potential intervention on the subsequent development of breast cancer that recruits women may actually find little effect simply because the volunteer participants are at a higher risk of developing breast cancer anyway.

Individuals who volunteer to participate in studies may also differ from the general population by education, gender, socioeconomic status, and other demographic characteristics. The result of this is that it can limit the generalizability of results to the general population. In situations where the possibility of volunteer bias is a concern, random sampling methods can be used to recruit study participants and therefore minimize the volunteer effect.

—Kate Bassil

See also Bias; Participation Rate; Sampling Techniques; Target Population; Validity

Further Readings

Rychtarik, R. G., McGillicuddy, N. B., Connors, G. J., & Whitney, R. B. (1998). Participant selection biases in randomized clinical trial of alcoholism treatment settings and intensities. *Alcoholism: Clinical and Experimental Research*, *22*(5), 969–973.

Stohmetz, D. B., Alterman, A. I., & Walter, D. (1990). Subject selection bias in alcoholics volunteering for a treatment study. *Alcoholism: Clinical and Experimental Research*, *14*(5), 736–738.

W

War

War is generally defined as armed conflict conducted by nation-states. The term is also used to denote armed action by a group within a nation against governmental or occupying forces; such armed actions are often termed *civil wars*, *wars of liberation*, or *revolutionary wars*. This entry examines the physical and psychological impact of war, terrorism, and other forms of armed violence, and the role that epidemiology can play in understanding and preventing violence.

War accounts for more death and disability than many major diseases. War destroys families, communities, and sometimes entire nations and cultures. War siphons limited resources away from health and other human services and damages the infrastructure that supports health. War violates human rights. The mindset of war—that violence is the best way to resolve conflicts—contributes to domestic violence, street crime, and other kinds of violence. War damages the environment.

An estimated 191 million people died during wars in the 20th century, more than half of whom were civilians. The exact figures are unknowable because of poor recordkeeping during wartime. Over the course of the 20th century, an increasing percentage of people killed in war were civilians; in some wars in the 1990s, possibly 90% of the people killed were noncombatant civilians. Most of them were caught in the crossfire of opposing armies or were members of civilian populations specifically targeted during war.

During most years of the past decade, there were approximately 20 wars, mainly civil wars that were infrequently reported by the news media in the United States. For example, more than 3.8 million people died in the civil war in the Democratic Republic of Congo during the past several years. As another example, more than 30 years of civil war in Ethiopia led to the deaths of 1 million people, about half of whom were civilians.

Several of these civil wars have been considered to be genocidal. In the Iraq War, which began in 2003, more than 2,500 U.S., British, and other Coalition troops had been killed as of March 2006, and more than 16,000 were wounded. An unknown number of Iraqi civilians have died as a result of the war; estimates range to more than 650,000, based on a cluster sample survey of Iraqi households that was performed in September 2006.

Many people survive wars only to be physically scarred for life. Millions of survivors are chronically disabled from injuries sustained during wars or the immediate aftermath of wars. Landmines are a particular threat; in Cambodia, for example, approximately 1 in 250 people is an amputee as a result of a landmine explosion. Approximately one third of the soldiers who survived the civil war in Ethiopia were injured or disabled; at least 40,000 had lost one or more limbs.

Millions more are psychologically impaired from wars, during which they have been physically or sexually assaulted, have been forced to serve as soldiers against their will, witnessed the death of family members, or experienced the destruction of their communities or entire nations. Psychological trauma may be demonstrated in disturbed and antisocial behavior, such as aggression toward others, including family members. Many military personnel suffer

from post-traumatic stress disorder (PTSD) after military service.

Rape has been used as a weapon in many wars. In acts of humiliation and revenge, soldiers have raped female family members of their enemies. For example, at least 10,000 women were raped by military personnel during the war in the 1990s in Bosnia and Herzegovina. The social chaos brought about by war also creates conditions permitting sexual violence.

Children are particularly vulnerable during and after wars. Many die as a result of malnutrition, disease, or military attacks. Many are physically or psychologically injured. Many are forced to become soldiers or sexual slaves to military personnel. Their health suffers in many other ways as well, as reflected by increased infant and young-child mortality rates and decreased rates of immunization coverage.

The infrastructure that supports social well-being and health is destroyed during many wars, so that many civilians do not have access to adequate and safe food, water, medical care, and/or public health services. For example, during the Persian Gulf War in 1991 and the years of economic sanctions that followed, the United Nations Children's Fund (UNICEF) estimated that at least 350,000 more children than expected died, with most of these deaths due to inadequate nutrition, contaminated water, and shortages of medicines. Many of these deaths were indirectly related to destruction of the infrastructure of civilian society, including health care facilities, electricity-generating plants, food-supply systems, water-treatment and sanitation facilities, and transportation and communication systems. The Iraq War has further damaged this health-supporting infrastructure.

In addition, many civilians during wartime flee to other countries as refugees or become internally displaced persons within their own countries, where it may be difficult for them to maintain their health and safety. Refugees and internally displaced persons are vulnerable to malnutrition, infectious diseases, injuries, and criminal and military attacks. Many of the 35 million refugees and internally displaced persons in the world were forced to leave their homes because of war or the threat of war.

War and the preparation for war divert huge amounts of resources from health and human services and other productive societal endeavors. This is true for many countries, including the United States, which ranks first among nations in military expenditures and arms exports, but 38th in infant mortality rate and 45th in life expectancy. In some less developed countries, national governments annually spend $10 to $20 per capita annually on military expenditures, but only $1 per capita on all health-related expenditures. The same types of distorted priorities also exist in more developed, or industrialized, countries.

War often creates a cycle of violence, increasing domestic and community violence in the countries engaged in war. War teaches people that violence is an acceptable method for settling conflicts. Children growing up in environments in which violence is an established way of settling conflicts often choose violence to settle conflicts in their own lives. Teenage gangs mirror the activity of military forces. Returning military servicemen commit acts of violence against others, including their wives or girlfriends.

War and the preparation for war have profound impacts on the environment. Specific examples include the following:

- destruction of mangrove forests in Vietnam by the defoliant Agent Orange or by bombs, with the resultant bomb craters filling with stagnant water and becoming breeding sites for mosquitoes that spread malaria and other diseases;
- destruction of human environments by aerial carpet bombing of major cities in Europe and Japan during World War II; and
- ignition of more than 600 oil well fires in Kuwait by retreating Iraqi troops in 1991.

Less obvious are other environmental impacts of war and the preparation for war, such as the huge amounts of nonrenewable fossil fuels used by the military before and during wars, and the environmental hazards of toxic and radioactive wastes, which can contaminate air, soil, and both surface water and groundwater.

In the early 21st century, new geopolitical, tactical, and technological issues concerning war are arising that have an impact on health and health services. These include use of new weapons, use of "suicide-bombers," use of drone (unmanned) aircraft and high-altitude bombers, and newly adopted United States policies on "preemptive" wars and on "usable" nuclear weapons. An example of the introduction of new weaponry has been the use of shell casings hardened with depleted uranium (DU), a toxic and radioactive material employed for its density and ability to ignite on impact. DU has been used by the United States in the Persian Gulf War, the wars in the Balkans

and Afghanistan, and (also by the United Kingdom) in the Iraq War. An estimated 300 metric tons of DU remain in Iraq, Kuwait, and Saudi Arabia. Because of the lack of data on the number of troops and civilians exposed, the levels of exposure, and the short-term and long-term consequences, epidemiological studies of the health impact have been fragmented and inadequate.

Terrorism

Closely related to war are forms of armed violence, often termed *terrorism*, in which individuals or groups, often clandestine, use politically motivated violence or the threat of violence, especially against civilians, with the intent to instill fear. These individuals or groups may be seen as "resistance fighters" or "freedom fighters" by those who support their actions or as "terrorists" by those who oppose them. Nation-states generally refuse to consider as "terrorism" their own governmental military actions that target civilians with the intent to instill fear.

Although there is much discussion about the possibility of chemical, biological, radiological, and nuclear weapons being used in terrorist attacks, the vast majority of terrorist attacks have used conventional weapons, mainly explosives. Since the 9/11 attacks in the United States and the transmittal of anthrax bacteria through the U.S. mail shortly thereafter, the U.S. government has conducted what it terms a *war on terror*. This has led to the restriction of civil liberties of U.S. citizens and the arrest, detention without charges, and violation of the human rights of noncombatant citizens of other countries whom the U.S. government has suspected of being "terrorists" or having "terrorist" ties. The United States has given much attention and devoted many human and financial resources to terrorism preparedness since 2001, often at the cost of reducing attention and resources for major public health problems, such as tobacco, alcohol, and other forms of substance abuse; gun-related deaths and injuries; and HIV/AIDS, cardiovascular disease and cancer.

The Roles of Epidemiology

Epidemiology has an important role in documenting and understanding the adverse health effects of war, terrorism, and other forms of armed violence, and thereby helping to prevent these effects. Epidemiologic surveillance, research, and analysis can, for example, help document and describe the morbidity and mortality due to the direct and indirect effects of war. Epidemiology can also elucidate the adverse health effects of a variety of chemical, biological, physical, and psychosocial exposures that occur during war among both combatants and noncombatants. It can also help us better understand the long-term physical, psychological, and social consequences of war. Epidemiologists have responsibilities not only to perform this type of surveillance, research, and analysis but also to share their findings with the scientific community and the general public. The results of, and conclusions drawn from, epidemiologic surveillance and research on war, terrorism, and other forms of armed violence are useful for developing and implementing policies to prevent armed violence and its adverse health consequences and for better providing for needs of those adversely affected.

—Victor W. Sidel and Barry S. Levy

See also Bioterrorism; Firearms; Genocide; Immigrant and Refugee Health Issues; Post-Traumatic Stress Disorder; Violence

Further Readings

Austin, J. E., & Bruch, C. E. (Eds.). (2000). *The environmental consequences of war*. Cambridge, UK: Cambridge University Press.

Cukier, W., & Sidel, V. W. (2006). *The global gun epidemic: From Saturday night specials to AK-47s*. Portsmouth, NH: Praeger Security International.

Levy, B. S., & Sidel, V. W. (Eds.). (2008). *War and public health* (2nd ed.). New York: Oxford University Press.

Levy, B. S., & Sidel, V. W. (Eds.). (2003). *Terrorism and public health: A balanced approach to strengthening systems and protecting people*. New York: Oxford University Press.

Levy, B. S., & Sidel, V. W. (Eds.). (2006). *Social injustice and public health*. New York: Oxford University Press.

Power, S. A. (2002). *A problem from hell: America and the age of genocide*. New York: Basic Books.

Sidel, M. (2004). *More secure, less free? Antiterrorism policy and civil liberties after September 11.* Ann Arbor: University of Michigan Press.

Stockholm International Peace Research Institute. (2006). *SIPRI yearbook 2006: Armaments, disarmament and international security*. Oxford, UK: Oxford University Press.

Taipale, I., Makela, P. H., Juva, K., Taipale, V., Kolesnikov, S., Mutalik, R., et al. (Eds.). (2002). *War or health? A reader*. London: Zed Books.

Waterborne Diseases

In the mid-1800s, physician John Snow recommended removal of the handle from a water pump in a London neighborhood, ending an outbreak of cholera that had killed more than 500 people in a 10-day period. In the past 150 years, much progress has been made in understanding and preventing the transmission of infectious waterborne diseases. Even so, waterborne pathogens continue to be transmitted to humans via recreational water contact and contaminated drinking water supplies throughout the world, resulting in morbidity and mortality that is preventable. Infections that result from contact with waterborne pathogens can result in either endemic or epidemic disease. Most waterborne diseases are endemic in a population—there is some baseline level of disease that occurs normally in a population. An epidemic is defined when cases occur in excess of the normal occurrence for that population.

The Centers for Disease Control and Prevention (CDC) estimates that each year infectious waterborne diseases account for approximately 2 billion episodes of diarrhea leading to an estimated 1 million deaths worldwide. Most of these diarrheal deaths occur among children in developing countries, but the elderly and immunocompromised populations are also at an increased risk for waterborne infections. Table 1 lists the primary agents of infectious waterborne disease worldwide. These bacteria, viruses, and protozoa typically cause gastrointestinal symptoms, although some may cause a variety of other health effects, including neurological disorders (e.g., primary amoebic meningoencephalitis caused by *Naegleria fowleri*) and respiratory illness (e.g., pneumonia caused by *Legionella* species.

The CDC reports biennial estimates of waterborne outbreaks attributed to bacteria, viruses, and protozoa in the United States. In 2001 and 2002, 31 drinking water outbreaks were reported in 19 states, resulting in 1,020 cases of illness. Sixty-five recreational water outbreaks were reported by 23 states, resulting in 2,536 cases of illness. The number of cases from recreational and drinking water outbreaks for selected pathogens is shown in Table 2. Endemic waterborne disease is more difficult to quantify, but the 1996 Amendments to the Safe Drinking Water Act mandated that the U.S. Environmental Protection Agency (EPA) and the CDC jointly develop a national estimate of waterborne disease occurrence in the United States. Preliminary estimates by the EPA indicate that of all cases of acute gastrointestinal illness occurring in the U.S. population served by community water systems, approximately 8.5% may be attributed to their drinking water ($\sim$ 16.4 million cases per year).

Waterborne Disease Surveillance Systems

Public health surveillance systems are critical for detection and control of waterborne diseases. In many developed countries, governmental systems are in place requiring laboratories, hospitals, and clinicians to report certain diseases to a central agency. In the United States,

Table 1 Primary Agents of Infectious Waterborne Diseases

Bacteria	*Viruses*	*Protozoa*
Campylobacter jejuni	Hepatitis A	*Balantidium coli*
Escherichia coli	Norwalk virus	*Cryptosporidium* species
Francisella tularensis	Rotavirus	*Cyclospora cayetanensis*
Legionella species		*Entamoeba histolytica*
Leptospira species		*Giardia* species
Mycobacterium species		*Naegleria fowleri*
Salmonella typhi		
Shigella species		
Vibrio cholerae		

Source: Heymann (2004).

Table 2 Two-Year Case Counts for Primary Agents Associated With Selected Infectious Waterborne Disease Outbreaks in the United States 2001–2002

	*U.S. Reported Cases**	
Pathogen	*Recreational Water*	*Drinking Water*
Bacteria		
Campylobacter jejuni	–	25
Escherichia coli	78	2
Legionella species	68	80
Shigella species	78	–
Viruses		
Norwalk virus	146	727
Protozoa		
Cryptosporidium species	1474	10
Giardia species	2	18
Naegleria fowleri	8	2
Unknown	145	117

Sources: Blackburn et al. (2004) and Yoder et al. (2004).
* Does not include infectious waterborne outbreaks due to *Pseudomonas aeruginosa*, *Bacillus species*, *Staphylococcus species*, or Avian schistomsomes.

individual states require different "notifiable diseases" to be reported to public health officials, which are later compiled in the National Notifiable Diseases Surveillance System (NNDSS) by the CDC and the Council of State and Territorial Epidemiologists (CSTE). These notifiable diseases include a variety of bioterrorism-related conditions, as well as many potential waterborne diseases, such as cryptosporidiosis, giardiasis, and legionellosis. These types of data can provide some insight into endemic disease, but reporting requirements are not limited to waterborne infections. For example, giardiasis may be transmitted by contaminated water or food, or it may be sexually transmitted. The NNDSS is an example of a passive surveillance system since these data are voluntarily reported to the CDC by state, territorial, and local public health agencies.

In contrast to passive surveillance, active surveillance relies on solicitation of disease reports from laboratories, hospitals, and clinicians (e.g., the CDC's FoodNet Program). Active surveillance overcomes the problem of underreporting by health care providers and can be specifically tailored to identify secondary complications associated with infections. Syndromic surveillance is a specific type of surveillance that may be useful in detecting waterborne disease outbreaks. Syndromic surveillance involves the systematic gathering of population behavior and health data, such as antidiarrheal sales or emergency room visits, to identify anomalous trends. Syndromic surveillance may increase timely detection of outbreaks before laboratory or clinically confirmed diagnostic information is available, but empirical evidence of the efficacy of syndromic surveillance to mitigate the effects of waterborne disease outbreaks through earlier detection and response is lacking.

In the United States, epidemics may be identified through the Waterborne Diseases Outbreak Surveillance System (WBDOSS)—a database of drinking and recreational water outbreaks maintained by the CDC, the EPA and the CSTE. The ability of the WBDOSS to accurately capture the disease burden associated with waterborne outbreaks may be limited due to the difficulty of outbreak detection and inherent limitations of passive surveillance system data. These limitations include the following:

- Waterborne infectious disease often manifests as gastroenteritis or other self-limiting illnesses with mild symptoms; therefore, only a small proportion of cases may seek medical attention.
- Waterborne outbreaks that result in mild symptoms, have low attack rates, or are not caused by an easily identifiable etiologic agent may go unrecognized because the medical community never has an opportunity to make a formal diagnosis.
- The ability to detect waterborne outbreaks depends on the capacity of local public health agencies and laboratories to identify cases of illness and link these in a timely manner to a common source of exposure or to an etiologic agent.
- Water service system type, source water type, and size of the population served by the contaminated water system may affect the likelihood that an outbreak is attributed to a waterborne source.

Prevention of Waterborne Diseases

Drinking water sources are vulnerable to contamination from point and nonpoint sources. Point sources, such as improperly treated sewage discharged to a water source, can lead directly to infectious waterborne

disease transmission. Nonpoint sources, such as agricultural or urban runoff, can introduce pathogens to surface waters or to groundwater under the influence of surface water.

Prevention of waterborne disease transmission in public drinking water systems can be accomplished in several ways, including (1) watershed management and protection of source waters, (2) use of treatment techniques intended to remove or inactivate pathogens prior to distribution, (3) presence of residual disinfectant and implementation of routine pipe flushing programs to prevent growth of disease-causing organisms in the water distribution system, and/or (4) implementation of measures to prevent cross-connections between wastewater and drinking water in the distribution system. Often, a combination of actions is employed to help ensure multiple barriers of protection of water supplies. If monitoring results suggest contamination of drinking water, actions can be taken to prevent waterborne spread. These include issuing of "boil water" advisories, implementing additional treatment, or ceasing use of the water source.

In recreational waters, disease transmission can be prevented by careful monitoring and sanitation practices. In public swimming pools and spas, practices that prevent transmission of disease include (1) maintenance and monitoring of disinfectant and pH levels, (2) policies to require showers prior to entering the pool or spa, and (3) measures to prevent accidental fecal release. Prevention of accidental fecal release is especially important at facilities that serve young children; measures such as implementation of bathroom breaks as part of the pool schedule and providing separate pools for young children are effective at reducing fecal-oral transmission. In the United States, states and local jurisdictions enforce environmental health laws governing sanitation practices for pools, spas, and other recreational waters. In marine and fresh recreational waters, eliminating potential sources of pathogens such as wastewater discharge near public beaches is an important step to help reduce recreational waterborne infections. Water monitoring and beach closure postings are also essential for reducing potential exposure to pathogens.

Routine monitoring is another effective measure employed to ensure that drinking and recreational water is safe to use. In the United States, regulations require monitoring for biological, chemical, and radiological parameters in public drinking water supplies. Routine monitoring for both drinking and recreational waters usually includes tests for coliform bacteria, which are used to indicate the possibility of fecal contamination that could spread waterborne diseases. Additional monitoring and treatment requirements for more specific pathogens, such as *Cryptosporidium* and *Giardia*, are required for drinking waters with surface water sources in the United States. In recreational waters, monitoring for specific pathogens is expensive and may require specialized equipment not readily available, so indicator bacteria such as total coliform, *Escherichia coli*, and *Enterococci* are typically used as surrogates for the pathogenic organisms that might be present. One drawback to the use of indicators is that they may not always be well correlated with specific pathogen occurrence or risk of illnesses. An additional disadvantage of current pathogen testing methods (i.e., bacteria, viruses, and protozoa) is that these tests require 24 hr or more to generate results, so there is a lag in time between when exposure is potentially occurring and when any preventative actions may be initiated. These problems are likely to diminish as rapid testing methods and new indicators are developed and become more widely available.

—June M. Weintraub and J. Michael Wright

See also Centers for Disease Control and Prevention; Notifiable Disease; Outbreak Investigation; Public Health Surveillance; Snow, John

Further Readings

Berger, M., Shiau, R., & Weintraub, J. M. (2006). Review of syndromic surveillance: Implications for waterborne disease detection. *Journal of Epidemiology and Community Health*, *60*, 543–550.

Blackburn, B. G., Craun, G. F., Yoder, J. S., Hill, V., Calderon, R. L., Chen, N., et al. (2004). Surveillance for waterborne-disease outbreaks associated with drinking water: United States, 2001–2002. *Morbidity and Mortality Weekly Report Surveillance Summaries*, *53*, 23–45.

Centers for Disease Control and Prevention. (2006, June). 12 Steps for prevention of recreational water illnesses (RWIs). Retrieved August 11, 2006, from http://www.cdc.gov/healthyswimming/twelvesteps.htm.

Fewtrell, L., & Colford, J. M., Jr. (2005). Water, sanitation and hygiene in developing countries: Interventions and diarrhoea—a review. *Water Science and Technology*, *52*, 133–142.

Gerba, C. P., Hunter, P. R., Jagals, P., & Peralta, G. L. (Eds.). (2006). Estimating disease risks associated with drinking water microbial exposures. *Journal of Water and*

Health, *4*(Suppl. 2), 1–2. Retrieved July 27, 2007, from http://www.epa.gov/nheerl/articles/2006/waterborne_disease.html.

Heymann, D. L. (Ed.). (2004). *Control of communicable diseases manual* (18th ed.). Washington, DC: American Public Health Association.

U.S. Environment Protection Agency. (1986). *Ambient water quality criteria for bacteria: 1986*. EPA440/5-84-002. Washington, DC: Office of Water Regulations and Standards Division.

Wade, T. J., Calderon, R. L., Sams, E., Beach, M., Brenner, K. P., Williams, A. H., et al. (2006). Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness. *Environmental Health Perspectives*, *114*, 24–28.

Wade, T. J., Pai, N., Eisenberg, J. N., & Colford, J. M., Jr. (2003). Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. *Environmental Health Perspectives*, *111*, 1102–1109.

Wheeler, J. G., Sethi, D., Cowden, J. M., Wall, P. G., Rodrigues, L. C., Tomkins, D. S., et al. (1999). Study of infectious intestinal disease in England: Rates in the community, presenting to general practice, and reported to national surveillance. *British Medical Journal*, *318*, 1046–1050.

Yoder, J. S., Blackburn, B. G., Craun, G. F., Hill, V., Levy, D. A., Chen, N., et al. (2004). Surveillance for waterborne-disease outbreaks associated with recreational water: United States, 2001–2002. *Morbidity and Mortality Weekly Report Surveillance Summaries*, *53*, 1–22.

Women's Health Issues

Women constitute 51% of the U.S. and 50% of the world's population. The familiar paradox of women's health, that women live longer than men but have poorer health throughout their lives, continues to be true. In most developed countries, women live about 6.5 years longer than men, on average. Women's mortality advantage has been reduced somewhat in recent years, reflecting decreased heart disease and cancer death rates among men, but not women. Women's morbidity and mortality are influenced by a variety of conditions that preferentially affect them, as noted below.

Women's health is a broad topic that has gained recognition as a discipline. Multiple definitions have been proposed with more recent definitions focusing on the variety of factors that influence a woman's health during her life span. For example, the National Academy on Women's Health Medical Education defines women's health as devoted to facilitating the preservation of wellness and prevention of illness in women; it includes screening, diagnosis, and management of conditions that are unique to women, are more common in women, are more serious in women, or have manifestations, risk factors, or interventions that are different in women. As a discipline, women's health also recognizes (a) the importance of the study of gender differences; (b) multidisciplinary team approaches; (c) the values and knowledge of women and their own experience of health and illness; (d) the diversity of women's health needs over the life cycle and how these needs reflect differences in race, class, ethnicity, culture, sexual preference, and levels of education and access to medical care; and (e) the empowerment of women to be informed participants in their own health care.

One issue for women's health research, reporting, and interpretation is the conflation of the terms sex and gender. Sex is a biological phenomenon, whereas gender is a social construction resulting from culturally shaped norms and expectations for behavior. Biological differences may not be taken into account because they are regarded as a product of cultural influences; on the other hand, differences in the socialization of women are sometimes not taken into account in the exploration of sex differences. Thus, the conflation of sex and gender is problematic and may obscure questions such as whether women experience pain differently than men—a sex difference—or have been trained to seek care more frequently—a gender difference. Nonspecific use of the terms *sex* and *gender* has had an impact on the equitable treatment of women in biomedical research and clinical medicine and on how sex differences have been conceived, studied, and addressed in biomedicine.

Not long ago, women were routinely excluded from large-scale clinical trials. For instance, most trials for the prevention of heart disease studied middle-aged males and excluded women because of a complex and sometimes conflicting set of assumptions. On the one hand, women's hearts were assumed to be the same as men's; therefore, it was unnecessary to include both sexes in the trial. On the other hand, women were assumed to be sufficiently different from men (because of hormonal and reproductive factors, for instance) to justify their exclusion from trials. This paradoxical attitude toward sex difference in clinical trials persists

today and highlights the complexities of addressing sex differences in health. Human subject guidelines, and the National Institutes of Health grant requirements, have mandated women's inclusion in clinical trials and research, yet the question remains as to how similarities and differences between men and women will be conceived, studied, and compared. The way research questions are posed will dictate the answers investigators obtain and will have implications for women's treatment and overall health. For instance, it is well recognized that woman are diagnosed with depression in greater numbers than are men. Is this difference a sex difference (Are women at higher risk of depression by virtue of being women?), a gendered difference (Are women more likely to seek care for depression than men?), or is it something else (Are doctors more likely to diagnose depression in women than in men?)? Precision of language and thought demands that we focus on the ways we measure and report differences between men and women and allows us to specify what these differences mean for biomedical research and ultimately for patients care.

In epidemiology, the subject of women's health is frequently organized by life course, by anatomical feature, or by chronic/infectious states. This entry incorporates and supplements these schemas by considering women's health issues by life stage, by anatomy, and by social origin, focusing primarily on the experience of women in the United States.

Women's Health Issues: By Life Stage

Adolescence and Menarche

Adolescence is a period of tremendous change in women's lives, resulting in physical and sexual maturation. The rate at which maturation occurs depends on ethnicity, nutritional status, and physical activity. According to the National Survey of Family Growth (1988), median age at menarche for American women in 1988 was 12.5 years. Health issues of greatest concern during adolescence include development through puberty; menarche, or first menstrual bleeding; menstrual disorders; premenstrual syndrome; and adolescent pregnancy.

Reproductive Years

Women's reproductive years span the time from puberty to menopause, usually characterized as from 15 to 44 years. Hormones are clearly connected with health during the reproductive years and fluctuate cyclically, except during pregnancy and lactation. During the reproductive period of a woman's life, the health issues of greatest concern include the following: sexual dysfunction; fecundability, or ability to conceive; contraception; abortion, both induced and spontaneous; pregnancy; labor and delivery; infertility; assisted reproductive technology; adverse birth outcomes, such as preterm birth, low birthweight, small for gestational age, and birth defects; myoma and leiomyomata (benign tumors of the uterine muscle, also known as uterine fibroids); abnormal uterine bleeding; pelvic pain; pelvic floor relaxation; and endometriosis (which results from endometrial cells growing outside the uterus).

Perimenopause, Menopause, and Postmenopause

The most frequently used definition of menopause comes from the World Health Organization and defines menopause as the permanent cessation of menstruation resulting from the loss of ovarian follicular activity. Perimenopause, or climacteric, includes the period prior to menopause when endocrinological, clinical, and biological changes associated with menopause are occurring. Postmenopause is defined as the period after menopause and begins 12 months after spontaneous amenorrhea. Health tends to decline as women age, so many of the health issues we identify as chronic health conditions become of concern during this period of life. Most of these will be considered in the following section on anatomy. One health decision directly associated with reproductive senescence is whether to take hormone replacement therapy (HRT). HRT has been implicated as both beneficial and harmful to the health and quality of life for menopausal and postmenopausal women. Definitive information from randomized controlled trials is forthcoming.

Elderly and Frail Elderly Years

Almost 60% of people above the age of 65 years in 1999 were women. Although older women face many risks related to chronic disease as they age, most are healthy well into their later years. In 1994, 75% of women aged between 65 and 74 rated their health as good to excellent. The presence of chronic

conditions increases with age. Nearly half of women aged 75 or older reported activity limitations resulting from chronic conditions. Two nonchronic health conditions faced by women in older age are sensory impairment (age-related loss of vision, hearing, and chemical senses such as taste and smell) and Alzheimer's disease, which is a degenerative disease of the brain, associated with the development of abnormal tissues and protein deposits in the cerebral cortex and characterized by confusion, disorientation, memory failure, speech disturbances, and the progressive loss of mental capacity.

Women's Health Issues: By Anatomy

Another organizational schema for discussing women's health issues is by anatomical feature. Men and women share the bulk of their anatomy, but the extent to which these anatomical structures operate, age, and degenerate identically is unknown. Most of the conditions grouped by anatomy are chronic in nature and will subsequently be more prevalent in older ages than among younger women.

Bone and Musculoskeletal Disorders

Two bone/skeletal disorders of particular concern for women's health are osteoporosis and osteoarthritis. Osteoporosis is a disease of bone in which the bone mineral density (BMD) is reduced, bone micro-architecture is disrupted, and the noncollagenous proteins in bone is altered, resulting in increased propensity to fracture. Arthritis is a group of conditions where there is damage caused to the joints of the body; it is the leading cause of disability among women above the age of 65 years.

Cancer

In the United States, cancer is the second leading cause of death for both men and women. Although there are more than 40 forms of cancer, six sites accounted for more than 60% of all deaths due to cancer among American women: breast, lung, colorectal, cervical, uterine, and ovarian. Cancer rates have been relatively stable among white women but have increased among nonwhites. Endogenous hormones, such as estrogen, have been implicated in the growth and development of several of these cancers (breast, cervical, uterine, and ovarian).

Cardiovascular Disease (CVD)

CVD is the leading cause of death among women, including young women. Age-specific CVD death rates lag about 10 years behind those for men. Until 1990, there was little information on women and heart disease because women were excluded from almost all major CVD randomized trials. It is unclear what role women's hormones, including those taken via hormone replacement therapy, play in protecting women against CVD.

Digestive Diseases

Gastrointestinal disorders represent some of the most poorly understood conditions in the female body. Many share a constellation of symptoms that currently have little structural or biochemical explanation. Despite the lack of clear diagnostic and physiologic understanding, digestive diseases have significant impacts on women's health. Relevant gastrointestinal issues for women's health include irritable bowel syndrome, characterized by abdominal pain or discomfort associated with a change in bowel function; gallstone disease (the formation of stones in the gallbladder or bile ducts); peptic ulcer disease, which refers to a discrete mucosal defect in the portions of the gastrointestinal tract exposed to acid and pepsin secretion; dyspepsia or stomach pain; aerophagia (swallowing too much air and the resultant belching), rumination (regurgitation of previously consumed food); functional constipation; functional diarrhea; and functional abdominal pain.

Immunity and Autoimmune Diseases

Diseases of the immune system demonstrate sex differences in incidence, natural history, and disease severity. These differences are illustrated in the cytokines measured, the degree of immune responsiveness, and the presence of sex hormones. Furthermore, the degree of immune responsiveness differs between men and women. The common theme connecting autoimmune disorders is the presence of an autoimmune response based on genetic risk factors that interact with environmental triggers. These triggers might be exposure to infection, chemicals, physical stress, or other unknown exposures. Immune and autoimmune diseases that affect women's health include asthma; allergic diseases; multiple sclerosis, a disease that affects the central nervous system in

which the protective myelin around nerve fibers is lost, leaving scar tissue called sclerosis; rheumatoid arthritis; thyroid diseases; systemic lupus erythematosus (a chronic, inflammatory autoimmune disease that targets various organs); and Sjögren's syndrome, a chronic disease in which white blood cells attack the moisture-producing glands, including the mouth, eyes, and organs.

Oral Health

Changing hormonal levels over the life course, particularly during puberty, menses, and menopause, have been implicated in frequency of cold sores, gingivitis during puberty, dry mouth, taste changes, increased risk of gum disease, and bone weakness around menopause. Oral health issues of particular concern for women's health include periodontal disease and temporomandibular disorders, a variety of disorders that causes pain and tenderness in the temporomandibular joint.

Urologic and Kidney Conditions

Urinary incontinence, the inability to hold urine until arriving at a toilet, is often temporary and always results from an underlying medical condition. Women experience incontinence twice as often as men. Pregnancy and childbirth, menopause, and the structure of the female urinary tract account for this difference. Another urologic condition is interstitial cystitis, which is a long-lasting condition also known as painful bladder syndrome or frequency-urgency-dysuria syndrome, which results from the wall of the bladder becoming inflamed or irritated. This irritation affects the amount of urine the bladder can hold and causes scarring, stiffening, and bleeding in the bladder. Diabetes is the most important of the kidney conditions facing women's health. The overall age-adjusted prevalence of physician-diagnosed diabetes is the same in women and in men (5.2% vs. 5.3%), but its sequelae, including kidney disease and end-stage renal disease, are serious. Diabetes is the seventh leading cause of death in the United States.

Other Conditions

A number of poorly understood conditions adversely affect women's health in the United States, including migraines, fibromyalgia, and chronic fatigue syndrome.

Migraine headache is a severe pain felt on one side, and sometimes both sides, of the head, and lasting from a few hours up to 2 days. Migraines may be accompanied by nausea, vomiting, and light sensitivity. Migraines are more common in women than in men and are most common in women between the ages of 35 and 45. Hormones have been implicated in migraine prevalence; more than half of women with migraine report having them right before, during, or after their period.

Fibromyalgia, formerly known as fibrositis, is a chronic condition causing pain, stiffness, and tenderness of the muscles, tendons, and joints. Fibromyalgia is also characterized by restless sleep, awakening feeling tired, fatigue, anxiety, depression, and disturbances in bowel function. Its cause is currently unknown. Fibromyalgia affects predominantly women between the ages of 35 and 55.

Chronic fatigue syndrome (CFS) is an illness characterized by profound disabling fatigue lasting at least 6 months accompanied by symptoms of sleep disturbance, musculoskeletal pain, and neurocognitive impairment. A unifying etiology for CFS is yet to emerge. More women than men are diagnosed with CFS, but it is unclear if this differential results from women seeking care for CFS more frequently than men or some underlying predilection.

Neuroscience and Women's Health

In addition to the health issues noted above, the ways in which neuroscience intersects women's health, for example, sex differences in cognition, sex differences in drug behavior, sex differences in manifestations of brain disorders, sex differences in sensory perception and pain, sex differences in balance and the vestibular system, and so on, will remain an important area for improving understanding of women's health.

Women's Health Issues: Social Origin

Many issues that affect women's health result from the intersection of women and social interactions or social structure.

Sexually Transmitted Diseases

In the United States from 1973 to 1992, more than 150,000 women died of causes related to sexually transmitted diseases (STDs), including human

immunodeficiency virus (HIV). Women bear a rate and burden of STD that is disproportionate compared with men. Much of this overrepresentation can be attributed to increased biological and behavioral susceptibility. As women age, the female ecology becomes less susceptible to colonization by STD-causing agents, but among all women, adolescents are at particularly high risk for STDs. STDs of particular concern for women's health include chlamydia, gonococcal infection, syphilis, genital herpes, human papilloma virus (HPV, which causes genital warts and is associated with the development of cervical and other genital tract squamous precancerous lesions and cancers), HIV/acquired immunodeficiency syndrome (AIDS), and pelvic inflammatory disease (PID, a general term that describes clinically suspected infection with resultant inflammation of the female upper reproductive tract, including the fallopian tubes, ovaries, and uterine lining).

Mental Health Issues

Men and women differ in the prevalence and severity of mental illness, but the validity of this finding is undermined by mental illness diagnosis (based largely on subjective symptoms), gender-based behavioral training (women more likely to seek help), and physician assumptions (physicians may expect to see more mental disorders among women). The mental health picture is further complicated by hormonal effects on brain function, brain maturation, and pharmacological response. Despite the limited information on mental health issue etiology, ascertainment, diagnosis, and treatment, multiple mental health issues affect women's health, including the following: depression, gender and mood disorders, anxiety disorders, post-traumatic stress disorder (PTSD), eating disorders and body image concerns (including anorexia, bulimia, overweight, obesity), and addictive disorders (tobacco, alcohol, and other drug abuse). As currently understood, many of these mental health issues preferentially influence women's health.

Violence

Research indicates that women experience many forms of violence, and both physical and psychological violence can occur from perpetrators who are strangers, acquaintances, family members, or partners. In 1995, the National Crime Victimization Survey, conducted by the U.S. Department of Justice, estimated that approximately 4% of women in the United States reported experiences of violent crime during the past year. The estimated number of incidents perpetrated against women is more than 4.7 million, and in 1994, women were raped at a rate of 4.6 per 1,000 women. Violence has long-reaching effects, including both mental and physical health consequences.

Global Issues in Women's Health

Women living outside the United States, particularly those living in nondeveloping and developing countries, are faced with many of the same health issues as women living in the United States, but some have different concerns as well. Three areas that are of particular concern for women living outside the United States are high maternal morbidity and mortality associated with childbearing, female genital mutilation/circumcision, and HIV/AIDS.

It is important to note that women's health issues are differentially distributed; not all issues affect all women equally. In the United States and elsewhere, social hierarchies exist that contribute to significant disparities between racial groups, particularly between white and nonwhite women. In addition, the health needs of lesbians are barely understood and ill appreciated by the health care and research community. Health disparities arise from differential exposure, diagnosis, and treatment, and virtually all women's health issues can be considered through a health disparities framework.

—Lynne C. Messer

See also Chronic Disease Epidemiology; Health Disparities; Life Course Approach; Maternal and Child Health Epidemiology

Further Readings

Boston Women's Health Book Collective. (2005). *Our bodies, ourselves for the new century*. New York: Simon & Schuster.

Carlson, K. J., Eisenstat, S. A., & Ziporyn, T. (2004). *The new Harvard guide to women's health*. Boston: Harvard University Press.

Gay, K. (2001). *Encyclopedia of women's health issues*. Portsmouth, NH: Greenwood.

Goldman, M. B., & Hatch, M. C. (Eds.). (2000). *Women and health.* San Diego, CA: Academic Press.

Loue, S., & Sajotovic, M. (Eds.). (2004). *Encyclopedia of women's health*. New York: Kluwer Academic.

Web Sites

Centers for Disease Control and Prevention, Office of Women's Health: http://www.cdc.gov/women.
Health Resources and Services Administration, Maternal and Child Health Bureau: http://mchb.hrsa.gov.
U.S. Department of Health and Human Services, Office on Women's Health, The National Women's Health Information Center: http://www.4women.gov.
World Health Organization, Women's Health: http://www.who.int/topics/womens_health/en.

World Health Organization

The World Health Organization (WHO) is a specialized agency of the United Nations (UN). It was established on April 7, 1948, to promote international cooperation in improving health conditions. The WHO administrative office is headquartered in Geneva, Switzerland; however, WHO has six regional offices for Africa, the Americas, Southeast Asia, Europe, the Mediterranean, and the Western Pacific, located in major cities in each area.

The WHO mission is improving the health of all people of the world. The WHO definition of health is broad: As defined in the WHO constitution, "health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." A major aspect of this mission is to combat disease, especially infectious diseases, through the development and distribution of vaccines and by coordinating international efforts in monitoring outbreaks of diseases such as malaria and AIDS. However, WHO is also concerned with matters that affect health less directly, such as improving living conditions.

WHO work may be divided into three categories: (1) gathering and dissemination of health information through research services, (2) disease control through providing vaccination and medication, and (3) consultation and education through organizing conferences.

WHO has 192 member states: These include all UN member states, except Liechtenstein, and two non-UN members, Niue and the Cook Islands. These member states appoint delegations to the World Health Assembly, which is WHO's highest decision-making body. The World Health Assembly, which generally meets in May of each year, elects 32 members who are qualified in the field of health and are representatives from WHO's member states to be appointed to the Executive Board for 3-year terms. The main functions of the Board are to give effect to the decisions and policies of the Assembly, to advise it, and generally to facilitate its work.

The Assembly's main duty includes the supervision of the financial budgets, reviewing proposed projects, and appointing the Director General. WHO is financed based on annual contributions made by member governments on the basis of relative ability to pay and on the allocated resources that were assigned by the UN after 1951.

The day-to-day work of the WHO is carried out by its Secretariat, with regional offices throughout the world. These offices are staffed with health care workers who carry out the health projects designed for improving the human beings' health status in that specific region. In addition, WHO is represented by its Goodwill Ambassadors, who work independently and freely. These Ambassadors are usually celebrities, appointed in a nondiplomatic position to use their talent and fame in advocating for the health and well-being of human beings and in supporting WHO's goals and purposes.

WHO operates in 147 country and liaison offices in all its regions. The presence of a country office is motivated by a specific need and must be requested by that country. The country office is headed by a WHO Representative (WR), who is not a national of that country. The office consists of the WR and several health experts, both foreign and local. The main functions of WHO country offices include being the primary adviser of that country's government in matters of health and pharmaceutical policies, as well as coordinating a role for the action of other international and/or nongovernmental organizations when health is concerned and to provide leadership and coordination for emergency and disaster medical relief efforts.

—Najood Ghazi Azar

See also Disaster Epidemiology; Epidemiology in Developing Countries; Health, Definitions of; Vaccination

Web Sites

World Health Organization: http://www.who.org.

Y

Yellow Fever

Yellow fever is a hemorrhagic fever that has a viral etiology. The virus is transmitted to humans by infected mosquitoes (*Aedes aegyptii*). It is called *yellow* fever due to the jaundice that affects some patients causing yellow eyes and yellow skin. The disease itself may be limited to mild symptoms or may cause severe illness or even death. Yellow fever occurs exclusively in Africa and South America. Annually, it is estimated to cause 200,000 cases, and the death toll is estimated to be around 30,000. It is a notifiable disease under the International Health Regulations of the World Health Organization (WHO), and member states are officially obliged to notify yellow fever cases to the WHO.

History

The first descriptions of a disease such as yellow fever can be found in historic texts as early as 400 years ago. It was especially common in American seaports. For instance, Philadelphia experienced an epidemic in 1793, which killed 10% of the city's population, and during which almost half of Philadelphia's residents fled the city. Yellow fever accounted for a significant number of casualties in the American army in the Spanish-American war of 1898: The impact was severe enough to warrant the setting up of the United States Army Yellow Fever Commission, also known as the Reed Commission, in 1900. Carlos Finlay, a Cuban physician, had proposed the mosquito-vector theory in 1881. In collaboration with the Reed Commission and with the help of a few human volunteers, his theory was confirmed. By adopting mosquito control measures, yellow fever was controlled within 6 months in Havana.

At present, yellow fever is considered to be endemic in 33 African countries, 9 South American countries, and several Caribbean islands. Yellow fever has never been reported in Asia.

Transmission

Yellow fever affects mainly humans and monkeys. The virus spreads by *horizontal transmission* (from one person to another by mosquitoes) or by *vertical transmission* (transovarially in infected mosquitoes). There are three transmission cycles depending on whether the mosquitoes are domestic (urban yellow fever), wild (jungle or sylvatic yellow fever), or semidomestic (intermediate yellow fever). In South America, only the first two transmission cycles are found. Sylvatic yellow fever usually causes sporadic cases of yellow fever, mainly in young males who work in the forest. Intermediate yellow fever causes small-scale epidemics in African villages. Epidemics of urban yellow fever may occur when a person from an endemic area migrates to a nonendemic area (especially crowded urban areas) and introduces the virus in an unvaccinated population.

Clinical Manifestations

The virus (a *flavivirus*) enters the body through the bite of a female mosquito. This is followed by an

incubation period of 3 to 6 days during which there are no signs or symptoms. Yellow fever has two phases. The acute phase is characterized by fever, chills, myalgia, headache, nausea, vomiting, anorexia, and general exhaustion. In the majority of the cases, these symptoms subside, and patients improve in 3 to 4 days. A few (about 15%), however, enter a toxic phase within 24 hr of the acute phase. This is characterized by reappearance of fever and deterioration of major organ systems of the body, mainly the liver and the kidneys signaled by jaundice, blood in stools, albuminuria, anuria, and so on. Half of the patients who enter the acute phase die within 10 to 14 days, while the others recover without significant residual organ damage. Yellow fever may be confused with many other diseases, especially in the initial stages, and it can be confirmed by serologic assays and other blood tests. As there is no specific drug for treating yellow fever, treatment is symptomatic.

Prevention

Since yellow fever is a zoonosis and cannot be eradicated, vaccination is the single most effective control measure in populations where it is endemic and for people traveling to endemic areas. Mosquito control can be used as an adjunct to vaccination. In case of low vaccination coverage, surveillance is very critical for early detection and rapid containment of outbreaks.

Yellow fever vaccine is a live viral vaccine that can be administered at 9 months of age, with a booster every 10 years. In countries where yellow fever is endemic, the WHO has recommended incorporation of yellow fever vaccine in routine childhood immunizations. It is contraindicated in cases of egg allergy, immune deficiency, pregnancy, and hypersensitivity to previous dose. Rare adverse reactions have included encephalitis in the very young, hepatic failure, and even death due to organ failure. The vaccine is not given before 6 months of age. In the past, intensive immunization campaigns have been successful in greatly decreasing the number of yellow fever cases.

—Sangeeta Karandikar

See also Epidemiology in Developing Countries; Insect-Borne Disease; Notifiable Disease; Public Health Surveillance; Reed, Walter

Further Readings

Atwater Kent Museum of Philadelphia. (2007). City history lesson 'Yellow Fever Epidemic of 1793.' Retrieved November 18, 2006, from http://www.philadelphia history.org/akm/lessons/yellowFever.

Chaves-Carballo, E. (2005). Carlos Finlay and yellow fever: Triumph over adversity. *Military Medicine*, *170*(10), 881–885.

Health Sciences Library at the University of Virginia. (2004). *Yellow fever and Reed Commission, 1898–1901*. Retrieved November 18, 2006, from http://www.healthsystem.virginia.edu/internet/library/historical/medical_history/yellow_fever.

Tomori, O. (2004). Yellow fever: The recurring plague. *Critical Reviews in Clinical Laboratory Sciences*, *41*(4), 391–427.

World Health Organization. (2001). *Yellow fever fact sheet*. Retrieved November 18, 2006, from http://www.who.int/mediacentre/factsheets/fs100/en.

Youth Risk Behavior Surveillance System

The Youth Risk Behavior Surveillance System (YRBSS) is a project of the Centers for Disease Control and Prevention (CDC). Developed in 1990, the YRBSS consists of local, state, and national school-based surveys of youth to monitor the prevalence of and trends in health risk behaviors. Specific behaviors of interest include substance abuse, risky sexual practices, and unhealthy physical and dietary behaviors. These risk behaviors, often developed during childhood and adolescence, contribute to the leading causes of mortality and morbidity among youth and adults in the United States. The goals of YRBSS include documenting the prevalence and co-occurrence of health risk behaviors; monitoring trends in the prevalence of these behaviors; generating comparable national, state, and local data, as well as data on youth subpopulations; and tracking progress toward the objectives of Healthy People 2010 and other programs.

Surveys are conducted every 2 years and are representative of public and private high school students at the national level and public high school students at the local and state levels. Students are not tracked over time, and these data are publicly available. In addition to these biennial surveys, YRBSS also includes other national CDC surveys, including the

National College Health Risk Behavior and Alternative High School Youth Risk Behavior Surveys.

Methodology

National YRBSS surveys employ three-stage cluster sampling, while local and state surveys use two-stage cluster sampling. In the first stage of sampling in national surveys, primary sampling units (PSUs) are selected. PSUs are usually large-sized counties or groups of adjacent, smaller counties. During the second sampling stage, public and private schools are selected from the PSUs. Black and Hispanic students are oversampled at this stage to gain enough data to conduct analyses of these subgroups separately. Finally, classes are sampled in each grade of each selected high school. All students in selected classes are eligible for the survey. Local and state surveys use two stages of sampling: first, to select schools, and second, to sample classes within the selected schools.

The data are collected in a similar manner in the national, state, and local surveys. Parental permission is obtained in accordance with local standards for all YRBSS surveys. Students' participation in the self-administered surveys is anonymous and voluntary. Between 1991 and 2003, student response rates ranged from 83% to 90%.

Limitations

There are several limitations of the YRBSS. Key limitations are all data are self-reported; the surveillance includes only in-school youth, who are less likely to engage in risky health behaviors; parental consent procedures are inconsistent; state-level data are not available for all 50 states; and the system is not able to evaluate specific interventions or programs.

—Anu Manchikanti

See also Behavioral Risk Factor Surveillance System; Child and Adolescent Health; Healthy People 2010

Further Readings

Centers for Disease Control and Prevention. (2004). Methodology of the youth risk behavioral survey system. *Morbidity and Mortality Weekly Report*, *53*(RR-12), 1–16.

National Center for Chronic Disease Prevention and Health Promotion. (2006). YRBSS: Youth risk behavior surveillance system. Retrieved August 15, 2006, from http://www.cdc.gov/HealthyYouth/yrbs/index.htm.

Z

Zoonotic Disease

Infectious or communicable diseases of humans can be divided into those that are communicable only between humans and those that are communicable to humans by nonhuman vertebrate animals (those with backbones such as mammals, birds, reptiles, amphibians, and fish, referred to in this entry simply as ''animals''). The latter diseases are called zoonoses or zoonotic diseases. Because of the large number of domestic and wild animals that can serve as a source of zoonotic diseases, and the numerous means of transmission including vectors, zoonotic diseases are often difficult to control. Public health veterinarians have a critical role in zoonotic disease surveillance, prevention, and control, but risk reduction increasingly requires application of multidisciplinary teams and a unified concept of medicine across human and animal species lines.

Zoonotic Disease Classification

All classes of disease agents cause zoonotic disease. These include bacteria (e.g., listeriosis), chlamydia (e.g., psittacosis), rickettsia (e.g., Rocky Mountain spotted fever), viruses (e.g., Hendra), parasites (e.g., leishmaniasis), and fungi (e.g., histoplasmosis).

Zoonoses can be subdivided into those transmitted from animals to humans (zooanthroponoses) or from humans to animals (anthropozoonoses, also called reverse zoonoses). *Mycobacterium tuberculosis* has been spread from humans to cattle and elephants, and methacillin-resistant *Staphylococcus aureus* (MRSA) has been transmitted from people to horses and then back to people. Diseases that are rarely transmitted between animals and humans are sometimes included, such as foot-and-mouth disease in cattle.

Zoonoses transmitted through direct contact are orthozoonoses (e.g., rabies). Cyclozoonoses (e.g., echinococcosis) require more than one vertebrate host for development. Metazoonoses are transmitted by an infected invertebrate vector (e.g., scrub typhus from mite bites). Zoonoses transmitted through physical contact with food, soil, or vegetation are saprozoonoses, sapronoses, or geonoses (e.g., fungal infections). Some diseases fit more than one category (e.g., tularemia from fly or tick bites, direct contact, or environmental exposure).

As noted by Enserink in *Science* in 2000, of 1,709 identified human disease agents, 832 (49%) are classified as zoonotic. Of the 156 ''emerging'' diseases, 114 (73%) are zoonotic. Thus, zoonotic diseases are disproportionately represented among those spreading into new areas.

Zoonotic disease agents account for most of the organisms in Categories A, B, and C of the U.S. government's Select Agent List of likely organisms for bioterrorism attacks. The diseases caused by Category A select agents include smallpox, anthrax, plague, tularemia, botulism, and viral hemorrhagic fevers.

Select agents in Categories B and C cause bacterial, chlamydial, and rickettsial diseases, including brucellosis, Q Fever, glanders, melioidosis, foodborne/waterborne disease, psittacosis, and typhus. Select viral agents include smallpox, Nipah, hanta, and the

encephalitides viruses. Select agents that can lead to intoxications include *Staphylococcus* enterotoxin B, ricin, and *Clostridium perfringens* Epsilon toxin. All these select agents are considered zoonotic except for smallpox.

Populations at Increased Risk

Anyone who comes into contact with infected animals, vectors, or contaminated areas can become infected with zoonotic diseases; however, the risk of clinical signs and death is not uniformly distributed. The proportion who remain asymptomatic and the case fatality rate (proportion of ill persons who die) vary with certain risk factors.

Age is often associated with disease severity. Of those infected with *Escherichia coli* O157:H7 from contact with animals or their environment, children are more likely to develop potentially fatal hemolytic uremic syndrome (HUS). Younger, healthier people appear to be more susceptible to serious illness from the highly pathogenic avian influenza (HPAI) strain of Asian H5N1, compared with human influenza strains that differentially cause severe illness and death in older people. Similarly, hantavirus infection was first identified in physically fit young adults, and the very young and very old still seem to be relatively unaffected. Although the factors leading to these age differences are not understood, infection with both the Asian H5N1 HPAI virus and hantaviruses lead to pathologic changes caused by the body's own immune response to the viruses.

For some diseases, immunosuppression from disease or medication is a risk factor. Cryptosporidiosis is a common coinfection with acquired immunodeficiency syndrome (AIDS). Those without a functioning spleen have an increased risk of illness and death from *Capnocytophaga canimorsus* infection after dog bites. Those who take chloroquine for malaria prophylaxis concurrently with rabies preexposure immunizations are less likely to develop a sufficient immunologic response to survive a rabies exposure.

Other populations at risk include those who are cognitively impaired and cannot recognize or report bites from rabid bats. Pregnant women are at risk of fetal congenital malformations with lymphocytic choriomeningitis virus (LCMV) infection. Solid-organ transplant recipients have died from rabies and LCMV infections transmitted from the donors.

Zoonotic Disease Control

Zoonotic diseases are particularly difficult to control because of their animal reservoirs. Zoonoses are unlike diseases that can be eradicated with intensive human vaccination campaigns, such as smallpox and polio. It may be possible to eliminate some zoonotic disease variants from certain regions, as campaigns with oral rabies vaccines have attempted to do by distributing vaccine baits for vaccination of foxes, coyotes, and raccoons.

Global movement of animals has increased problems with zoonotic disease control. Inexpensive puppies of certain breeds are in great demand, and occasionally they are rabid when imported. Raccoon rabies was introduced to the central east coast of the United States by deliberate human movement of raccoons from the southeastern United States for hunting purposes, with subsequent spread to the entire east coast of the United States, as well as the midwestern United States and parts of Canada. Monkeypox virus infection of African rodents imported for the pet trade led to pet prairie dog and human cases and restrictions on the pet prairie dog trade. The spread of Asian H5N1 HPAI (avian flu) from Asia to Europe, the Middle East, and Africa appears to be a result of human movement of domestic birds as well as wild bird migration.

Zoonotic disease risk is increased when humans live in close proximity to domestic animals such as poultry and livestock. This allows efficient use of limited land resources and constant care and protection of the animals. But this practice increases the risk of humans becoming infected with disease agents such as HPAI.

Even in areas with greater separation between human homes and animal barns, zoonotic diseases still pose a risk because of human contact with animals. *Salmonella* infections (sometimes with multidrug resistant strains) have occurred from pets in homes, including reptiles and amphibians (turtles, iguanas, snakes), exotic pets (hedgehogs, sugar gliders), pocket pets (hamsters, mice, rats), pet birds (chicks, ducklings), dogs and cats, and from pet treats. Transmission can be directly from animal handling or by exposure to environmental contamination.

These transmission routes also apply to disease agents spread from livestock to people in public settings. Large *E. coli* O157:H7 outbreaks have been associated with dairy farms, children's day camps

conducted in farm settings, social events in buildings previously used for animal exhibitions, fair petting zoos, and contaminated fair water systems. Critical control methods in homes and public settings include animal management to reduce disease burden, management of animal and human contacts, and education to reduce exposure particularly by handwashing.

Limiting contact between humans and wild animals is also critical to reducing risk. Most human rabies deaths in the United States are due to bites from bats, frequently in home settings. Although the human immunodeficiency virus (HIV) that causes AIDS is not zoonotic, it apparently evolved from similar monkey viruses through the practice of hunting and consuming bush meat (monkeys). Unprotected cleaning up of rodent feces is associated with hantavirus infection, and plague infection is associated with use of woodpiles and with outdoor recreational activities that bring people into contact with wild rodents and their fleas.

Contact between wild and domestic animals also increases the threat of zoonotic diseases for people. One example is Nipah virus, first identified in Indonesian pigs and pig farmers. Preliminary information about a possible association with fruit bats led to removal of fruit trees overhanging pig pens, which eliminated the pig and human cases in that area. This control method is reminiscent of John Snow's interruption of an 1854 London cholera outbreak after he identified a statistical association between use of a well pump and cases of cholera, even though cholera transmission was not fully understood.

Because zoonotic disease agents can be found in humans, animals, the environment, and vectors, management requires the collaboration of many types of health and disease control specialists. Disease control may include vector control programs for ticks (Lyme disease), fleas (plague), or mosquitoes (West Nile virus), and environmental cleanup or protection may be required to address disease agents that remain viable from days to years on surfaces (*Salmonella*), in soils (*Anthrax*), or in the water (*Leptospira*). In most state health agencies, public health veterinarians are available to assist in this critical disease control coordination.

—Millicent Eidson

See also Avian Flu; Bioterrorism; Influenza; Plague; Snow, John; Vector-Borne Disease

Further Readings

Beran, G. W., & Steele, J. H. (Eds.). (1994). *Handbook of zoonoses* (2nd ed.). Boca Raton, FL: CRC Press.

Enserink, M. (2000). Emerging diseases: Malaysian researchers trace Nipah virus outbreak to bats. *Science*, *289*, 518–519.

Hugh-Jones, M. E., Hubbert, W. T., & Hagstad, H. V. (2000). *Zoonoses: Recognition, control, and prevention*. Ames: Iowa State University Press.

Krauss, H., Weber, A., Appel, M., Enders, B., Isenberg, H. D., Scheifer, H. G., et al. (Eds.). (2003). *Zoonoses: Infectious diseases transmissible from animals to humans.* Washington, DC: ASM Press.

Palmer, S. R., Soulsby, L., & Simpson, D. I. H. (Eds.). (1998). *Zoonoses: Biology, clinical practice, and public health control.* Oxford, UK: Oxford University Press.

Z Score

Z scores are also called standard units or standard scores. A z score standardizes values of a random variable from a normal (or presumed normal) distribution for comparison with known probabilities in a standard normal probability distribution table or z table. A z score is unitless and is simply a measure of the number of standard deviations a value is from the mean. The z score is especially useful to indicate where a particular data point is relative to the rest of the data. A positive z score indicates that the point is above the mean, while a negative z score indicates that the point is below the mean. The z score removes varying units (e.g., pounds or kilograms) and allows for easy determination of whether a particular result is unusual.

Most significance testing, hypothesis testing, and confidence intervals are based on an assumption that the data are drawn from an underlying normal distribution. However, the probabilities are dependent on the mean, μ, and standard deviation, σ, of the distribution. Since it is physically impossible to calculate the probabilities associated with infinitely many pairs of μ and σ, the z score allows a researcher to compare the sample data with the standard normal distribution. The standard normal distribution is a normal distribution with $\mu = 0$ and $\sigma = \sigma^2 = 1$, also denoted $N(0,1)$.

The z score finds the point z on the standard normal curve that corresponds to any point x on a nonstandard normal curve. To convert an x value from

the original scale to a z score, center the distribution by subtracting the mean and then rescale by dividing by the standard deviation. The formula for this conversion is

$$z = \frac{\text{Data point} - \text{Mean}}{\text{Standard deviation}}.$$

In a population, the formula becomes $z = (x - \mu)/\sigma$. In a sample, the formula becomes $z = (x - \bar{x})/s$. This standardization ensures that the resulting distribution has a mean of 0 and standard deviation of 1. See Figure 1 for an illustration of how the x scales and z scores compare for a normal distribution with mean $\mu = 5$ and standard deviation $\sigma = 2$.

If the assumption of normality for the data is not grossly violated, then the z score will allow the researcher to compare the data with known probabilities and draw conclusions. If x did not have at least an approximately normal sampling distribution, then further use of the z score may result in erroneous conclusions. The central limit theorem does not apply since the researcher is interested in individual x values, not the mean of the x values.

Using the relationship between the z and x scales, a researcher can use a standard normal table or z table to find the area under any part of any nonstandard normal curve. To calculate the area or probability of an x value occurring between two numbers a and b in the x scale, use

$$P(a < X < b) = p\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = p\left(\frac{a-\mu}{\sigma} < z < \frac{b-\mu}{\sigma}\right).$$

In Figure 1 ($\mu = 5$, $\sigma = 2$), to find the probability of an x value between 3 and 10, use

$$P(3 < X < 10) = p\left(\frac{3-5}{2} < \frac{X-\mu}{\sigma} < \frac{10-5}{2}\right) = p(-1 < z < 2.5) = .3413 + .4938 = .8351.$$

A z score is sometimes confused with the z statistic, also denoted "z," which is used in the z test. The z statistic differs in that it measures the number of standard deviations a sample statistic (the mean, $\bar{x}$) is from some hypothesized value of a population parameter (a numerical value, μ_0). The differences between that formula and the z score formula are subtle,

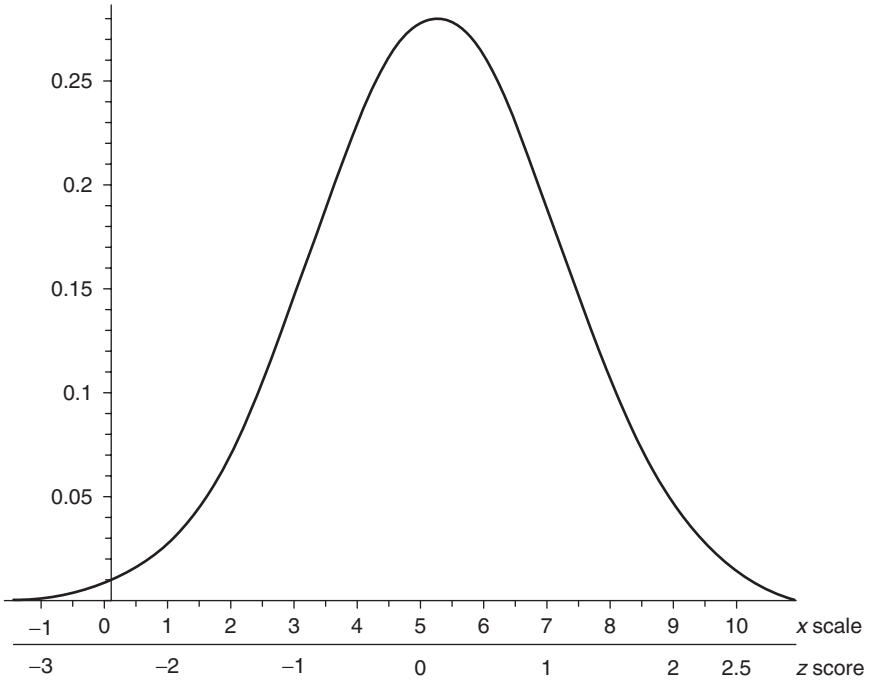


Figure 1 Comparison of x Scales and z Scores

understanding of the area. Following each entry is a Further Readings section that presents the interested reader with a carefully selected set of references should they wish to inquire more deeply into the topic.

How the *Encyclopedia of Epidemiology* Was Created

The *Encyclopedia* was developed in seven basic steps.

Step 1. Leading epidemiologists and public health scientists were invited to serve on the Advisory Board. The Advisory Board includes faculty members from major epidemiology programs, in addition to a biostatistician and a health economist from two of the leading medical schools in the United States. Together they present broad and varied experience in the field, including considerable overseas research and training.

Step 2. A master list of topics for the *Encyclopedia* was created. First an initial list of topic headwords was created by the Editor in Chief, using as a guide the content of major epidemiologic textbooks, epidemiology and public health journals, and the curricula of accredited epidemiology and public health programs in the United States. This list was revised by the Advisory Board, with a particular effort made to include topics of current relevance, such as bioterrorism, SARS (severe acute respiratory syndrome), and violence as a public health issue.

Step 3. Contributors were identified and invited, based on personal knowledge of their expertise, references from colleagues, and publication record. The authors represent a mix of academics and practitioners from around the world and include people from all career stages, from new researchers with particular subject expertise in an emerging field to emeritus faculty drawing on many years of experience.

Step 4. Contributors were given basic instructions and guidelines regarding their entries. In particular, they were encouraged to be thorough in coverage, nontechnical in approach, and balanced in the presentation of different views on any controversial issue.

Step 5. The Editor in Chief and Advisory Board members, who are all subject experts in epidemiology and related fields, reviewed all articles and requested revisions as necessary.

Step 6. A Sage Developmental Editor, who was not an epidemiologist, reviewed entries once they had been passed by the subject experts, with an eye to consistency and comprehensibility for the general reader. Further revisions were requested at this stage if necessary.

Step 7. The volumes were finalized and the extra material was compiled, including the Reader's Guide and final matrix of cross-references and blind references.

Acknowledgments

A number of people played major roles in producing this *Encyclopedia*. First I would like to thank my Advisory Board, who went far beyond the call of duty in suggesting current topics, locating authors, and participating in the manuscript review process. I would also like to thank my editorial team at Sage, in particular, my developmental editor, Diana Axelsen, and my production editor, Melanie Birdsall. And of course I would like to thank all the scholars and professionals who authored the entries.

I would also like to thank my literary agent, Neil Salkind at Studio B, for his guidance and support.

On a personal level, I would like to thank my husband, Dan Peck, for his unwavering love and support, my professors at City University of New York and Saint Louis University for setting me down the path that led to this *Encyclopedia*, and my colleagues at Washington University Medical School and BJC HealthCare in Saint Louis for sharing their collective wisdom and experience.

—Sarah Boslaugh

including the change from the use of the standard deviation of x, σ, to the use of the standard error, $\sigma_{\bar{x}}$, which is also known as the standard deviation of the sampling distribution. The z statistic formula is

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sigma / -\sqrt{n}}.$$

—Stacie Ezelle Taylor

See also Central Limit Theorem; Normal Distribution; Random Variable; Sampling Distribution

Further Readings

Munro, B. H. (2005). *Statistical methods for health care research* (5th ed.). Philadelphia: Lippincott Williams & Wilkins.

Petruccelli, J. D., Nandram, B., & Chen, M. (1999). *Applied statistics for engineers and scientists.* Upper Saddle River, NJ: Prentice Hall.

Index