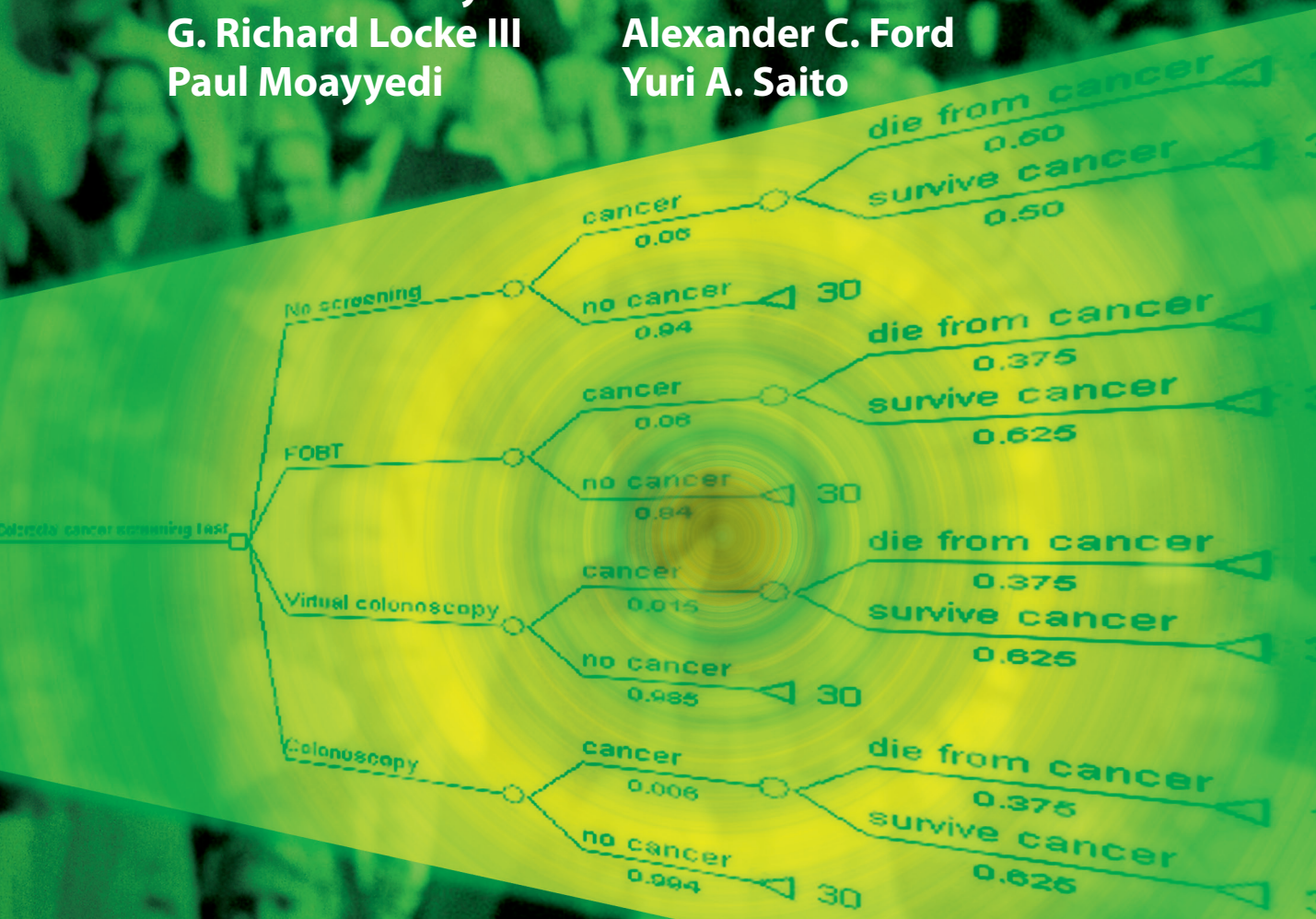
GI Epidemiology

Diseases and Clinical Methodology

SECOND EDITION

Edited by
Nicholas J. Talley
G. Richard Locke III
Paul Moayyedi
Joe West
Alexander C. Ford
Yuri A. Saito



with website

WILEY
Blackwell

GI Epidemiology

GI Epidemiology
Diseases and Clinical Methodology

Second Edition

Edited by

Nicholas J. Talley, MD, PhD, M Med Sci (Clin Epi), FRACP, FAFPHM, FRCP, FACP, FACG, AGAF
Pro Vice-Chancellor and Dean (Health and Medicine), and Professor
Senior Staff Specialist (Gastroenterology), John Hunter Hospital
University of Newcastle
Callaghan, NSW, Australia

G. Richard Locke III, MD
Professor of Medicine
GI Epidemiology/Outcomes Unit
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

Paul Moayyedi, BSc, MB ChB, PhD, MPH, FRCP, FRCPC, AGAF, FACG
Acting Director of the Farncombe Family Digestive Health Research Institute
Director of Division of Gastroenterology
McMaster University
Hamilton, ON, Canada

Joe West, BMedSci, BM BS, MRCP, MSc, PhD, PGDip
Associate Professor and Reader in Epidemiology; Honorary Consultant Gastroenterologist
Division of Epidemiology and Public Health
University of Nottingham
Nottingham, UK

Alexander C. Ford, MBChB, MD, FRCP
Associate Professor and Honorary Consultant Gastroenterologist
Leeds Teaching Hospitals Trust
Leeds, West Yorkshire, UK

Yuri A. Saito, MD, MPH
Assistant Professor of Medicine
Director, GI Epidemiology/Outcomes Unit
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

WILEY Blackwell

This edition first published 2014 © 2014 by John Wiley & Sons, Ltd; 2007 by Blackwell Publishing

Registered office: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell

Library of Congress Cataloging-in-Publication Data

GI epidemiology : diseases and clinical methodology / edited by Nicholas J. Talley, G. Richard Locke III, Paul Moayyedi, Joe West, Alexander C. Ford, Yuri A. Saito. – Second edition.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-67257-0 (hardback : alk. paper) – ISBN 978-1-118-72707-2 – ISBN 978-1-118-72708-9 (ePdf) – ISBN 978-1-118-72709-6 (ePub) – ISBN 978-1-118-72710-2 (mobi)
I. Talley, Nicholas Joseph, editor of compilation. II. Locke, G. Richard, III, editor of compilation. III. Moayyedi, Paul, editor of compilation. IV. West, Joe, editor of compilation. V. Ford, Alexander C., editor of compilation.
[DNLM: 1. Gastrointestinal Diseases–epidemiology. 2. Epidemiologic Methods. WI 140]
RC801
616.3′3–dc23

2013018951

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover design by Nathan Harris

Set in 9/11.5pt Sabon by Aptara Inc., New Delhi, India

1 2014

Contents

Contributors

Guruprasad P. Aithal, BSc, MBBS, MD, FRCP, PhD
Co-Director, NIHR Nottingham Digestive Diseases Biomedical Research Unit
Nottingham University Hospitals NHS Trust and University of Nottingham
Nottingham, UK

David Armstrong, MA, MB BChir, FRCP, FRCPC, AGAF, FACG
Professor of Medicine
Farncombe Family Digestive Health Research Institute & Division of Gastroenterology
McMaster University
Hamilton, ON, Canada

Sumeet K. Asrani, MD
Hepatology
Baylor University Medical Center
Dallas, TX, USA

Alan Barkun, MD, MSc
Division of Gastroenterology
Montreal General Hospital Site
The McGill University Health Centre
Montreal, QC, Canada

Adil E. Bharucha, MD
Professor of Medicine
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

Timothy R. Card, PhD, FRCP
Associate Professor and Honorary Consultant Gastroenterologist
Division of Epidemiology and Public Health
Nottingham City Hospital
University of Nottingham
Nottingham, UK

Margaret D. Carroll, MSPH
Division of Health and Nutrition Examination Surveys
National Center for Health Statistics
Centers for Disease Control and Prevention
Hyattsville, MD, USA

Suresh T. Chari, MD
Professor of Medicine
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

Rok Seon Choung, MD, PhD
Department of Internal Medicine
Institute of Digestive Diseases and Nutrition
Korea University
Seoul, South Korea

Abhijit Chowdhury, MBBS, MD, DM
Division of Hepatology
School of Digestive and Liver Diseases
IPGME & R, Kolkata, India

Seth D. Crockett, MD, MPH
Assistant Professor, Division of Gastroenterology
University of North Carolina School of Medicine
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

Colin J. Crooks
Division of Epidemiology and Public Health
Nottingham City Hospital
University of Nottingham
Nottingham, UK

Ilona Csizmadi, MSc, PhD
Research Scientist, Department of Population Health Research
Alberta Health Services-Cancer Care
Departments of Community Health Sciences and Oncology
Faculty of Medicine
University of Calgary
Calgary, AB, Canada

Kshaunish Das, MBBS, MD, DM
Division of Gastroenterology
School of Digestive and Liver Diseases
IPGME & R, Kolkata, India

Jessica A. Davila, PhD
Associate Professor of Medicine
Program Chief, Health Services Research Methodology and Statistics Core
Baylor College of Medicine
Michael E. DeBakey VA Medical Center
Houston, TX, USA

Anne Duggan, PhD, FRACP, B.Med, MHP
Conjoint Professor
Department of Gastroenterology
John Hunter Hospital
Newcastle, NSW, Australia

Hashem B. El-Serag, MD, MPH
Chief, Gastroenterology and Hepatology
Chief, Clinical Epidemiology and Outcomes
Baylor College of Medicine
Houston, TX, USA

Guy D. Eslick, DrPH, PhD, FACE, FFPH
Associate Professor of Surgery and Cancer Epidemiology
The Whiteley-Martin Research Centre
Discipline of Surgery
The University of Sydney
Sydney, Australia

James E. Everhart, MD, MPH
Chief, Epidemiology and Clinical Trials Branch
Division of Digestive Diseases and Nutrition
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, MD, USA

Tala H.I. Fakhouri, PhD, MPH
National Center for Health Statistics
Centers for Disease Control and Prevention
Hyattsville, MD;
Epidemic Intelligence Service
Centers for Disease Control and Prevention
Atlanta, GA, USA

Katherine M. Flegal, PhD
Office of the Director
National Center for Health Statistics
Centers for Disease Control and Prevention
Hyattsville, MD, USA

Kate M. Fleming, MA, MSc, PhD
Lecturer in Gastrointestinal Epidemiology
Division of Epidemiology and Public Health
Nottingham City Hospital
University of Nottingham
Nottingham, UK

Alexander C. Ford, MBChB, MD, FRCP
Associate Professor and Honorary Consultant Gastroenterologist
Leeds Teaching Hospitals Trust
Leeds, West Yorkshire, UK

Rodney Givney, BScMed (Hons), MBBS, FRCPA, PhD, MPH
Division of Microbiology
Hunter Area Pathology
Pathology North & Newcastle University
New South Wales, Australia

Matthew J. Grainge, MSc, PhD
Associate Professor
Division of Epidemiology and Public Health
Nottingham City Hospital
University of Nottingham
Nottingham, UK

Magnus Halland, BMed, BMedSci (Hons), MPH
Conjoint Lecturer
University of Newcastle
Callaghan, NSW, Australia

David J. Humes, BSc, MBBS, MRCS, PhD
Lecturer in Surgery
Nottingham Digestive Disease Centre Biomedical Research Unit
Division of Surgery
University of Nottingham
Nottingham, UK

Richard H. Hunt, FRCP, FRCPEd, FRCPC, MACG, AGAF
Professor, Farncombe Family Digestive Disease Research Institute and Division of Gastroenterology
McMaster University Health Science Centre
Hamilton, ON, Canada

John M. Inadomi, MD
Cyrus E. Rubin Professor of Medicine
Head, Division of Gastroenterology
University of Washington
Seattle, WA, USA

Steven J. Jacobsen, MD, PhD
Director of Research
Research and Evaluation
Southern California Permanente Medical Group
Pasadena, CA, USA

Linda E. Kelemen, MSc, ScD
Research Scientist, Department of Population Health Research
Alberta Health Services-Cancer Care
Departments of Medical Genetics and Oncology
University of Calgary
Calgary, AB, Canada

Hannah P. Kim
Division of Gastroenterology and Hepatology
University of North Carolina School of Medicine
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

W. Ray Kim, MD
Associate Professor of Medicine
GI Epidemiology/Outcomes Unit
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

Brian K. Kit, MD, MPH
Division of Health and Nutrition Examination Surveys
National Center for Health Statistics
Centers for Disease Control and Prevention
Hyattsville, MD, USA

Brian E. Lacy, PhD, MD
Professor of Medicine
Dartmouth Medical School
Director, GI Motility Laboratory
Section of Gastroenterology and Hepatology
Dartmouth-Hitchcock Medical Center
Lebanon, NH, USA

Grigorios I. Leontiadis, MD, PhD
Associate Professor
Department of Medicine
Division of Gastroenterology
McMaster University
Hamilton, ON, Canada

John M. Levenick, MD
Clinical Instructor in Medicine
Section of Gastroenterology and Hepatology
Dartmouth-Hitchcock Medical Center
Lebanon, NH, USA

Joseph Lipscomb, PhD
Professor of Public Health and Georgia Cancer Coalition Distinguished Cancer Scholar
Department of Health Policy & Management
Rollins School of Public Health
Emory University
Atlanta, GA, USA

G. Richard Locke III, MD
Professor of Medicine
GI Epidemiology/Outcomes Unit
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

Edward V. Loftus, Jr, MD
Professor of Medicine
Director, Inflammatory Bowel Disease Interest Group
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

Jonas F. Ludvigsson, MD, PhD
Professor in Clinical Epidemiology
Senior Pediatrician
Karolinska Institutet and Örebro University Hospital
Stockholm, Örebro, Sweden

Sanjiv Mahadeva, MRCP, MD
Lecturer and Consultant Gastroenterologist
Division of Gastroenterology, Department of Medicine
Faculty of Medicine
University of Malaya
Kuala Lumpur, Malaysia

Salaheddin Mahmud, MD, PhD, FRCPC
Assistant Professor of Medicine
Community Health Sciences, University of Manitoba;
Epidemiology, CancerCare Manitoba
Winnipeg, MB, Canada

L. Joseph Melton III, MD, MPH
Professor of Epidemiology
Department of Health Sciences Research
Mayo Clinic College of Medicine
Rochester, MN, USA

Paul Moayyedi, BSc, MB ChB, PhD, MPH, FRCP, FRCPC, AGAF, FACG
Acting Director of the Farncombe Family Digestive Health Research Institute
Director of Division of Gastroenterology
McMaster University
Hamilton, ON, Canada

Joselito M. Montalban, MD, MCHM
Departments of Internal Medicine and Community Health Sciences
University of Manitoba
Winnipeg, MB, Canada

Joseph A. Murray, MD
Professor of Medicine
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

Olof Nyrén, MD, PhD
Professor of Clinical Epidemiology
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Stockholm, Sweden

Cynthia L. Ogden, PhD, MRP
Division of Health and Nutrition Examination Surveys
National Center for Health Statistics
Centers for Disease Control and Prevention
Hyattsville, MD, USA

Judith M. Podskalny, PhD
Director, Research Fellowship and Career Development
and Digestive Disease Centers Programs
Division of Digestive Diseases and Nutrition
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, MD, USA

Dawn Provenzale, MD, MS
Associate Professor of Medicine
Director, Durham Epidemiologic Research and Information Center
Director, GI Outcomes Research
Duke University Medical Center
Durham, NC, USA

Linda Rabeneck, MD, MPH, FRCPC
Vice President, Prevention and Cancer Control
Cancer Care Ontario
Professor of Medicine, University of Toronto
Toronto, ON, Canada

Enrique Rey, MD, PhD, AGAF
Professor of Medicine
Functional GI Disorders Unit
Division of Digestive Diseases
Hospital Clinico San Carlos, Universidad Complutense
Madrid, Spain

Alberto Rubio-Tapia, MD
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

Yuri A. Saito, MD, MPH
Assistant Professor of Medicine
Director, GI Epidemiology/Outcomes Unit
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

William Sanchez, MD
Assistant Professor of Medicine
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine
Rochester, MN, USA

William J. Sandborn, MD
Professor of Medicine and Adjunct Professor of Surgery
Chief, Division of Gastroenterology
Director, UCSD IBD Center
University of California San Diego and UC San Diego Health System
La Jolla, CA, USA

Philip Schoenfeld, MD, MSEd, MSc (Epi)
Professor of Medicine
University of Michigan School of Medicine
Anne Arbor, MI, USA

Eldon A. Shaffer, MD, FRCPC
Professor of Medicine
Division of Gastroenterology
University of Calgary
Calgary, AB, Canada

Nicholas J. Shaheen, MD, MPH
Professor of Medicine and Epidemiology
Director, Center for Esophageal Diseases and Swallowing
University of North Carolina School of Medicine
Chapel Hill, North Carolina, USA

Harminder Singh, MD, MPH
Assistant Professor of Medicine
Departments of Internal Medicine and Community Health Sciences, University of Manitoba;
Division of Gastroenterology
University of Manitoba IBD Clinical and Research Centre;
Department of Medical Oncology and Haematology, CancerCare Manitoba;
Winnipeg, MB, Canada

Robin Spiller
Nottingham Digestive Disease Biomedical Research Unit
University of Nottingham, Queen's Medical Centre
Nottingham, UK

Crenguta Stepan, MD
University of Washington
Valley Medical Center
Seattle, WA, USA

Aravind Sugumar, MD
Assistant Professor of Medicine
Department of Gastroenterology and Hepatology
University of Kansas Medical Center
Kansas City, KS, USA

Joseph Sung, MD, PhD
Chairman and Professor of Medicine
Department of Medicine & Therapeutics
Prince of Wales Hospital, Shatin
The Chinese University of Hong Kong
NT, Hong Kong

Christina M. Surawicz, MD
Professor of Medicine
Section Chief, Gastroenterology, Harborview Medical Center
Assistant Dean for Faculty Development
University of Washington School of Medicine
Seattle, WA, USA

Nicholas J. Talley, MD, PhD, M Med Sci (Clin Epi), FRACP, FAFPHM
Pro Vice-Chancellor and Dean (Health and Medicine), and Professor
University of Newcastle
Callaghan, NSW, Australia

Laila J. Tata, BSc, MSc, PhD
Associate Professor of Epidemiology
Division of Epidemiology and Public Health
University of Nottingham
Nottingham, UK

Frances Tse, MD
Division of Gastroenterology
McMaster University
McMaster University Medical Centre
Hamilton, ON, Canada

Santhi Swaroop Vege
Professor of Medicine
Division of Gastroenterology and Hepatology
Mayo Clinic
Rochester, MN, USA

Joe West, BMedSci, BM BS, MRCP, MSc, PhD, PGDip
Associate Professor and Reader in Epidemiology; Honorary Consultant Gastroenterologist
Division of Epidemiology and Public Health
University of Nottingham
Nottingham, UK

Ingela Wiklund, PhD
Senior Research Leader
Evidera
London, UK

Dhiraj Yadav, MD, MPH
Division of Gastroenterology and Hepatology
University of Pittsburgh Medical Center
Pittsburgh, PA, USA

Hematram Yadav, MBBS, MPH, MBA, MRSH, FAMM
Department of Community Medicine
Faculty of Medicine
International Medical University
Kuala Lumpur, Malaysia

Foreword

This volume presents an authoritative overview of current understanding regarding the epidemiology of gastrointestinal diseases and will serve well those working in research or clinical medicine who are seeking to answer questions regarding the causes of such diseases. All major gastrointestinal disease entities are covered in 23 topic-orientated chapters, each with a set of key points and some testing multiple choice questions, and the reader can jump straight into their disease of interest knowing that the state of the art in epidemiology will be presented in a clear and concise manner. But this book offers much more. It is really two books in one: in addition to the topic-orientated chapters, an extensive series of introductory chapters outlines the major study designs in epidemiology and summarizes the main areas of methodology underlying each design. This provides an important grounding in critical appraisal to guide those wishing to delve deeper into the literature. Many of the methodological complexities are reviewed further in the web supplement to the book. These elements combined with the "how to do clinical research" sections provide a primer in gastrointestinal epidemiology on a par with many standard general epidemiology textbooks.

Professional epidemiologists frequently complain that clinicians in a particular field, who may spend a decade or more honing their diagnostic and therapeutic skills, often think they can acquire an understanding of epidemiology as an incidental by-product without ever seriously considering the necessary methods involved. This book can act as a corrective to such tendencies and it is to be hoped that all those seeking an understanding of the epidemiology of a specific gastrointestinal disease will read the first 12 chapters of the volume (along with the supplement) so they can evaluate the strengths and limitations of the methods employed and hence the certainty or otherwise of the conclusions. Unfortunately no textbook can force a reader to prepare him/herself in this way but this volume leaves no excuse and stands in contrast to many others seeking to cover the same ground.

As medical knowledge expands at an ever increasing pace, the broad understanding of disease distribution and dynamics provided by epidemiology remains of fundamental importance if prevention is going to be placed on the agenda. Many gastrointestinal diseases are fully preventable while others remain enigmatic in their aetiology. This volume covers the full spectrum and ultimately addresses the key public health question for each disease – can it be prevented on the basis of present knowledge and, if so, how? Readers are, however, equipped not only with the answer but the weight and texture of evidence leading to the answer. The editors are to be congratulated on assembling a group of expert gastroenterologists/epidemiologists who can pull this evidence together.

David Forman
Head, Section of Cancer Information
International Agency for Research on Cancer
Lyon, France
October 2013

Preface

A rich tradition of epidemiologic research exists in gastroenterology, and the 2nd edition of *GI Epidemiology* aims to provide a comprehensive expert roadmap. Why is it critically important to study and understand the epidemiology of gastrointestinal and liver diseases? Health professionals strive to cure disease, and epidemiology can provide vital clues about disease pathogenesis and etiology. Case-control and cohort studies as well as clinical trials and meta-analyses inform gastroenterology practice but to interrogate the information requires skills in epidemiology. To rationally apply testing, physicians need to understand the prevalence of disease in their practices. Undertaking rigorous clinical research relies on appropriate study design and this is the core of epidemiology. For priorities in the health system, policy makers rely on knowledge of the burden of illness, while government and nongovernment funders of research use such information to help determine resource distribution.

Knowledge continues to explode; this new edition of *GI Epidemiology* has been completely revised and updated by experts from around the world. We have proudly built on the success of the 1st edition, which has become the standard textbook in the field, with a more global focus, expanded methodological guidance, increased illustrations, summaries of key points, and coverage of all major diseases and syndromes. In addition to the 35 chapters in print, a further 10 chapters online cover additional background material including further insights into specific methodological issues and how to secure research funding. Multiple choice questions have been included to aid learning.

We hope you will enjoy reading *GI Epidemiology*. The best and brightest minds in gastroenterology and epidemiology have contributed to this volume, and we are very grateful for their diligent efforts. Despite an increasing interest in and understanding of the epidemiology of gastrointestinal diseases, many vital gaps remain. We look forward to many of the readers of this book being inspired to fill these gaps. The Editors remain passionate about the discipline of GI Epidemiology and we welcome you to the field.

Nicholas J. Talley, MD, PhD,
on behalf of the Editors

Companion website

This book is accompanied by a website:

ADDITIONAL CHAPTERS: AVAILABLE ONLINE AT:
www.wiley.com/go/talley/giepidemiology.com

Part 1: Gastrointestinal Diseases and Disorders: The Public Health Perspective

Part 2: Methodological Issues in GI Epidemiology

PART ONE

Gastrointestinal Diseases and Disorders: The Public Health Perspective

1 The burden of gastrointestinal and liver disease around the world

Hannah P. Kim[1], *Seth D. Crockett*[2], *& Nicholas J. Shaheen*[3]

[1]Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2]Division of Gastroenterology, University of North Carolina School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]Center for Esophageal Diseases and Swallowing, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA

Key points

- Gastrointestinal and liver diseases are among the most common diseases worldwide, with diarrheal disease, malignancies, and liver disease having a substantial toll on worldwide mortality.
- Many of these diseases are preventable and possibly curable.
- There is wide variability in the incidence, management, and mortality associated with these disease states throughout the world.
- Understanding trends in GI illness and the factors responsible for variability in incidence and outcomes will allow clinicians, public health professionals, policy makers, and healthcare organizations to intervene in a more logical way and allocate resources to meet the needs of afflicted patients and decrease the burden of gastrointestinal and liver diseases.

Introduction

Gastrointestinal and liver diseases represent a significant global health problem, and cause approximately 8 million deaths per year worldwide [1]. In developed countries, GI malignancies are among the leading causes of death. In developing countries, diarrheal disease and viral liver infections are highly prevalent and are responsible for significant mortality. These and other diseases are tracked by international and regional health organizations. These tracking measures allow for some assessment of the global burden of GI disease, and may allow identification of important temporal trends.

Below we highlight sources of burden of GI illness internationally. Using international databases, we will highlight some important trends in diarrheal disease and childhood mortality, explore the burden of gastrointestinal malignancies, and discuss the toll of several selected liver diseases. Because valid international estimates are not available for some gastrointestinal conditions, we report regional data with respect to the toll of other selected GI diseases.

Much of the data demonstrated below has been collected as part of various projects conducted by the World Health Organization (WHO). Geographical regions that are discussed throughout this chapter are based on the six officially delineated WHO regions: Africa, the Americas, Eastern Mediterranean, Europe, Southeast Asia, and Western Pacific. A map delineating each region can be found at: http://www.who.int/about/regions/en/index.html.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Diarrheal disease

Global burden

An estimated 2.5 billion cases of diarrhea occur annually in children under five years of age [2], with an estimated frequency of 2–3 episodes per child per year in developing countries [3]. Diarrheal disease is the second leading cause of mortality in this age group worldwide, after pneumonia. Responsible for over 15 % of deaths of children less than five years of age, diarrheal disease accounts for more than 1.3 million deaths each year. It is also responsible for more deaths than HIV/AIDS, malaria, and measles combined [1].

Figure 1.1 displays the number of under-5 deaths secondary to diarrheal disease by WHO region. Diarrheal death is much more common in the developing world, with over 56 % of deaths occurring in Africa. Africa and Southeast Asia combined account for nearly 80 % of all under-5 diarrhea-related deaths. Furthermore, 75 % of childhood deaths attributable to diarrheal disease can be found in just 14 developing countries, led by India, Nigeria, and the Democratic Republic of the Congo [4]. This is largely due to contamination of drinking water and compromised sanitation in these countries. Children in these countries develop nutritional deficiencies, and are more susceptible to repeated episodes of diarrhea and severe dehydration, also contributing to the high incidence of mortality due to diarrhea in developing nations [2].

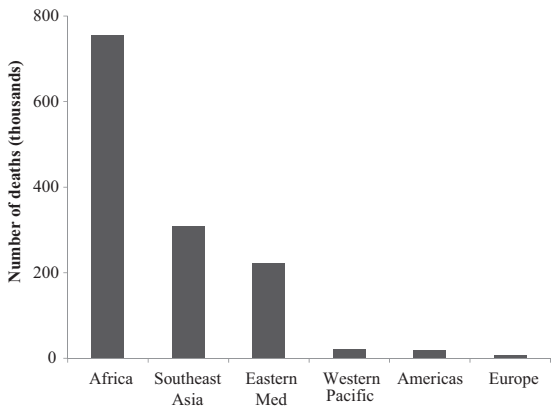


Figure 1.1 Deaths secondary to diarrheal disease among children aged <5 years by WHO region, 2008. Source: WHO Health statistics and health information systems – child mortality by cause.

Efforts to reduce the number of childhood deaths secondary to diarrheal disease in the 1970s and 1980s have favorably impacted the burden of diarrheal disease. These efforts included increasing oral rehydration therapy and the implementation of programs to educate caregivers on proper treatment. While the overall incidence rates of diarrheal disease have remained stable throughout the past three decades, there has been a decrease in diarrhea-associated deaths [3]. Estimates have shown a steady decline with 4.6 million deaths per year in the 1960s and 1970s, 3.3 million deaths per year in the 1980s, 2.5 million deaths per year in the 1990s, and 1.5 million deaths in 2004 [2, 5–7]. Despite this improvement, diarrhea continues to be an unacceptably common cause of childhood death, especially in developing countries.

Gastrointestinal malignancies

Global burden

Cancer is the leading cause of death in developed nations and is the second leading cause of death in developing nations [8]. GLOBOCAN is a WHO project which estimates the international burden of cancer using population-based cancer registries [9]. Gastrointestinal cancers were responsible for nearly one-third of new cancer cases in 2008. Table 1.1 displays incidence of, and mortality from, gastrointestinal cancers worldwide, as well as their rank among all major cancer sites. Colorectal cancer continues to have the highest incidence rate among gastrointestinal malignancies and is the third most commonly occurring cancer worldwide, with over 1.2 million new cases estimated in 2008. Hepatocellular, esophageal, and pancreatic cancers are of particular importance because of their high mortality; in fact, mortality-to-incidence ratios approach one internationally. Colorectal cancer is associated with a much better prognosis, with a mortality-to-incidence ratio of approximately 0.5. Assessment of the three most commonly occurring gastrointestinal malignancies worldwide demonstrates marked variation in incidence and mortality. Colorectal and gastric cancers will be discussed in the following two sections and liver cancer will be discussed in a later section.

Table 1.1 Incidence and mortality of gastrointestinal cancers worldwide, 2008

				Incidence			Mortality		
Rank among GI sites	Rank among all sites	Cancer site	ICD-10 code	Numbers	Crude rate†	ASR§	Numbers	Crude rate†	ASR§
1	3	Colorectum*	C18-21	1,235,108	18.3	17.3	609,051	9.0	8.2
2	4	Stomach	C16	988,602	14.6	14.1	737,419	10.9	10.3
3	6	Liver	C22	749,744	11.1	10.8	695,726	10.3	10.0
4	8	Esophagus	C15	481,645	7.1	7.0	406,533	6.0	5.8
5	13	Pancreas	C25	278,684	4.1	3.9	266,669	4.0	3.7
6	15	Lip, oral cavity	C00-08	263,020	3.9	3.8	127,654	1.9	1.9
7	21	Gallbladder	C23-24	145,203	2.2	2.0	109,587	1.6	1.5

Source: GLOBOCAN 2008.
*Includes anal cancer.
†Crude rates are per 100,000.
§ASR, age-standardized rates per 100,000.

Colorectal cancer

Colorectal cancer is the third highest incident cancer, and fourth most common cause of death from cancer worldwide, with over 609,000 deaths estimated in 2008. Approximately 60 % of colorectal cancer cases are found in developed regions; however, only approximately 53 % of deaths attributable to colorectal cancer are found in these same regions. Of note, the incidence rate of colorectal cancer in Africa is a small fraction of that in Europe, but is associated with cancer-related mortality in nearly all cases.

In the last three decades, the United States has witnessed a decrease in the incidence rate of colorectal cancer and an even greater decrease in the mortality rate. The extent to which decreasing colorectal cancer mortality can be attributed to earlier detection of colorectal cancer and improved methods of treatment is debated [10]. Unfortunately, those in less developed regions, where proper resources are lacking, suffer poorer prognoses.

Gastric cancer

Gastric cancer is the second most common gastrointestinal cancer and the fourth most common cancer worldwide. It was responsible for nearly 1 million new cancer cases and approximately 737,000 cancer deaths in 2008, making it the number one GI-related cancer killer worldwide. More than 70 % of the new cases and more than 75 % of deaths occurred in less developed regions. The incidence rate of gastric cancer is greatest in the Western Pacific, with nearly half of all cases being found in China (463,000 cases) and with highest incidence rates among the Republic of Korea and Japan. The lowest rates of gastric cancer can be found in Africa, Southeast Asia, and the Eastern Mediterranean regions. Regional variation may be partially attributed to differences in dietary patterns and the prevalence of *Helicobacter pylori* infection [8]. While gastric cancer is one of the leading causes of cancer death, individuals with gastric cancer in the Western Pacific tend to have better prognoses than those in other regions, possibly due to the increased use of screening methods and earlier detection of cancer [11].

Selected diseases of the liver

Hepatitis B

An estimated 2 billion people worldwide have been infected with the hepatitis B virus (HBV). More than 350 million people have chronic liver infections, and approximately 600,000 persons die annually due to acute or chronic consequences of the virus. Hepatitis B is estimated to be the cause of 30 % of cirrhosis and 53 % of hepatocellular carcinoma [12]. Hepatitis B is endemic in China and other parts of Asia, with most infections occurring during childhood, and 8–10 %

of the adult population being chronically infected. In contrast, less than 1 % of the population in Western Europe and North America is chronically infected [13].

In developing countries, HBV is largely transmitted during childbirth and early childhood infections. In developed countries, transmission is primarily through high-risk sexual behavior and IV drug use, as well as from migration of infected individuals from high prevalence areas [14]. Those infected at a young age are most likely to develop chronic infections. Whereas about 90 % of infants <1 year infected with HBV will develop chronic infections, about 90 % of healthy adults who are infected will completely recover within six months. Approximately 25 % of adults who become chronically infected during childhood die from HBV-related liver cancer or cirrhosis [15].

Hepatitis C

An estimated 3–4 million people are infected with hepatitis C virus (HCV) each year with a total of 130–170 million people chronically infected internationally. Additionally, more than 350,000 people die from hepatitis C-related liver diseases annually. Hepatitis C is estimated to be the cause of 27 % of cirrhosis and 25 % of hepatocellular carcinoma worldwide [12]. Although HCV infection is found worldwide, high rates of infection are found in Egypt (22 %), Pakistan (4.8 %), and China (3.2 %) [16]. The main mode of transmission in these countries is secondary to injections using contaminated needles. Other modes of transmission include contaminated blood transfusions, organ transplants, IV drug use with contaminated needles, and pre- or perinatal transmission from an HCV-infected mother.

Viral hepatitis in the United States

It is clear that the toll of hepatitis B and hepatitis C infections is significant worldwide. Interestingly, data from the US Centers for Disease Control and Prevention (CDC) demonstrates a decrease in reported cases and incidence of hepatitis B and C in the United States (Table 1.2) [17]. The incidence per 100,000 population of acute hepatitis B has decreased from 3.8 in 1998 to 1.3 in 2008. Also, the incidence per 100,000 population of acute hepatitis C has decreased from 1.3 in 1998 and has been ≤ 0.3 since 2003. The cause of these secular trends remains unclear, but may reflect changing practices in the IV drug user community, or a cohort effect.

Table 1.2 Incidence per 100,000 population of acute hepatitis B and hepatitis C in the United States by year, 1998–2008

	Hepatitis B		Hepatitis C	
Year	Number	Incidence	Number	Incidence
1998	10,258	3.8	3,518	1.3
1999	7,694	2.8	3,111	1.1
2000	8,036	2,9	3,197	1.1
2001	7,844	2.8	1,640[c]	0.7[c]
2002	8,064	2.8	1,223[d]	0.5[d]
2003	7,526	2.6	891[d]	0.3[d]
2004	6,212	2.1	758	0.3
2005	5,494	1.8	694	0.2
2006	4,713[a]	1.6[a]	802	0.3
2007	4,519	1.5	849	0.3
2008	4,033[b]	1.3[b]	878[b]	0.3[b]

Source: CDC Viral Hepatitis Statistics and Surveillance.
[a]Excludes cases from Arizona.
[b]Excludes cases from Delaware.
[c]Excludes cases from New Jersey and Missouri.
[d]Excludes cases from Missouri.

Liver cancer

Liver cancer is the third most common gastrointestinal cancer and the fifth most common cancer worldwide. Almost 750,000 new liver cancer cases and 700,000 deaths are estimated to have occurred in 2008, with over 80 % of new cases and deaths occurring in less developed regions. There were an estimated 694,000 deaths from liver cancer in 2008, and because of its high fatality (overall ratio of mortality to incidence of 0.93), liver cancer is the third most common cause of death from cancer worldwide. Within liver cancers, hepatocellular carcinoma constitutes the major histological subtype, accounting for 70–85 % of the total liver cancer toll worldwide. Cholangiocarcinomas (intra- and extrahepatic bile duct cancers) are relatively rare, but high rates have been found in areas such as Thailand and other parts of eastern Asia secondary to endemic liver fluke infection [8].

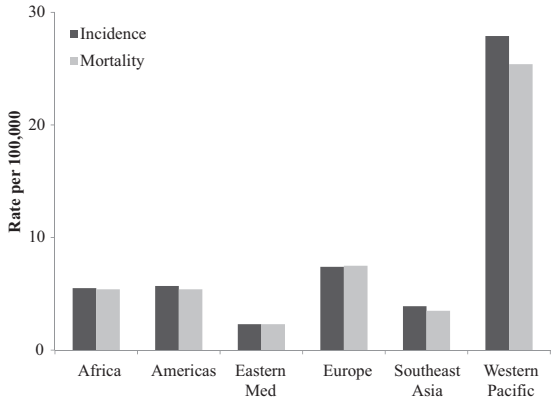


Figure 1.2 Incidence and mortality rates of liver cancer by WHO region, 2008. Source: GLOBOCAN 2008.

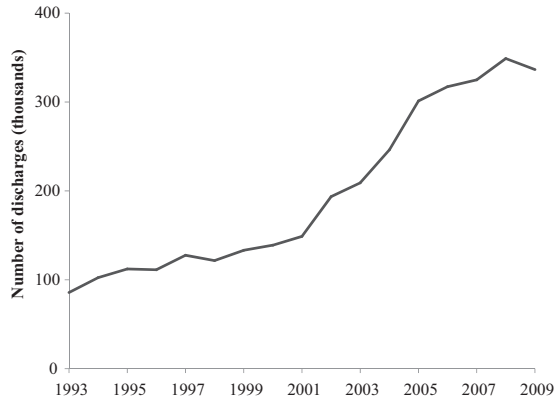


Figure 1.3 Trend of *Clostridium difficile* infection discharge diagnoses from hospital admissions, 1993–2009. Source: HCUP Nationwide Inpatient Sample (NIS), 1993–2009.

Figure 1.2 shows the distribution of liver cancer incidence and mortality by WHO region. The highest incidence and mortality rates are found in the Western Pacific, with more than half of new cases and deaths occurring in China [9]. Incidence and mortality rates are significantly lower in all other regions. The significantly higher incidence of liver cancer in the Western Pacific is largely due to the elevated prevalence of chronic hepatitis B virus (HBV) infection. HBV infection is responsible for approximately 60 % of total liver cancer in developing countries and for about 23 % of total liver cancer in developed countries [18]. Similarly, chronic hepatitis C virus (HCV) infection accounts for about 33 % and 20 % of total liver cancers in developing countries and developed countries, respectively.

Selected gastrointestinal diseases

Clostridium difficile infections

Clostridium difficile is a spore-forming, gram-positive bacillus that can cause disease ranging from mild diarrhea to fulminant colitis and death. This pathogen is recognized as the most common infectious cause of healthcare-related diarrhea [19]. Mutations that confer antibiotic resistance, increase toxin production, or facilitate sporulation have substantially increased the prevalence and virulence of this opportunistic pathogen [20]. During the mid and late 1990s, the reported incidence of *C. difficile* infection (CDI) in acute care hospitals in the United States remained stable at 30–40 cases per 100,000 population. In 2001, this number rose to almost 50 and continued to increase, resulting in 84 per 100,000 reported cases in 2005, a nearly threefold increase since 1996 [21]. Figure 1.3 displays the trend of US hospital discharge diagnoses of CDI over a 17-year period (1993–2009). Parallel to the increasing prevalence of this disease is its increasing severity and fatality. For example, in England, CDI was listed as the primary cause of death for 499 patients in 1999, 1998 patients in 2005, and 3393 patients in 2006 [21].

In addition, while CDI has traditionally affected elderly or severely ill hospital and nursing home patients, a 2005 US CDC advisory noted increased infection in populations not previously considered at risk, including young and healthy persons who have not been exposed to a hospital or healthcare environment or antimicrobial therapy [22]. Transmission in such cases may be attributable to close contact with patients who have CDI and direct person-to-person spread.

Gastroesophageal reflux disease

A major trend in gastroesophageal reflux disease (GERD) is an observed increase in its prevalence over the past two decades. Europe and North America have shown an increase in the prevalence of reflux symptoms, and studies of the same source population

over time have demonstrated an increase in prevalence in the United States, Singapore, and China [23]. Prevalence in Western countries has been estimated at 10–20 %, using criteria of at least weekly heartburn and/or acid regurgitation [24]. According to a review using the US National Ambulatory Medical Care Survey (NAMCS), the rate of US ambulatory care visits for GERD increased from 1.7 per 100 persons to 4.7 per 100 persons from 1990–1993 to 1998–2001 and continues to be a frequent cause of consultation in primary care [25].

The incidence of a GERD diagnosis and the demographic factors associated with the diagnosis were assessed using the UK General Practice Research Database [26]. In this study, 7159 patients were identified with a new GERD-related diagnosis in 1996, corresponding to an incidence among individuals aged 2–79 years of 4.5 new diagnoses per 1000 person-years. The incidence was age-related and increased with age until 69 years, with a slight decrease thereafter. Women had a slightly higher risk of developing GERD than men in patients over 50 years of age (rate ratio = 1.3).

Inflammatory bowel disease

Although a major cause of gastrointestinal illness and healthcare utilization, reliable data on inflammatory bowel disease rates are not available in most regions of the world. When examining the age-adjusted time trends of US physician visits secondary to Crohn's disease and ulcerative colitis (UC) from 1960–2006, physician visits for Crohn's disease increased almost fourfold over a 30-year period from the early 1960s to the early 1990s, from about 120 to 400 physician visits per 100,000 people. Since then, the rates of Crohn's disease visits appear to have leveled off. Physician visits for UC actually slightly decreased during the same 30-year period from about 400 to 300 physician visits per 100,000 people. With respect to sex differences, physician visits for Crohn's disease remained about 1.4-fold more frequent in women than men. Between 1960 and 1984, physician visits for UC were 1.3-fold more frequent by women than by men; however, during more recent periods, the rates of physician visits for UC by men and women have become more similar [27].

From 1951 to 2005, there has been a nearly 80 % decrease in mortality from UC from approximately 5.6 to 1.2 deaths per million population in a total of 21 countries [28]. On the other hand, from 1951 to 1975, mortality from Crohn's disease increased almost twofold from 0.8 to 1.5 deaths per million population. Since then, mortality from Crohn's disease has been decreasing and paralleling the trend of UC.

Gastrointestinal diseases responsible for hospitalization

While gastrointestinal illness is a major cause of hospitalization throughout the world, reliable data on hospitalization rates for various illnesses are not available internationally. Table 1.3 demonstrates the most common gastrointestinal and liver causes of hospitalization, ordered by number of reports at discharge, using the National Inpatient Sample, a 20 % stratified sample of US community hospitals. Acute pancreatitis, gallstone diseases, diverticulitis without hemorrhage, and acute appendicitis were each responsible for over 200,000 hospitalizations. Aspiration pneumonia was the fifth cause of hospitalization, and was also in the overall top 30 causes of hospitalization for any disease entity.

Limitations of the data

The data that were used for the above analyses are of the highest quality information available to assess the overall global burden of gastrointestinal diseases. However, there are some limitations that merit attention.

Ideally, all data would come from vital registries with complete coverage and medical certification of cause of death. For countries with incomplete or no vital registration system, epidemiologic studies, systematic reviews, and statistical modeling were used. For countries with incomplete data or no data regarding cause of death, the distribution of deaths was estimated using statistical models, proportional mortality, and natural history models. The 2008 estimates made available by the WHO were created using WHO's extensive databases and based on information provided by Member States, as well as on systematic reviews and analyses carried out by CHERG (the Child Health Epidemiology Reference Group).

Table 1.3 Most common gastrointestinal principal discharge diagnoses from hospital admissions, 2009*

Rank among GI dx	Rank among all dx	ICD-9-CM code(s)	Principal diagnosis	Total # Admissions	% Δ from 2000	Median LOS (days)	Total hospital days† (thousands)	Median costs (USD)	Aggregate cost (USD, thousands)	In-hospital deaths n (%)
1	21	577.0	Acute pancreatitis	274,119	+30	4.0	1,409	6,096	2,599,686	2,631 (1.0)
2	41, 76	574.0, 574.1	Cholelithiasis with cholecystitis‡	226,216	−14	3.0§	819	8,322§	2,208,531	959 (0.4)
3	27	562.11	Diverticulitis without hemorrhage	219,133	+41	4.0	1,099	6,077	2,115,989	1,235 (0.6)
4	29	540.9	Acute appendicitis	207,345	+22	1.0	362	6,592	1,491,402	90 (0.04)
5	30	507.0	Aspiration pneumonia	188,930	+6	6.0	1,475	9,030	2,523,299	22,273 (11.8)
6	37	558.9	Noninfectious gastroenteritis/colitis	151,856	+36	2.0	462	4,090	775,020	486 (0.3)
7	44	578.9	Gastrointestinal hemorrhage NOS	140,497	+22	3.0	612	6,090	1,155,971	4,914 (3.5)
8	46	560.9	Intestinal Obstruction NOS	134,431	+44	3.0	600	5,098	1,018,437	2,812 (2.1)
9	47	278.01	Morbid obesity	132,448	+314	2.0	288	10,689	1,642,293	137 (0.1)
10	57	8.45	*Clostridium difficile* infection	110,553	+237	5.0	761	6,774	1,119,213	4,038 (3.7)
11	73	560.81	Intestinal adhesions with obstruction	83,183	+23	7.0	736	11,853	1,453,238	2,265 (2.7)
12	94	8.8	Viral gastroenteritis	66,842	+29	2.0	181	3,677	298,507	108 (0.2)
13	96	530.81	Esophageal reflux	65,634	−32	2.0	162	4,366	386,229	n/a**
14	100	562.12	Diverticulosis with hemorrhage	64,222	−6	3.0	291	5,818	552,906	713 (1.1)

*Weighted national estimates from HCUP Nationwide Inpatient Sample (NIS), 2009, Agency for Healthcare Research and Quality (AHRQ). Total number of weighted discharges in the United States based on HCUP NIS = 39,434,956.

†Total hospital days per year for all persons with each diagnosis, estimated by the product of number of discharges and mean LOS.

‡ICD-9-CM codes for "calculus of gallbladder with acute cholecystitis" (574.0) and "calculus of gallbladder with other cholecystitis" (574.1) combined for this diagnosis. Total number of discharges, aggregate charges and costs, and in-hospital deaths represent sum from both.

§Median LOS and median costs presented for most common ICD-9-CM codes in this category (574.0).

**Too few events to generate a stable estimate in this category [(standard error/weighted estimate) > 0.30].

Dx, diagnosis; ICD-9-CM, *International Classification of Diseases*, 9th edition, Clinical Modification; %Δ, percent change; LOS, length of stay; USD, US dollars; NOS, not otherwise specified.

Incidence data for cancers are associated with some level of delay as this type of data requires time to be compiled; while the numbers within this chapter are the most current available, there is a several year time lag. More recent data about individual regions may be found in reports from the registries themselves. Information from most of the developing countries may be considered of relatively limited quality, but this information remains the only source of information for these regions. Mortality statistics collected and made available by the WHO have the advantages of national coverage and long-term availability; however, some datasets are of lesser quality than others. For some countries, coverage of the population is incomplete, resulting in low estimated mortality rates. In other countries, the quality of cause of death information is poor. While almost all the European and American countries have comprehensive death registration systems, most African and Asian countries (including the populous countries of Nigeria, India, and Indonesia) do not. Of course, a major concern regarding data from developing countries is detection bias. In countries with limited medical technology and resources, the burden of undiagnosed cancer is likely substantial and is not quantifiable.

Data for some of the selected gastrointestinal illnesses was not readily available from regional databases; therefore, the data in the above discussion is largely from studies that have accessed such primary databases and performed their own analyses.

Data derived from administrative databases, such as the NIS data, may suffer from the use of data primarily for billing purposes. Therefore, the fidelity of coding data to clinical information must be considered. The median and aggregate costs are estimates, calculated from hospital charges, and the data are by level of discharge (e.g. a single patient could be represented by multiple discharges). Also, in analyzing the trends, some trends may represent epiphenomena. For example, an increase in morbid obesity discharges may be due to increasing popularity of obesity surgery, for which morbid obesity is the principal coded discharge diagnosis.

Implications

Gastrointestinal and liver diseases are responsible for significant morbidity and mortality worldwide. The above statistics attest to the toll of these diseases. Beyond merely describing the terrible impact of these diseases, an understanding of the epidemiology of gastrointestinal and liver disease allows consideration of improvement of systems-based practices and public policy. Many individuals suffer from preventable disease states such as childhood diarrhea, malignancy, and various liver diseases. Millions of children annually die preventable deaths due to diarrheal disease. Cancer prognosis may be poorer in developing countries due to late detection and lack of access to resources and standard treatment. Numerous cases of gastrointestinal cancers could be prevented by vaccinations for viral hepatitis and improved screening, as well as by promoting physical activity, implementing programs for tobacco control, and healthier dietary intake. In addition, data should be updated regularly in order to track progress, as well as to spot temporal trends in disease burden that might merit reallocation of resources to address the changes.

Multiple choice questions

1 Which of the following is not associated with an increased incidence of childhood diarrhea?

A Inconsistent access to a clean water supply
B Residing in the WHO Africa or Southeast Asia region
C Chronic nutritional deficiencies
D Availability of oral rehydration solutions

2 Which GI-related malignancy resulted in the most estimated number of deaths in the year 2008?

A Colorectal cancer
B Stomach cancer
C Liver cancer
D Esophageal cancer
E Pancreatic cancer

3 Which gastrointestinal principal discharge diagnosis has had the greatest percentage increase from 2000 to 2009?

A *Clostridium difficile*
B Acute pancreatitis
C Morbid obesity
D Intestinal obstruction NOS
E Diverticulitis without hemorrhage

Appendix 1.A

Sources

Diarrheal disease

Estimates used in this section are based on data from the Global Health Observatory (GHO), a repository that provides access to over 50 datasets on priority health topics including mortality and burden of disease, produced by the World Health Organization (WHO) (http://apps.who.int/ghodata/). Estimates for the distribution of causes of death among children aged <5 years can be accessed through "World Health Statistics" → "Cause-specific mortality and morbidity" → "Causes of death among children" of the GHO data repository. Measurement and estimation methods can be found at: http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=89.

In collaboration with the Child Health Epidemiology Reference Group (CHERG), the WHO Department of Health Statistics and Informatics prepared country-level estimates of child deaths under 5 years of age by cause for the year 2008. These estimates are derived from WHO databases, information provided by Member States, as well as systematic reviews and analyses carried out by CHERG. Country-level data was combined to achieve data for each WHO region. Mortality data on diarrheal disease and other causes of death in children aged <5 years, as well as the methods used to obtain these estimates can be accessed at: http://www.who.int/healthinfo/statistics/mortality_child_cause/en/index.html.

Gastrointestinal malignancies

The estimates used in this section are based on GLOBOCAN 2008, a standard set of worldwide estimates of cancer incidence and mortality produced by the International Agency for Research on Cancer (IARC) under the auspices of WHO. This project was developed to provide up-to-date estimates of the incidence of, and mortality from major cancers for all nations in the world. GLOBOCAN allows individuals to obtain current estimates for major cancers categorized by region, sex, and age groups.

Incidence data were derived from population-based cancer registries, either national or subnational areas. In developing countries, incidence data is often available only from major cities. Mortality data was collected and provided by WHO. While not all datasets are complete and of the same quality (coverage of the population is incomplete or cause of death is inaccurate), it is the most accurate and thorough information available. Provisional estimates of age- and sex-specific deaths from cancer for 2008 have been used for regions without death information or where statistics are considered unreliable. National population estimates for 2008 were extracted from the United Nation (UN) population division's 2008 revision using geographical definitions as defined by the UN. The methods used to estimate incidence and mortality of cancers for each country can be found at GLOBOCAN data sources and methods: http://globocan.iarc.fr/.

Selected diseases of the liver

The data used in the discussions of hepatitis B and C are derived from WHO estimates of burden of disease. The WHO media center has over 100 fact sheets on various health-related topics such as different infections, disease states, and health risks, which can be found at: http://www.who.int/mediacentre/factsheets/en/. The hepatitis B and hepatitis C fact sheets were last updated in July 2012. The data included in the discussion about liver cancer is derived from GLOBOCAN 2008, discussed in the previous sources section.

Hepatitis B and hepatitis C trend data were obtained from the Centers for Disease Control and Prevention (CDC) – Viral Hepatitis Statistics and Surveillance, which can be found at: http://www.cdc.gov/hepatitis/Statistics/index.htm.

Gastrointestinal diseases responsible for US hospitalization

The most common inpatient gastroenterology and hepatology discharge diagnoses for the United States may be compiled using the Nationwide Inpatient Sample (NIS). The NIS is one of the databases in the Healthcare Cost and Utilization Project (HCUP) (http://hcupnet.ahrq.gov/). NIS is the only national hospital database with charge information on all patients, regardless of payer, including persons covered by Medicare, Medicaid, private insurance, and the uninsured. The most recent version, NIS 2009, contains all discharge data from 1050 hospitals located in 44 states, representing a 20 % stratified

sample of US community hospitals. The sampling frame for the 2009 NIS sample is a sample of hospitals that comprises approximately 95 % of all hospital discharges in the United States.

The NIS database was queried for the rank order of the principal discharge diagnosis (i.e. *International Classification of Diseases Clinical Modification*, 9th edition (ICD-9-CM) for all patients in all hospitals. From the top 100 diagnoses, we identified the gastroenterology and hepatology diagnoses among them, which were subsequently rank-ordered after combining related diagnosis codes. We then performed a separate query for each individual ICD-9-CM code (or group of codes) to acquire data on mean and median length of stay (LOS), median charges and costs, aggregate charges (i.e. "the national bill") and aggregate costs, and number of inpatient deaths associated with each diagnosis or diagnosis group. We also performed a temporal analysis for the number of admissions for the top principal GI diagnoses between the years 2000 and 2009 to identify relevant trends.

Total hospital days were estimated by the product of the mean LOS and the number of discharges for each diagnosis. Total charges were converted to costs by HCUP using cost-to-charge ratios based on hospital accounting reports from the Centers for Medicare & Medicaid Services (CMS). Cost data are presented preferentially, as costs tend to reflect the actual costs of production, while charges represent what the hospital billed for the case.

References

1 Global Health Observatory [Internet database]. World Health Organization (2008). Available from: http://apps.who.int/ghodata/ (accessed July 27, 2011).

2 Johansson E, Wardlaw T. (2009) *Diarrhoea: Why children are still dying and what can be done*, UNICEF/World Health Organization, New York/Geneva.

3 Boschi-Pinto C, Lanata CF, Black RE. (2009) The global burden of childhood diarrhea, in *Maternal and Child Health: Global Challenges, Programs, and Policies* (ed. J Ehiri), Springer, pp. 225–43.

4 Child mortality by cause [Internet database]. World Health Organization (2008). Available from: http://www.who.int/healthinfo/global_burden_disease/cod_2008_sources_methods.pdf (accessed July 27, 2013).

5 Snyder JD, Merson MH. The magnitude of the global problem of acute diarrhoeal disease: a review of active surveillance data. *Bull World Health Organ* 1982;60(4):605–13.

6 Bern C, Martines J, de Zoysa I, Glass RI. The magnitude of the global problem of diarrhoeal disease: a ten-year update. *Bull World Health Organ* 1992;70(6):705–14.

7 Kosek M, Bern C, Guerrant RL. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bull World Health Organ* 2003;81(3):197–204.

8 Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA: Cancer J Clin* 2011;61(2):69–90.

9 Ferlay J, Shin HR, Bray F, et al. (2010) GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet database]. International Agency for Research on Cancer, Lyon, France. Available from: http://globocan.iarc.fr (accessed August 11, 2011).

10 Sandler RS. Editorial: colonoscopy and colorectal cancer mortality: strong beliefs or strong facts? *Am J Gastroenterol* 2010;105(7):1633–5.

11 Lee KJ, Inoue M, Otani T, et al. Gastric cancer screening and subsequent risk of gastric cancer: a large-scale population-based cohort study, with a 13-year follow-up in Japan. *Int J Cancer* 2006;118(9):2315–21.

12 Perz JF, Armstrong GL, Farrington LA, et al. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol* 2006;45(4):529–38.

13 Hepatitis B [Internet database]. WHO Media centre. Available from: http://www.who.int/mediacentre/factsheets/fs204/en/index.html (last accessed May 6, 2013).

14 Ahmed F, Foster GR. Global hepatitis, migration and its impact on Western healthcare. *Gut* 2010;59(8):1009–11.

15 Te HS, Jensen DM. Epidemiology of hepatitis B and C viruses: a global overview. *Clin Liver Dis* 2010;14(1):1–21, vii.

16 Hepatitis C [Internet database]. WHO Media centre. Available from: http://www.who.int/mediacentre/factsheets/fs164/en/index.html (accessed August 19, 2011).

17 Viral Hepatitis Statistics and Surveillance [Internet database]. Centers for Disease Control and Prevention (CDC). Available from: http://www.cdc.gov/hepatitis/Statistics/index.htm (accessed September 08, 2011).

18 Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 2006;118(12):3030–44.

19 Dubberke ER, Wertheimer AI. Review of current literature on the economic burden of *Clostridium difficile* infection. *Infect Control Hosp Epidemiol* 2009;30(1):57–66.

20 Freeman J, Bauer MP, Baines SD, et al. The changing epidemiology of *Clostridium difficile* infections. *Clin Microbiol Rev* 2010;23(3):529–49.
21 Kelly CP, LaMont JT. *Clostridium difficile* – more difficult than ever. *New Engl J Med* 2008;359(18):1932–40.
22 Centers for Disease Control and Prevention (CDC). Severe *Clostridium difficile*-associated disease in populations previously at low risk – four states, 2005. *MMWR Morb Mortal Wkly Rep* 2005;54(47):1201–5.
23 El-Serag HB. Time trends of gastroesophageal reflux disease: a systematic review. *Clin Gastroenterol Hepatol* 2007;5(1):17–26.
24 Dent J, El-Serag HB, Wallander MA, Johansson S. Epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* 2005;54(5):710–7.
25 Altman KW, Stephens RM, Lyttle CS, Weiss KB. Changing impact of gastroesophageal reflux in medical and otolaryngology practice. *Laryngoscope* 2005;115(7):1145–53.
26 Ruigomez A, Garcia Rodriguez LA, Wallander MA, et al. Natural history of gastro-oesophageal reflux disease diagnosed in general practice. *Aliment Pharmacol Ther* 2004;20(7):751–60.
27 Sonnenberg A, Chang J. Time trends of physician visits for Crohn's disease and ulcerative colitis in the United States, 1960–2006. *Inflamm Bowel Dis* 2008;14(2):249–52.
28 Sonnenberg A. Time trends of mortality from Crohn's disease and ulcerative colitis. *Int J Epidemiol* 2007;36(4):890–9.

Answers to multiple choice questions

1. D
2. B
3. C

PART TWO

How to Critically Read the Gastrointestinal Epidemiology Literature

Introduction and overview

Joe West
Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK

This series of chapters will take the reader through the process of critically appraising the epidemiologic literature with specific reference to aspects relevant to gastrointestinal epidemiology. The chapters have focused on reading papers that report the findings of cohort studies, case-control studies, randomized controlled trials, and systematic reviews, and interpreting the results in the context of clinical practice. As outlined in the final chapter, the notion of evidence-based medicine relies heavily on these study designs, hence our scrutiny of their methodology. Each of the first four chapters gives an example approach to making an assessment of whether the paper you are reading is of sufficient quality for its findings to be credible. Within each chapter this systematic approach will be used to appraise an original piece of research in the field of aspirin and colorectal cancer as a practical example. The final chapter, Chapter 6, will take an overview of the whole process while challenging the reader to form their own opinion of the value of the work they are appraising with respect to their clinical and research practice. For all of the study types, we will consider the first question one should ask of any paper being read, namely "What is the research question?" All epidemiologic studies require a clear and precise question that includes a description of what the study is trying to achieve, in whom, and for what purpose. If this is not clear from the introduction to the paper you may as well put it in the recycling bin before reading further as it is often a good indication of the quality (or lack of quality) of the work described thereafter. However, should you be persuaded to read the paper in its entirety, the methods outlined in the chapters that follow give a structured approach to deciding on the quality of the work presented and the specific issues that arise in each of the designated study designs, with particular reference to the gastrointestinal epidemiology literature.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

2 How to read a cohort study

Laila J. Tata
Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK

Key points

- The cohort is the basis of all epidemiologic study designs as it is the closest way to study the natural progression of people's life course over which the temporal relationship between exposures and outcomes can be assessed.
- Health outcomes are compared between people grouped on the basis of whether or not they have certain exposures or between groups with different levels of exposure.
- Selection bias, ascertainment bias, and follow-up bias are common and their potential impact must be considered along with confounding and chance.
- A good understanding of the cohort design provides the foundation for appraising all study designs particularly experimental trials.

Brief introduction to cohort studies

The term cohort (from the Latin *cohors*, originally describing a specific unit of soldiers in the Roman military) has been widely adopted by epidemiologists and throughout medical science. A cohort is often considered as a specified group of people who are identified at a similar place and time and then followed over a certain period. The general concept of tracking over time has resulted in loose terminology, such as follow-up study, longitudinal study or prospective study, often used as synonyms of cohort study [1]. When epidemiologists discuss the design and conduct of a cohort study, however, they are most often interested in comparing the occurrence of disease between groups of people who have or do not have a certain exposure, or between groups with different levels of exposure. Cornerstone to the definition of a cohort study is the measurement of outcomes over time and, for etiologic purposes, the identification of exposures and exposure timing before outcomes have occurred. In this sense, a cohort study is the closest way to study the natural progression of people's life course and in fact the ultimate cohort study would do just that, measure and record all exposures that may lead to the occurrence of any number of events over a lifetime, and even better, over generations.

The great advantage of cohort studies for studying etiology of disease and the relative effects of risk factors on later development of health outcomes is that certainty about temporal relationships can be established. The use of follow-up time between exposures and outcomes not only permits comparisons of how much disease will occur between different cohorts, but also provides understanding of whether one cohort

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

is more or less likely to develop disease sooner than other cohorts. We could say with certainty that death will occur in 100 % of all cohorts, although it would be useful to know whether people in one cohort are more likely to die within the next 10 years or whether their median survival time differs greatly from another cohort.

New occurrence of disease (e.g. celiac disease diagnosis), recurrent events (e.g. episodes of gastrointestinal bleeding), and different types of events (e.g. all-cause mortality and cause-specific mortality) can all be measured in cohort studies. When studying diseases such as cancer that usually develop a long time after exposure, costs and logistics of running cohort studies escalate which has led to the establishment of large research cohorts, enabling a number of different exposures and outcomes to be studied. The British birth cohorts [2], the 1949 Framingham Heart Study [3], the 1951 British Doctors study [4], the Nurses' Health Study [5], and the more recent European Prospective Investigation into Cancer (EPIC) study [6] are a few examples. These are nevertheless cost- and labor-intensive observational studies and the need to wait for outcomes to occur also means that cohort studies may be impractical for studying very rare outcomes, unless preexisting complete data are available on large numbers of people. Alternatively, the cohort design is particularly useful for studying rare exposures because participants can be chosen based on their exposure. For example, if only 5 % of people in a given population had an exposure, selection of 100 people at random from the population would likely result in only 5 exposed people to follow up, whereas one could select people based on their exposure to obtain 50 exposed and 50 unexposed.

Cohort study principles are the same whether a population was identified and followed up prospectively or whether an historical cohort was identified using existing records of information on exposures and outcomes in the past, such as government and occupational records or linkage of different data sources. In an historical cohort study with the aim of studying etiology of disease, it is still important to ensure that exposures occurred, and were ideally recorded, before the outcomes.

An understanding of cohort studies will aid the appraisal of all study designs including experimental trial designs. All randomized controlled trials (RCTs) are essentially cohort studies [7], the only difference being that in a trial the researcher assigns exposures to groups in the study population whereas in a cohort study the researcher must use existing exposures to divide the population into groups. Issues surrounding selection bias, losses to follow-up, ascertainment bias and blinding can all be learned from cohort study methods. These and other issues that are important for evaluating the quality of cohort studies will be discussed in this chapter through use of a 10-point checklist. This will then be practically applied in the appraisal of a study of regular aspirin use for the prevention of colorectal cancer [8].

Biases commonly seen in cohort studies

The collection of exposure information before outcome information in a cohort study avoids a battery of problems introduced when people have to recall or report information from the past. Selection bias, ascertainment bias and follow-up bias, however, may all have important impacts on cohort studies. Selection bias may arise when there are important differences in study subjects other than the exposure of interest (see point 1 of checklist). Ascertainment bias may occur if outcomes are obtained differently between exposed and unexposed groups (see point 3 of checklist). Follow-up bias may occur if there are differences in losses to follow-up that are related to participants' exposure status (see point 5 of checklist).

10-Point checklist of important issues when reading a report of a cohort study

Determining whether a study is in fact a cohort design can be a more difficult task than one first assumes, primarily because many studies that are *called* cohort studies by the authors are in fact not cohort studies. The first point below should help with this initial hurdle.

1 How has the study population been identified and selected?

If a study's participants were selected based on whether or not they have an outcome, which is the outcome one is attempting to predict the risk of in the study, it is not a cohort study.

Cohort participants may be first identified and selected as a group representative of a population with one or more common characteristics (e.g. year of

birth, sex, area of residence, people with diabetes) and then subdivided based on their exposures or degree of exposure, a common approach in large population-based cohorts. Alternatively, if the study was originally designed only to address a specific exposure-outcome relationship, individuals may have been identified by their exposure experience, which is often necessary for studies of rare exposures such as occupational radiation or history of childhood X-ray exposure. In this situation a cohort without the specific exposure of interest would also need to be selected for comparison. Using either method of selection, the aim is to achieve the following principles:

- ***Exposed and unexposed cohorts should be assessed for similarities and differences other than the exposure of interest***

The comparison group without the exposure is often misleadingly called a "healthy control group" or "healthy comparisons" – in fact the unexposed comparison group should not be a group that is generally healthier than the exposed group as this could lead to a bias, making the exposure of interest appear worse than it truly is.

Paramount in selection of the study population is to obtain exposed and unexposed cohorts that are ideally similar in all ways other than the exposure of interest. While this is in fact very difficult in practice (e.g. regular exercisers have many important differences to nonexercisers), the reader should question whether, *other than the exposure under study, could the method of selecting the study participants have led to other important differences between the groups?* If there is evidence of this *selection bias*, the degree to which this will affect the relationship of interest must be addressed.

- ***All study participants must be at risk of the outcome***

When first identified cohort participants should be alive and at risk of the outcome, but not typically have the outcome of interest already, as the objective is to study the outcome's occurrence over the study period. Ensuring that women are not selected for a study of prostate cancer or that women with hysterectomies are not selected for a study of uterine cancer is straightforward, as neither will be at risk of the outcomes. For some diseases, such as epilepsy, it will have been important to ensure that participants did not have preexisting disease as it is not possible to have this as new disease more than once. For outcomes that may recur, such as *Clostridium difficile* infection, whether participants have had infection in the past may not preclude them from being at risk in the study at hand.

2 How was the exposure defined and measured?

- ***Exposures should be clearly defined and measureable, with explicit information on exposure timing***

Whether there are one or more exposures in a study, each should be specific, clearly defined, and measurable among all study participants. The most useful clinical information on the effects of an exposure ideally includes knowing "how much, for how long, and when" all of which are measureable in real time during a cohort study. Participants may be defined by their past (e.g. childhood X-ray exposure), present or future exposures, by a specific time window (e.g. first trimester exposure to antacid drugs), or by levels of exposure.

- ***Objective measures should be used where possible***

Exposures may be measured using standardized diaries kept by study participants, or questionnaires at regular intervals such as weekly or yearly, but as for all research, the most objective measures from which adequately complete and specific information can be obtained should be used. Portable exercise monitors, hospital records of surgery, blood measurement, or medication use confirmed by prescription records are some examples.

3 What is the outcome and how was it ascertained?

- ***Outcomes should be clearly defined and measureable, with explicit information on whether they are first or recurrent events***

Outcomes should be clearly defined and measurable in all study participants, using the most objective methods possible. When recurrent outcomes are possible during the study period, such as gastrointestinal bleed or myocardial infarction, methods should have been devised to distinguish new events from previous or historical events. This can present difficult clinical and methodological decisions, particularly with cancer outcomes which may recur in one or more sites. First occurrence of disease is often easier to define, measure, and interpret in context of an exposure of interest, although risk of recurrent disease will be important in certain clinical contexts.

• ***Outcomes should be defined, ascertained, and measured in the same way for exposed and unexposed groups***

Procedures to identify outcomes should ideally be identical for all study participants because differences in outcome ascertainment between exposed and unexposed groups can introduce *ascertainment bias*. Wherever possible, objective measures of disease such as cancer registrations should be used. Bias can also be avoided if study researchers measuring outcomes, and ideally also the study participants, are blinded to the exposure status or the hypothesis being tested.

• ***Outcomes within a reasonable time after an etiologic exposure should be considered***

It is unlikely that starting to smoke or taking antacids this month will cause esophageal cancer next month, yet a vaccination today may cause a severe reaction tomorrow. Therefore, when studying etiologic exposures, the time when outcomes occurred in relation to the exposures should be clearly communicated with logical consideration of the etiologic time window. Estimation of this will include the *induction time*, which is between the initiation of the exposure to a causal agent and the initiation of the disease process, and the *latency period*, which is the time following exposure to diagnosis of the disease [1].

4 How has person time been dealt with in this study?

Whether it is days or years, the amount of time each person contributes to a cohort study will affect their opportunity to have an outcome, so it is crucial to understand whether person-time has been considered and whether it has been dealt with appropriately.

• ***Study entry and exit times must be clearly defined for all participants***

These may be defined in any number of ways but should be consistently defined. In a closed cohort study entry times will be the same for all participants (e.g. all born within a specific year, all hospitalized within a specific month) whereas open cohorts may have wide variation in entries across calendar time (e.g. new general practice registration sometime between 2000 and 2010). When studying time to first disease occurrence, the person would exit the study at the time of the event and any person-time after would be excluded. Alternatively, recurrence studies would also consider the available time after each event up to the defined end of the study period.

• ***Analytical methods incorporating person-time must be used when there is important variation in the amount of time participants contribute***

If each participant in a disease-free cohort was followed over the same period of time, the proportion with disease could be measured which would be equivalent to the incidence over this study period. For some special situations, such as in pregnancy cohorts, person-time may be less relevant to include in calculations of outcome occurrence. However, the dynamic nature of populations and the reality that loss to follow-up is common in cohort studies, mean that use of each participant's person-time to calculate incidence rates is often required. This can also maximize the value of precise information available on the time to an event.

5 What has been done about loss to follow-up?

Loss to follow-up can have similar study impacts to initial nonresponse in the study population and whilst there is no set cut-off for adequate study follow-up, the extent and the reasons for any losses should be reported.

• ***Methods to minimize loss to follow-up should be described***

It is particularly difficult to retain participants in studies carried out over years or decades; procedures to maintain contact should have been considered and weighed against study costs. Routinely recorded data (e.g. death registries) may provide useful outcome data at low cost. Alternatively, regular questionnaires followed by phone calls or house visits may be needed for more detailed updates on exposures and outcomes.

• ***Impacts of loss to follow-up on study power should be considered***

The amount of loss to follow-up will directly impact the amount of study power because data (person-time and potential outcomes) are being lost. It is thus important to know, for example, whether this is 10 % loss to follow-up or 50 % loss to follow-up.

• ***Impacts of loss to follow-up on bias should be considered***

The impact loss to follow-up can have on bias, often called *follow-up bias*, is just as important as losing study power, if not more so. If there are important differences in loss to follow-up between the exposed and unexposed groups, the exposure may be related to reasons for loss to follow-up which could distort the effect being studied. *Both the amount and ideally the*

reasons for losses should be described separately for the exposed and unexposed groups. In a cancer study with 30 % and 10 % losses to follow-up in exposed and unexposed groups respectively where there were many losses due to death, we would be more concerned about potential bias than in a study with 11 % and 12 % losses where most were because participants moved out of the study area.

6 What has been done about time-varying exposures and their effect on the results?

Exposures like ethnicity will of course never change over a lifetime, but certain exposure measurements that do not exploit the use of time in a cohort study, ever use of aspirin for example, are of limited etiologic use. Where it is likely that exposures will change over time, particularly in long studies of diet and lifestyle characteristics, these exposures should be measured on a periodic basis to best characterize potential causal effects since the timing of exposure initiation and overall duration are often etiologically important. Changes in age and calendar time over a long period often have particularly important impacts on the occurrence of exposures and outcomes.

Exposure information collected over time can be described using summary measures. In more simplistic analyses, this may be an average or maximum duration of overall exposure. In more complex analyses, person-time may be divided into different levels of exposure status to obtain outcome incidence during different windows of exposure (e.g. person-time as a heavy, moderate, or light drinker).

7 Was information about potential confounders collected, and how?

While potential confounders may not be needed for some descriptive study questions (e.g. what is the incidence distribution of liver cancer across England?), studies of exposure-outcome relationships with no information on potential confounders should be interpreted with caution. Information on potential confounders should be clearly defined and comprehensively ascertained along similar principles to collection of exposure and outcome data.

- ***Potential confounders should be defined, ascertained, measured in the same way for exposed and unexposed groups***

Procedures to identify confounders should be identical for all study participants to avoid generating any biases between exposure groups. Objective measures should be used if possible. As with exposures, confounders may change over time, which should be considered when interpreting the study.

- ***Confounding should ideally be assessed in study analyses***

Numerical data on the distribution of potential confounders should be presented and assessed in the exposed and unexposed groups and incorporated into multivariate analyses with appropriate reasoning. Time-varying covariates can also be incorporated into analyses.

- ***Confounding by indication should be considered for studies of drug effects***

A particular type of confounding, *confounding by indication*, should also be considered when a drug of interest could be used to treat early symptoms of the outcome. This can also occur if the condition indicated for the drug treatment is independently associated with the outcome. These situations can be mistaken for a causal association between the drug and the outcome. Ensuring that the exposure occurs in a reasonable amount of time before the outcome can avoid this, but may be practically difficult.

8 How was the sample size determined, and was the study large enough to answer the question?

The number of people in the study, how long they are followed up and the number of outcomes all contribute to the study power. While sample size calculations are useful, each one is only relevant to a specific exposure-outcome combination and so a study with several exposures, outcomes, or subgroup analyses will not be accounted for in one calculation for a primary outcome. Reporting overall numbers of outcomes and person-time provides information that helps interpretation of study power for specific outcomes. If there are too few outcomes and particularly few outcomes in a small exposure group, the likelihood that the study results were due to chance may be high.

9 Were the data analyzed properly? (What statistics have been used and how do I interpret them?)

Virtually any measure can be calculated in a cohort study given that comprehensive data on exposures and outcomes are collected over time. Disease occurrence may be described as a risk or an odds, yet a rate (often called an incidence or an incidence rate) makes full use of denominator person-time data that are unique to cohort studies and should be used

whenever there is variation in follow-up time between study participants. Risk differences, risk ratios, odds ratios, rate differences, and rate ratios are all common measures of effect that are calculated by comparing disease occurrence between the exposure groups.

- ***The measure of effect used should be clearly described***

Determining what measure of effect has been used is crucial for correct understanding and for clinical or public health interpretation. Authors should unambiguously describe the measure of effect, avoiding terms such as "relative risk," and they should present sufficient data on numerators (outcome occurrence) and denominators to allow the reader to identify the effect measure used. In a study of bowel perforation associated with different surgical procedures, a risk ratio of 4 could represent 20 %/5 % perforation in surgery A versus B, a rate ratio of 4 could represent 8/2 perforations per 100,000 person-years in surgery A versus B, and a rate difference of 4 could represent 344-340 perforations per 10,000 person-years in surgery A compared with surgery B. It should be obvious that the clinical and public health importance of these three effect measures is interpreted differently.

- ***All analyses should be justified in relation to the data used***

Beyond basic effect measures, Poisson regression or Cox proportional hazards regression allow calculation of rate ratios (or hazard ratios) with adjustment for important confounders. Cox regression is particularly useful and necessary where there is evidence of the relative outcome incidence changing over the study period. Risk of infection, for example, may be very high in the hours after surgery and then decrease over days and weeks. Plots showing the change in the proportion of the cohort without the outcome over the follow-up time, often called survival curves, are helpful in visualizing whether incidence is changing over time and if there are important differences between exposure groups (Figure 2.1). Age and calendar time have considerable impact on the changing incidence of cancer in long follow-up studies, making hazard ratios, which have a similar interpretation to rate ratios, useful measures in these studies.

In situations where data are only collected for an exposed group [9] standardized mortality (or morbidity) ratios (SMRs) that compare outcome rates with those in a general population are also commonly used.

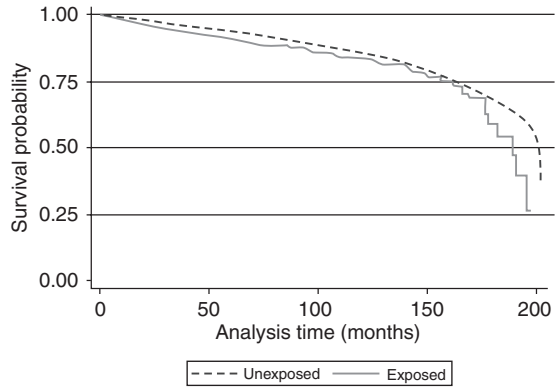


Figure 2.1 Kaplan-Meier curve showing the changing survival probability (or probability of not having an outcome) over time for an exposed and unexposed group. Individuals are censored at the point they have the outcome or if they are lost to follow-up. The median time to the outcome for each group, commonly called *median survival time*, is the plotted time when the survival function, or proportion on the *y*-axis is 0.50.

10 Were the conclusions properly drawn based on the results?

Conclusions should directly address the primary objective of the research and should be supported by the data as presented. As with any design, conclusions must be considered in the context of space and time and not in absolute terms. All cohort studies are carried out in a real-life setting, so dismissing a study based on a fast judgment of one limiting factor such as moderate follow-up bias is unreasonable. Instead, the reader should be considering the *extent* to which loss to follow-up or ascertainment bias may have affected the findings. We are often pressured to decide whether a study is good or bad, whether we do or do not accept the findings as truth, much akin to the reliance on statistical significance testing in science. Readers should not be victims to this over-simplification. Overall judgment should be holistic, taking into account the roles of chance, bias, and confounding, which should all be considered and described within a cohort study. If such information is lacking, interpretation will be limited and it may be difficult to determine whether conclusions were properly drawn. Introduction of the STROBE guidelines [10] will hopefully lead to more rigor in reporting standards for observational studies

as has been the aim of the CONSORT guidelines for trials.

Case study: Critical evaluation of cohort study "Long-term use of aspirin and nonsteroidal anti-inflammatory drugs and risk of colorectal cancer" [8]

Information on 82,911 women from the Nurses' Health Study [5] was used to assess prolonged aspirin use at various doses on the risk of incident colorectal cancer diagnoses. A total of 121,701 female nurses aged between 30 and 55 years from across the United States of America were initially registered using a postal questionnaire in 1976. Questions about aspirin use were introduced in 1980. The authors were able to assess dose, duration, and timing of aspirin use in relation to cancer outcomes up to the year 2000, making it one of the largest and longest studies of aspirin use. They found a reduced risk of colorectal cancer associated with regular aspirin use compared with nonregular use (rate ratio 0.77; 95 % confidence interval (CI) 0.67–0.88) over the whole study period; however, further analyses of doses and exposure duration showed that important risk reductions were seen only after 10 years of use, at doses of 14 tablets per week or greater. Non-aspirin NSAIDs also showed similar results. Adverse effects of aspirin (gastrointestinal bleeding) were additionally assessed, showing increased incidence with increasing aspirin doses. These were weighed against aspirin's potential protective effects in the authors' conclusion.

1 How has the study population been identified and selected?

As described above, the study participants were women selected from the Nurses' Health Study. Advantages of studying this population were that women were easily identified through memberships of State nursing boards. Female nurses are not in fact representative of the whole US population so it would be important to consider whether the findings of this study are generalizable to men and to other women in age groups not studied. There is no reason to believe, however, that the study's internal associations will be distorted.

Only 82,911 from the original 121,701 women (i.e. 68 % of the original cohort) were selected for the assessment of aspirin. The authors report that these were the women with adequate follow-up information and no reported history of cancer or other major bowel diseases. The authors' definition of adequate follow-up time and the proportion excluded for this reason in contrast to the proportion excluded because of history of cancer are not described. Without this information, it is not possible to determine whether exclusions may have introduced selection bias into the study.

- ***Exposed and unexposed cohorts should be assessed for similarities and differences other than the exposure of interest***

Because cohort participants were not specifically selected based on their exposures to aspirin, we can be reassured that there was no selection bias directly related to the exposure introduced at *initial* cohort selection. Aspirin exposure was categorized using information from the cohort's pre-collected questionnaire responses. While aspirin users may have had some important differences from non-aspirin users, any differences were not introduced by the initial selection methods. One reason the original cohort was set up to include only nurses was that they were considered to be more similar to one another than people in other professions or in a general population cohort.

- ***All study participants must be at risk of the outcome***

The authors report that any women with a reported history of cancer or other major bowel disease (we must presume that this is at the start of follow-up) were not included in analyses. While it is possible that this may have excluded some women unnecessarily, it is likely that this provided a naïve population of women at risk of colorectal cancer.

2 How was the exposure defined and measured?

Questionnaires were posted to women every 2 years from 1980 onwards (i.e. repeat cross-sectional data collection). Among the information items requested was aspirin and NSAID use, which was included in a list of several types and brands of vitamins and medications.

- ***Exposures should be clearly defined and measureable, with explicit information on exposure timing***

After obtaining questionnaire information on aspirin, exposure was defined in several ways to categorize dose, duration of use, and timing of use before colorectal cancer diagnoses. Based on the number of standard (325 mg) aspirin tablets women reported using weekly, they were first categorized as regular

aspirin users (≥2 standard tablets per week), nonregular users, or nonusers. Dose was characterized as 0.5–1.5, 2–5, 6–14 or >14 tablets per week. Years of aspirin use over the period were also calculated.

Categorizations were complex primarily because weekly aspirin use was (unfortunately) measured differently in sequential questionnaires. Women were initially asked in 1980 if they took "any of the following vitamins or medicines in most weeks" (which included aspirin and NSAIDs) and to record the number of tablets taken each week and the number of years of use. In subsequent years they were asked only to indicate a category of tablets taken per week as 1–3, 4–6, 7–14, or ≥15 (in 1982), the average number of days of aspirin use per month (in 1984), and further changes to the measurement of drugs were made in subsequent questionnaires. Although the overall drug categories seem well defined, it is important to recognize that these measurement changes could have introduced measurement error.

- ***Objective measures should be used where possible***

Women's self-reported aspirin and NSAID use was not an objective measure of exposure; however, this was likely the only practical way to measure this exposure over a long study follow-up period for a large number of participants. Accurate measures of drug exposure are notoriously difficult to obtain and purchase or prescription data are not often a better option.

3 What is the outcome and how was it ascertained?

The outcome was a new diagnosis of colorectal cancer during the study follow-up period. This was further subdivided into proximal colon, distal colon or rectal cancer, and cancers were staged.

- ***Outcomes should be clearly defined and measureable, with explicit information on whether they are first or recurrent events***

The outcome was clearly defined as a first event of colorectal cancer. It was ascertained biennially using the routine questionnaires sent to cohort members and was therefore measured as a self-reported diagnosis. Death information was obtained from the National Death Index or reported by next of kin. Diagnoses and deaths reportedly due to colorectal cancer were confirmed by a physician using medical and pathological reports, obtained with permission from participants or next of kin. Medical reports do not appear to have been consulted if women did not report cancer, however.

- ***Outcomes should be defined, ascertained, and measured in the same way for exposed and unexposed groups***

We can be assured that outcomes were defined and measured by researchers in the same way regardless of aspirin exposure categorization as this categorization was done after outcomes were obtained. The authors also report that the study physician who used medical and pathological reports to confirm the cancer site and stage, was blinded to the participants' exposure information.

One may consider whether there was any ascertainment bias due to the initial use of self-reported cancer outcomes from women. This should be considered in context of the methods and of the study findings. Firstly, women were not blinded to their exposures although they likely did not know the study hypothesis since little evidence relating to aspirin and cancer outcomes was available during the study period. Secondly, the findings indicate that women with higher aspirin use were at reduced risk of colorectal cancer which means that if a bias in reporting were present, aspirin users would have been less likely to report colorectal cancer, which seems unlikely. Considering that the study participants are nurses and that cancers are well diagnosed in the United States, it is unlikely that ascertainment bias was important in this study.

- ***Outcomes within a reasonable time after an etiologic exposure should be considered***

The authors do not discuss how the date of diagnosis of colorectal cancer was defined, which is a drawback with regard to assessing timing of the outcome. This may have been a date reported by the participants or from the medical notes, it may have been when participants first discovered their illness, at their first medical appointment for the illness, at the first confirmed diagnosis date, or the date of death for those dying from the illness.

Although consideration of an etiologic time window is not mentioned in their principle analyses and they include all outcomes over the study period, the authors do go on to assess duration of use by categorizing cumulative years of exposure, which attempts to address this.

- ***A secondary outcome of note***

This is one of the few studies of aspirin and cancer where adverse effects were also addressed. It is important to note, however, that all information on gastrointestinal bleeding was obtained by asking women

in 2004 to retrospectively recall and report whether they had any major episodes of gastrointestinal bleeding before the year 2000 and when these occurred. The increased risk of gastrointestinal bleed with aspirin use has external research evidence and is clinically recognized; it is therefore possible that recall bias may have affected adverse event findings in this study.

4 How has person-time been dealt with in this study?

Women were followed up from the month they returned the 1980 questionnaire up to the month they were diagnosed with colorectal cancer, death from other causes, or June 2000. All person-time women contributed was included in analyses.

5 What has been done about loss to follow-up?

The authors report that the follow-up of the overall cohort of the Nurses' Health Study (121,702 participants) exceeded 90 %. They did not specifically report the loss to follow-up for the women included in their current analysis (82,911 participants), nor whether loss to follow-up varied by aspirin use.

- ***Methods to minimize loss to follow-up should be described***

Although details are not discussed in the paper, the Nurses' Health Study used many ways to keep in contact with participant nurses, which is evident from their 90 % retained follow-up over 20 years.

- ***Impacts of loss to follow-up on study power should be considered***

The low loss to follow-up makes this less important to consider.

- ***Impacts of loss to follow-up on bias should be considered***

Although follow-up was quite complete, the authors should have reported whether there were differences in follow-up between aspirin users and nonusers and by levels of aspirin use in their analyzed population.

6 What has been done about time-varying exposures and their effect on the results?

As discussed in point 2, the authors measured aspirin exposure in several ways. To take into account changes over time in their analyses, they estimated an overall cumulative average intake of these drugs and the number of years of use using all questionnaire information up to the start of each 2-year follow-up interval.

7 Was information about potential confounders collected, and how?

A number of potential confounders including measures of age, ethnicity, smoking habit, exercise, and dietary intake were also self-reported biennially on questionnaires and thus were also time-varying.

- ***Potential confounders should be defined, ascertained, measured in the same way for exposed and unexposed groups***

As potential confounders were obtained in biennial questionnaires, we are assured that definitions, measurement, and ascertainment were consistent across the whole cohort.

- ***Confounding should ideally be assessed in study analyses***

The authors display the covariate distributions by aspirin use from the 1980 baseline questionnaire to characterize the cohort in their first data table. In their analyses they use Cox regression and adjust their models for several covariates, allowing these covariates to vary over time. For each 2-year interval, information was used from women's latest biennial questionnaire and hence they fully exploited their data on time-varying covariates.

- ***Confounding by indication should be considered for studies of drug effects***

Because aspirin was associated with a decrease in colorectal cancer risk confounding by indication is unlikely to be important in this study. If there were a medical condition for which people took a high dose of aspirin and this medical condition reduced the risk of colorectal cancer, this could result in confounding by indication.

8 How was the sample size determined, and was the study large enough to answer the question?

The authors do not present a sample size calculation. They do present the case numbers and the numbers of person-years for all aspirin exposure groups in their analyses, which give the reader information to interpret whether their study may have been underpowered. There were 962 cases of colorectal cancer over 1,592,017 person-years of follow-up and 196 of these cases occurred in women taking aspirin for over 10 years. This is one of the largest and longest follow-up studies of aspirin and NSAID use on the risk of colorectal cancer.

9 Were the data analyzed properly? (What statistics have been used and how do I interpret them)

- ***The measure of effect used should be clearly described***

The abstract describes only relative risks making it impossible to know whether rates, or person-time, were used in analyses. Fortunately, the methods assure

us that incident rates of cancer outcomes were calculated by dividing the number of new cases of colorectal cancer by the person-years contributed to each exposure category. Hazard ratios, which can broadly be interpreted as rate ratios, were then calculated to assess variation by aspirin exposure.

- ***All analyses should be justified in relation to the data used***

Cox proportional hazards regression was used to calculate hazard ratios and these analyses included time-varying exposures and time-varying covariates, which were appropriate for these time-to-event data. The multivariable results show that there was little confounding of the main effect estimates.

10 Were the conclusions properly drawn based on the results?

The authors' original objective was described rather broadly as "To examine the influence of aspirin and NSAIDs in prevention of colorectal cancer" [8]. Their initial conclusion was "Regular, long-term aspirin use reduces risk of colorectal cancer. Non-aspirin NSAIDs appear to have a similar effect" [8], with later elaboration to say that important effects are only seen after at least 10 years of use with maximal risk reduction at over 14 tablets per week. They caution that this dose is higher than that recommended for cardiovascular disease prevention and that risks of these doses should be weighed against adverse risks of gastrointestinal bleeding.

Several different measures of aspirin exposure representing dose, duration and timing of the exposure relating to the outcome were assessed in this paper and all showed coherent evidence that aspirin was associated with a reduced occurrence of colorectal cancer. Analytical methods took into account changes over time, both in the exposure and in potential confounders. Residual confounding, however, is possible in every study and women taking 14 aspirin tablets a week may have several important differences from less regular users. The study power was reasonable, although was likely limited for some of the specific estimates calculated. Nevertheless, this remains one of the largest and longest studies of colorectal cancer and aspirin, demonstrating how challenging it is to study disease epidemiology.

There is of course potential for misclassification of the exposures as these were self-reported by women, but because of the prospective data collection there is no reason to think that this may have directly introduced bias into the study. Follow-up was reasonably complete, and although further information could have been provided, it seems unlikely that follow-up bias would have had an important impact on the results. The authors have thus provided a complete and balanced conclusion from their data and analyses.

Of clinical consideration, however, is whether these results should be applied in current practice and if so, how and in what population. If we were to apply this to the whole population to prevent new cases of colorectal cancer, given a median diagnosis age of 60–70 years, people now in their 40s or ideally younger should start regular aspirin use at high doses. The cohort studied included only female nurses born between 1935–1970 in the United States, and whilst the findings may apply to this group, diet and lifestyle changed considerably over the period of this study. Colorectal cancer incidence also changed considerably so whether the findings reflect true chemoprevention or are related to indication or competing risks is difficult to answer, even given the clinical and epidemiologic rigor that should be awarded to this study.

Summary: 10-point checklist

1 **How has the study population been identified and selected?**
- *Exposed and unexposed cohorts should be assessed for similarities and differences other than the exposure of interest*
- *All study participants must be at risk of the outcome.*

2 **How was the exposure defined and measured?**
- *Exposures should be clearly defined and measureable, with explicit information on exposure timing*
- *Objective measures should be used where possible.*

3 **What is the outcome and how was it ascertained?**
- *Outcomes should be clearly defined and measureable, with explicit information on whether they are first or recurrent events*
- *Outcomes should be defined, ascertained, and measured in the same way for exposed and unexposed groups*
- *Outcomes within a reasonable time after an etiologic exposure should be considered.*

4 **How has person-time been dealt with in this study?**
- *Study entry and exit times must be clearly defined for all participants*

- *Analytical methods incorporating person-time must be used when there is important variation in the amount of time participants contribute.*

5 **What has been done about loss to follow-up?**

- *Methods to minimize loss to follow-up should be described*
- *Impacts of loss to follow-up on study power should be considered*
- *Impacts of loss to follow-up on bias should be considered.*

6 **What has been done about time-varying exposures and their effect on the results?**

7 **Was information about potential confounders collected, and how?**

- *Potential confounders should be defined, ascertained, measured in the same way for exposed and unexposed groups*
- *Confounding should ideally be assessed in study analyses*
- *Confounding by indication should be considered for studies of drug effects.*

8 **How was the sample size determined, and was the study large enough to answer the question?**

9 **Were the data analyzed properly? (What statistics have been used and how do I interpret them?)**

- *The measure of effect used should be clearly described*
- *All analyses should be justified in relation to the data used.*

10 **Were the conclusions properly drawn based on the results?**

Multiple choice questions

1 Which of the following statements is true about the cohort study design?

A The design is less useful for investigating rare exposures than for investigating rare outcomes.

B Historical cohort studies do not allow temporal relationships between exposures and outcomes to be established.

C Participants are followed over a period of time to measure incident outcomes.

D Historical cohort studies are particularly prone to recall-bias.

E The design is useful for studying new events of disease but cannot be used for studying recurrent events.

2 In a cohort study loss to follow-up

A affects the study power and can result in bias if it is different between exposed an unexposed cohorts.

B results in bias only when at least 20 % of the cohort has been lost to follow-up and there is a 10 % difference in losses between exposed and unexposed cohorts.

C is inevitable and always results in follow-up bias to a certain degree.

D is a result of poor initial recruitment of too few exposed study subjects.

E can occur as a result of a design flaw where the follow-up time was too short to allow for the length of the average disease induction and latency periods.

3 Which of the following is least important to consider in a cohort study?

A Ascertainment bias of the outcome

B Blinding participants to their outcome status

C Confounding

D Blinding researchers to the exposure status

E Selection bias

4 In a cohort study, it is possible to measure

A Disease prevalence

B Odds of disease

C Incidence rates of first occurrence of disease

D Incidence rates of disease

E All of the above

5 Which of the following statements is true about prospectively collected cohort data?

A Loss to follow-up in a cohort study can be a result of selection bias or ascertainment bias.

B Loss to follow-up in a cohort study never results in any study biases.

C It is better, in terms of statistical power, to have 1500 people each followed for 4 years than 2000 people each followed for 3 years.

D 1500 people each followed for 4 years will provide the same number of person-years as 2000 people each followed for 3 years.

E Prospective data must be collected at regular intervals over the study period.

References

1 Porta M (ed.) (2008) *A Dictionary of Epidemiology*, 5th edn, Oxford University Press, Oxford.

2 Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* 2006;35(1):34–41.

3 Dawber TR, Moore FE, Mann GV. Coronary heart disease in the Framingham Study. *Am J Public Health Nation's Health* 1957;47(4 Pt 2):4–24.
4 Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. Original publication: *BMJ* 1954;1(4877):1451–5; reprinted *BMJ* 2004;328(7455):1529–33.
5 Belanger CF, Hennekens CH, Rosner B, Speizer FE. The Nurses' Health Study. *Am J Nurs* 1978;78(6):1039–40.
6 EPIC – European Prospective Investigation into Cancer and Nutrition [Internet]. Available from: http://epic.iarc.fr/ (accessed June 10, 2011).
7 Rothman K. (2002) *Epidemiology: An Introduction*, Oxford University Press, New York.
8 Chan AT, Giovannucci EL, Meyerhardt JA, et al. Long-term use of aspirin and nonsteroidal anti-inflammatory drugs and risk of colorectal cancer. *JAMA* 2005;294(8):914–23.
9 Neal KR. Excess mortality rates in a cohort of patients infected with the hepatitis C virus: a prospective study. *Gut* 2007;56(8):1098–104.
10 von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370(9596):1453–7.

Answers to multiple choice questions

1. C
2. A
3. B
4. E
5. D

3 How to read a case-control study

Joe West, Laila J. Tata, & Timothy R. Card
Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK

Key points

- The case-control study has become one of the most commonly used designs in clinical medicine but despite its apparent simplicity close attention to the methods used are needed to appraise such studies.
- It is an efficient design for studying rare diseases because study subjects are specifically selected based on whether or not they have an outcome of interest.
- The terms 'case' and 'control' are ubiquitous in clinical research and readers must recognize that studies are often incorrectly labeled as case-control studies, where subjects have not been selected on an outcome of interest.
- The method of selection of the control population in a case-control study is crucial to interpreting the reported results.
- Potential impacts of selection bias, ascertainment bias, recall bias, confounding, and chance should always be considered.

Brief introduction to case-control studies

Case-control studies are an epidemiological study design used often for the purpose of identifying risk factors or putative etiological exposures for rare diseases. They have a number of advantages relative to cohort studies primarily related to efficiency. Often they can be smaller, less expensive, quicker to conduct and easier to analyze, yet despite their apparent simplicity pitfalls exist when attempting to critically appraise them. The essential facet of a case-control study is that cases with a disease or outcome of interest are identified from a population and then, from the same population, controls who are people without the disease are also selected. Then it is a matter of measuring the exposures of interest in both groups and comparing the frequency of occurrence of the exposures among cases with that among controls.

However, to the unwary reader, taking any result from a case-control study at face value is fraught with the possibility of mistakenly believing in the finding of an association as evidence of causality. The latter interpretation is a leap of faith beyond the ability of the study design, yet judging the accuracy of any reported association is not impossible. In this chapter we outline an approach to making that judgment by using a 10-point checklist and then apply this to critically appraise a published study [1].

Biases commonly seen in case-control studies

A bias commonly seen in case-control studies relates to the selection of cases and controls. Selection bias

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

will occur if selection of cases or controls is associated somehow with one or more of the exposures of interest. The aim when selecting the study population is for cases to be either all or a random sample of available cases, and for controls to represent a random sample of all those who do not have the disease, but would be cases if they did.

Information bias is always important to consider in a case-control study. This may arise if information on exposures is reported or ascertained in a systematically different way between cases and controls. When information on exposures is collected after cases have been identified (which is typical of this design), this takes the form of recall bias because participants must recall past exposures. Collecting information on past exposures without bias being introduced is challenging but can be achieved.

10-Point checklist of important issues when reading a report of a case-control study

Chapter 2 on cohort studies covers many aspects that are pertinent to the understanding and appraisal of case-control studies. Here, we focus on the issues we consider most crucial to forming an opinion of the methodological quality of a case-control study and the accuracy of its findings. In addition to our checklist below there are many places to read about the design and reporting of case-control studies and their evaluation. We recommend both the introductory chapter to Breslow and Day's IARC publication on the subject of case-control studies written by Philip Cole, and the STROBE guidelines for reporting observational studies with the accompanying explanatory article [2,3].

1 How were cases selected for the study?

- ***Cases should be clearly and consistently defined and identified***

It is crucial that the paper reports in detail how the cases were defined and identified as this gives the reader the opportunity to judge how representative the cases included in the study are of all cases of the disease in the target population. Cases can be identified in a number of different ways, for example clinical diagnoses, self-report, laboratory tests, death certificate, or from registries such as those of cancer. If possible, it is best if an objective measure of disease is used, for example consistent clinical criteria rather than self-report – that is, abnormal small bowel biopsy for celiac disease as opposed to self-report of having the disease.

Most case-control studies of chronic disease study new cases of disease, that is, incident cases. An alternative approach is to study prevalent cases; however, these may not be a good representation of all cases from the population as for a disease with high mortality, long-term survivors will be overrepresented among prevalent cases and they may be different from those who died earlier in some way that is related to the outcome of interest. Furthermore, prevalent cases may have altered their behavior since diagnosis rendering their exposures unrepresentative of those prior to disease onset.

A separate problem in case selection relates to generalizability. Cases selected from secondary or tertiary care, for example, are likely to be at the severe end of a disease spectrum while cases identified from primary care or from the general population are more likely to represent the full range of disease severity. Findings from hospital populations may only therefore be generalizable to similar populations.

- ***Any exclusion criteria should be described with sound reasoning that does not introduce bias***

Once the cases have been selected many studies end up excluding people from the analyses carried out. Such exclusions inevitably change the population studied and should be justified. Careful scrutiny of any exclusions is recommended as, following exclusions, a biased case series may remain.

2 From which population were the controls selected?

Controls for any case-control study must be drawn from the same population that the cases were obtained from. By this we mean that if a control were to develop the disease of interest, they should be able to become a case by being identified and included in the study.

- ***How were the controls defined, identified, and included?***

A control should not have the disease of interest being studied at the time they are identified and included in the study. Ideally this should be verified in a similar way to case verification with a result of a negative test (e.g. normal small bowel biopsy to rule out celiac disease), but practically, for clinical studies, controls

cannot for ethical and expense reasons undergo the same scrutiny of verification. Not presenting clinically is therefore usually accepted as the definition of not having the disease. However, by scrutinizing the definition, identification, and inclusion criteria of controls readers may be able to identify evidence of selection bias. A classic example of such selection bias is from the use of hospital controls. If when studying upper gastrointestinal bleeding controls had been recruited exclusively from a cardiology clinic, then they would be far more likely than the general population to be taking aspirin, and hence any risk of taking this drug might be minimized.

- ***Were any of the potential controls excluded and if so why?***

Exclusion criteria for controls should be the same as for cases. If this is not done then a study may find a spurious association between being a case and the exposure.

3 How do the cases and controls compare to one another?

Before assessing the exposure, comparisons between cases and controls should be made for important baseline characteristics such as age and sex. Often cases and controls have very different or very similar characteristics and these situations result in different challenges in the interpretation of the study.

- ***If cases and controls are different consider how this will affect the study?***

If controls are very different to cases for a certain characteristic, for example if they are on average 10 years older, the risk of exposure may also be different due to this characteristic. If an association is then found between the exposure and outcome of interest, it is important to recognize that this may simply have been due to the difference in age. This may occur for several characteristics so it is important that authors describe characteristics that could possibly be associated with the exposure or the disease. In general these problems can be handled by statistical adjustment; however, this requires that there is enough data. In one specific circumstance this will not be available. If cases and controls are differently distributed with respect to some nominal potential confounder (such as, for example, race), to such an extent that there are categories which contain many cases and no or almost no controls (or vice versa), then statistical techniques will be unable to correct for confounding.

- ***If cases and controls were similar how was this achieved? Consider how this will affect the analysis and interpretation of the study***

If cases and controls are very similar, for example in terms of their age, this may have been achieved through a design feature called "matching". This technique is used in an attempt to ensure that cases and controls are similar for some important variables (such as age and sex) that could potentially distort an association between an exposure and outcome (i.e. potential confounders). It is commonly assumed that the reason for matching is that if controls are matched for factors which are considered potential confounders of the relationship under study, for example age, then this should minimize the effect of the confounding. However, in general, confounding can be controlled for in analysis. The main advantage of matching is that it can improve efficiency. If one compares a group of mainly young adult cases with controls of all ages, then more controls are required (to provide the additional data to permit correction for confounding) than if only controls of similar age are used. Matching can be carried out on an individual basis whereby one or more controls are selected that closely matches the age of a single case. Alternatively, frequency matching is performed whereby a number of controls are randomly selected from a pool of potential controls of the same age stratum. Each method requires a different approach to analysis [4].

It is important that the report describes in some detail the variables used to match and the rationale behind their choices, the process by which this was done, and whether the matching has resulted in a representative control group. The main problem with matching is that if too many criteria are required to find a "matched" control, this can lead to the exclusion of a large number of potential controls possibly in a biased manner. For example, if age, sex, socioeconomic status, and smoking status are all considered necessary to be matched for in a study but the data on the latter two variables is not complete for many of the potential controls, then selection bias could be introduced in the control selection.

Finally, we should consider the possibility that one of the matching variables is itself too closely linked to exposure in such a way that variability in exposure is lost (overmatching). If in a study of upper

gastrointestinal bleeding we matched on the presence of ischemic heart disease then each case with ischemic heart disease would have controls with ischemic heart disease. As aspirin use would be very common in subjects with ischemic heart disease (and this is a common comorbidity in people of the age group who have upper gastrointestinal bleeding), we would be artificially constraining the exposure to be the same between cases and controls, and so hiding any effect of aspirin.

4 What is the exposure (or exposures) and how was it measured?

Case-control studies do not have to be restricted to studying only one exposure. Many exposures can be studied but it is crucial that all are clearly defined and measurable.

- ***Exposures should be defined, ascertained, and measured in the same way for cases and controls***

If there are systematic differences between cases and controls in terms of exposure measurement then this could introduce information bias into the study.

- ***Explicit information should be provided on when exposures were measured and when they occurred***

For an exposure to cause an outcome it must precede it; that is, an exposure must act prior to disease onset if it is causal. Hence when studying risk factors, we are often interested in exposures that occurred sometime in the past. The timing of some exposure onsets, for example obesity or smoking, however, may be difficult to measure in retrospect, and so for the reader it is important to consider how well the measured exposure represents the exposure present in the period during which it is thought to act.

5 What has been done to minimize bias?

Aside from the biases already discussed in relation to the selection of cases and controls and measurement of exposure there are other biases that can affect a case-control study. Particularly important in this study design are recall bias and ascertainment bias. It is intuitive to think that a person having just been diagnosed with a condition will scrutinize their past exposures more than a control individual who is currently healthy, and therefore report them differently when asked to recall them. Equally, in the process of being diagnosed, a case recording of potential exposures in the medical record may be more complete than for healthy controls (i.e. better ascertained). Minimizing these biases requires careful attention to detail in the method of exposure measurement to ensure that any likelihood of over- or underestimating the prevalence of an exposure in cases is decreased. One way of doing this is, if practically possible, blinding the researcher to the status of the study subjects such that they do not know if they are collecting exposure information on a case or on a control.

6 What has been done to minimize the problem of reverse causality?

Reverse causality occurs when an association is found between an exposure and an outcome in a case-control study not because the exposure caused the outcome, but because the outcome caused the exposure. In cohort studies where exposures are measured before outcomes this should not happen, but since in a case-control study information may be collected at or after the time of diagnosis of disease, it is clearly possible. Strategies to avoid this problem could include a retrospective review of records from before the occurrence of disease to assess exposure.

7 Was information about potential confounders collected, and how?

Information on potential confounders should be clearly defined and comprehensively ascertained along similar principles to collection of exposure data.

- ***Potential confounders should be defined, ascertained, measured in the same way for cases and controls***

Procedures to identify confounders should be identical for all study participants to avoid any biases between cases and controls. Objective measures should be used if possible. As with exposures, confounders may change over time, which should be considered when interpreting the study.

- ***Confounding should ideally be assessed in study analyses***

Numerical data on the distribution of potential confounders should be presented and assessed in the cases and controls, and incorporated into multivariate analyses with appropriate reasoning. The effect on the estimates of association of any confounding factors should be reported and explained.

- ***Confounding by indication should be considered for studies of drug effects***

A particular type of confounding, *confounding by indication*, should also be considered when a drug of interest could be used to treat early symptoms of

the disease that defines case status. This may be mistaken for an association between the drug and the outcome.

8 How was the sample size determined and was the study large enough to answer the question?

If a study is underpowered, then if no association is found there is a risk of a type 2 error, that is, finding an absence of association falsely. The number of cases and controls in the study and the number and frequency of exposures all contribute to the study power. While sample size calculations are useful, each one is only relevant to a specific exposure-outcome combination. Hence in a study with several exposures or with subgroup analyses, the power of these comparisons cannot be adequately described by a single calculation for a primary exposure-outcome relationship. In case-control studies there is an added complication in matched designs where the number of discordant pairs is of critical importance as only those matched pairs where there is a difference in exposure will contribute to a multivariate analysis. As readers we therefore need to gain indications of power from more than one source. Beyond the power calculation we should look at overall numbers of cases, controls, and exposure frequency, and in matched designs also at numbers of discordant pairs. We should also consider the size of confidence intervals around effect estimates. Where these are wide, power is low.

9 Were the data analyzed properly? (What statistics have been used and how do I interpret them?)

The most common measure of effect used in a case-control study is now an odds ratio due partly to its useful mathematical properties and relatively easy interpretation. Numerous other measures of effect have been reported in case-control studies over the years, but we will not consider them further here [4].

- ***All analyses should be appropriate to the data used***

A common error is not using the correct matched analysis technique despite a matched design. Failure to do so can mean that variation due to matching variables can, despite the matched design, distort the effect estimates. In an individually matched study therefore we should check that appropriate statistical techniques have been used. These may include paired t-tests for means, McNemar's test for proportions and conditional logistic regression for multivariate modeling. Frequency matching is analyzed more appropriately as if matching had not taken place.

- ***The handling of missing data should be described appropriately***

Exclusion of those with missing data from studies can cause bias. Such exclusions happen not only at the design stage but also in analysis if subjects with missing data are excluded, and so the handling of missing data should be described in the statistical methods.

- ***The reporting of results should include numbers of subjects, unadjusted and if relevant adjusted estimates of effect***

As outlined above the power of different parts of the study is not constant and the presence of apparent confounding can have more than one explanation. For a reader to interpret results therefore they must consider for themselves the numbers of subjects involved, and those exposed both to the primary exposure and confounding factors, and also the effect of correcting for these confounders. To fully interpret power we also need to consider the precision of effect estimates and so some measure of this such as a 95 % confidence interval is required.

10 Were the conclusions properly drawn based on the results?

Having read a study carefully the reader should be able to come to some view of the strengths and weaknesses of it and what the results mean. This should permit a critical consideration of the conclusions drawn by the authors.

Case study: Critical evaluation of a case-control study "Effect of aspirin and NSAIDs on risk and survival from colorectal cancer"

The paper by Din et al. [1] describes a population-based case-control study of colorectal cancer carried out in Scotland, UK to assess the risk of developing colorectal cancer if a person had previously taken aspirin or nonsteroidal anti-inflammatory drugs (NSAIDs). They reported that:

- *the lowest daily dose of aspirin, 75 mg, reduces the incidence of colorectal cancer;*
- *the reduction in colorectal cancer incidence is evident after only 5 years' use;*

• *the protective effect is apparent in the general population and not just in high-risk groups.*

1 How were cases selected for the study?

• ***Cases should be clearly and consistently defined and identified***

Cases were defined as all incident cases of colorectal cancer between the ages of 16 and 79 presenting to surgical units in Scotland (it is not clear during what time period). The report states that the study only recruited approximately 45 % of all incident cases that arose in Scotland during their study period so it is possible that some selection bias occurred if those presenting were in some way more or less likely to report taking aspirin or NSAIDs. Of those identified only 52 % agreed to participate, and of those only 82 % completed questionnaires sufficiently to permit analysis.

It may be that these findings are therefore only generalizable to those who were recruited and included in the study, namely those presenting to surgical units. We do not know from this report the differences between the study population and the entire set of incident colorectal cancers that occurred particularly by age, sex, and stage of disease. This could have been provided as presumably it would have been available from the relevant cancer registry.

• ***Any exclusion criteria should be described with sound reasoning that does not introduce bias***

Exclusion criteria were death prior to ascertainment, inability to consent or complete questionnaires. This suggests that those with the most severe disease, that is, Dukes D or metastatic disease, will have been excluded from the study. The rationale is sensible (as recruiting dead people in a questionnaire-based study is impossible, yet medical records could have been sought to ascertain exposures). Furthermore, only participants who fully completed the questionnaire were included for analysis.

Bias and restricted generalizability could have been introduced by these decisions but there is no information presented in the paper to confirm or refute this.

2 From which population were the controls selected?

Controls were identified from a population-based register. Details of this are not provided in the report.

• ***How were the controls defined, identified, and included?***

Controls were selected during the same time period as cases and were randomly drawn from a population-based register. Controls were matched to cases on age (+/− 1 year), sex, and residential area. Only 39 % of identified controls agreed to participate and of those 97 % completed questionnaires sufficiently to permit analysis.

• ***Were any of the potential controls excluded and if so why?***

There were no reported exclusion criteria for controls; however, only participants who fully completed the questionnaire were included for analysis. It is notable that while 97 % of controls met this criterion only 82 % of cases did. The 61 % of controls who effectively self-excluded themselves, however, do provide an opportunity for introducing bias into the study. If the participating control group were "more healthy" than the general population and therefore less likely to take aspirin or NSAIDs, this would have led to an underestimate of the effect of the drug.

3 How do the cases and controls compare to one another?

Cases and controls had some important differences and similarities. Despite matching on age the controls were older (statistically significantly so) presumably due to the differential participation. Otherwise both sex and deprivation (linked strongly to residential area) were very similar. In the nonmatched variables the controls were more physically active, smoked slightly less, consumed less energy, and drank similar amounts of alcohol.

• ***If cases and controls are different consider how this will affect the study?***

Most of the differences in this study were of a small magnitude although the difference in age could have affected the study. Older people are more likely to have been prescribed or take aspirin or NSAIDs and so this could have led to an overestimate of the effect of the drugs.

• ***If cases and controls were similar how was this achieved? Consider how this will affect the analysis and interpretation of the study***

The similarities between cases and controls in terms of sex, residential area (deprivation score) in this study are due to individual matching and this can be seen by the distribution of these characteristics in the case and control populations described in Table 1.

4 What is the exposure (or exposures) and how was it measured?

The main exposure of interest was aspirin and NSAID use. Participants completed a lifestyle and cancer

information questionnaire reporting their exposures 1 year prior to diagnosis (for cases) or recruitment (for controls). Intake of aspirin, other NSAIDs and analgesics were reported. For those reporting regular drug use, the date of drug commencement, number of months, and number of days per week ingested was recorded.

5 What has been done to minimize bias?

Little was done to assess or minimize bias in the exposure measurement, for example no attempt was made to verify reported drug use with medical prescriptions for these drugs. Given that the majority of participants were over 55 years of age it would be expected that a proportion of the reported drug use will have been medically prescribed.

6 What has been done to minimize the problem of reverse causality?

The report states that exposures were reported 1 year prior to either diagnosis or recruitment which should have avoided the problem of reverse causality; however this relied on good recall of exposures.

7 Was information about potential confounders collected, and how?

- ***Potential confounders should be defined, ascertained, measured in the same way for cases and controls***

Information on confounders appeared to be measured in the same way for cases and controls.

- ***Confounding should ideally be assessed in study analyses***

Numerical data on the distribution of potential confounders were presented and assessed in the cases and controls and were incorporated into multivariate analyses. There were changes in the effect estimates following adjustment but these were not explained.

- ***Confounding by indication should be considered for studies of drug effects***

It is possible that greater than 1 year prior to diagnosis a change in the prescription of analgesics occurred in the cases due to symptoms related to the disease (i.e. colorectal cancer). If so, that may have led to a reduction in the use of aspirin or NSAIDs and greater use of other analgesics. It is difficult to predict whether this did happen or not, and if so, what the effect this would have had on the study findings. This information was reportedly collected but is not presented for assessment in the results of the study.

8 How was the sample size determined and was the study large enough to answer the question?

There was no formal approach to determining sample size presented in the report and multiple comparisons have been made in the analysis. Aside from the ever versus never comparisons the confidence intervals around the point estimates are wide indicating that for these analyses power was somewhat low.

9 Were the data analyzed properly? (What statistics have been used and how do I interpret them?)

- ***All analyses should be appropriate to the data used***

In this study an incorrect analysis has been carried out with respect to the matched design. As the cases and controls were individually matched the authors should have used McNemar's test to assess the associations between exposures and case or control status. In the multivariate or adjusted analysis a conditional logistic regression model was not, but should have been, used. It is difficult to speculate on the impact of this choice of analysis as in the model used all of the matching factors have been inappropriately included. In addition, despite matching there was an important age difference between cases and controls.

- ***The handling of missing data should be described and appropriate***

It appears that the analysis has only included those participants with complete questionnaire information (as described in points 1 and 2 above). We do not know therefore the effect of excluding individuals with missing data from the analysis.

- ***The reporting of results should include numbers of subjects, unadjusted and if relevant adjusted estimates of effect***

The report does include this information.

10 Were the conclusions properly drawn based on the results?

The report concludes that aspirin confers a protective effect against the development of colorectal cancer with low dose aspirin after only 5 years of use in the general population. While this reflects the results presented in the report, there are enough concerns regarding the population studied, exposure measurement, and analysis carried out that the reader should be cautious in accepting the veracity of the authors' conclusion. Particular concerns are the difference in participation of the identified cases and controls (thus possible selection bias), the difference in age of the cases and controls, the inappropriate statistical analysis, and the lack of a clear dose-response relationship with duration of exposure. However, the results are

broadly in concordance with the other observational literature on this subject.

10-Point checklist of important issues when reading a report of a case-control study

1 **How were cases selected for the study?**
- *Cases should be clearly and consistently defined and identified*
- *Any exclusion criteria should be described with sound reasoning that does not introduce bias.*

2 **From which population were the controls selected?**
- *How were the controls defined, identified, and included?*
- *Were any of the potential controls excluded and if so why?*

3 **How do the cases and controls compare to one another?**
- *If cases and controls are different consider how this will affect the study*
- *If cases and controls were similar how was this achieved? Consider how this will affect the analysis and interpretation of the study.*

4 **What is the exposure (or exposures) and how was it measured?**
- *Exposures should be defined, ascertained, and measured in the same way for cases and controls*
- *Explicit information should be provided on when exposures were measured and when they occurred.*

5 **What has been done to minimize bias?**

6 **What has been done to minimize the problem of reverse causality?**

7 **Was information about potential confounders collected and how?**
- *Potential confounders should be defined, ascertained, measured in the same way for cases and controls*
- *Confounding should ideally be assessed in study analyses*
- *Confounding by indication should be considered for studies of drug effects.*

8 **How was the sample size determined and was the study large enough to answer the question?**

9 **Were the data analyzed properly? (What statistics have been used and how do I interpret them?)**
- *All analyses should be appropriate to the data used*
- *The handling of missing data should be described appropriately*
- *The reporting of results should include numbers of subjects, unadjusted and if relevant adjusted estimates of effect.*

10 **Were the conclusions properly drawn based on the results?**

Multiple choice questions

1 Which of the following statements is true about the case-control study design?

A The design is not prone to selection bias because cases are selected first.

B The design is useful for investigating population risk factors for rare outcomes.

C Exposures are always measured before study outcomes have occurred.

D Temporal relationships between exposure and the outcome can always be studied.

E The design is useful for studying incidence rates of newly occurring rare disease.

2 Which of the following is not important to consider in a case-control study?

A Recall bias

B Confounding

C Bias resulting from loss to follow-up

D Blinding researchers to the outcome status

E Selection bias

3 Subjects are selected for inclusion in a case-control study

A based on whether they are free of an exposure

B based on whether they do not have an intervention

C based on whether they have an intervention

D based on their exposure status.

E based on their outcome status

4 Which of the following statements is true about matched case-control studies?

A Matched case-control studies are an efficient study design.

B Controls should not have the outcome of interest.

C Controls can be matched to cases on variables that are potential confounders.

D Only one control can be matched to each case.

E The analysis should take account of the matched design.

References

1 Din FV, Theodoratou E, Farrington SM, et al. Effect of aspirin and NSAIDs on risk and survival from colorectal cancer. *Gut* 2010;59:1670–9.
2 Breslow NE, Day NE. (1980) *Statistical Methods in Cancer Research. Vol. I, The Analysis of Case-Control Studies*, International Agency for Research on Cancer, Lyon, France.
3 Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Int Med* 2007;147:W163–94.
4 Rothman KJ, Greenland S. (1998) *Modern Epidemiology*, Lippincott-Raven, Philadelphia, PA.

Answers to multiple choice questions

1. B
2. C
3. E
4. A, B, C, and E

4 How to read a randomized controlled clinical trial

Matthew J. Grainge
Division of Epidemiology and Public Health, Nottingham City Hospital, University of Nottingham, Nottingham, UK

Key points

- The principle behind an RCT is to eliminate confounding which is unavoidable in observational studies by ensuring that on average allocation groups are the same with respect to factors that could potentially influence the outcome of a study.
- Despite this advantage empirical evidence has shown that the absence of double blinding and failure to conceal allocation sequences in RCTs can both bias results away from the null by 30 % or more.
- Published trial reports also contain instances where investigators have misled through several mechanisms including inappropriate choices of comparison group and selective reporting of outcome measures.
- Greater emphasis on compulsory pre-registration of clinical trials and publication of trial protocols will inevitably lead to future improvements in the quality of trial conduct and reporting.

Brief introduction to randomized controlled trials

The randomized controlled trial (RCT) has become ubiquitously recognized as the gold-standard study design in clinical research. Evidence-based guidelines weight their recommendations heavily on the basis of data from RCTs at the expense of those that use alternative designs, whilst the well-established Cochrane collaboration relies almost solely on RCT data when commissioning reviews on the effectiveness of healthcare interventions. However, the quality of RCTs (both in terms of design, execution, and quality of reporting) vary considerably, and the question of whether a poorly designed and executed RCT can be considered more reliable than a high-quality observational study on the same topic is frequently debated with little consensus. Whilst it is undoubted that RCTs provide the highest level of evidence overall, when empirical evidence has shown that on average, the absence of double blinding and failure to conceal allocation sequences both bias treatment effects away from the null by 30 % or more [1–3], it becomes more difficult to argue the case that an RCT will

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

always provide higher quality evidence than its observational counterpart in every circumstance. In this chapter an approach to evaluating the quality of RCTs will be provided through use of a 10-point check list. This will then be used to appraise a research paper in an area of gastrointestinal cancer which has received much publicity over the past two decades.

Biases commonly seen in randomized controlled trials

The principle behind the RCT is simple, to eliminate confounding which is unavoidable in observational studies by ensuring that on average allocation groups are the same with respect to factors that could potentially influence the outcome of a study. Since the first RCT of streptomycin in the treatment of pulmonary tuberculosis reported in 1948 [4], the principle of randomization has been used effectively to provide the highest quality research evidence across all clinical specialties (the first recognized clinical trial in the discipline of Gastroenterology was published in 1955) [5]. However, this advantage can be compromised by errors in the design, conduct, data analysis, and reporting of a trial, prompting the need for greater critical awareness of RCT by those responsible for synthesizing evidence from RCTs and among those who put evidence they provide into practice [6]. We can broadly distinguish these methodological issues into those of internal and external validity. Internal validity relates to the extent to which bias is avoided in a clinical trial; this includes selection, performance and attrition bias as discussed later. External validity refers to whether results can be generalized to different groups, on the basis of factors such as age, gender, and severity of disease. Internal validity is a prerequisite for external validity but not vice versa [7].

The issue of human nature and fallibility creates the first set of problems when evaluating evidence from RCTs. Study investigators are able to manipulate the evidence (or cheat) by choosing an inappropriate comparison group, using a comparator at the wrong dose (for a drug trial), or carrying out the trial in a group of patients where the comparator is known to be ineffective or problematic (Hywel Williams, personal communication, 2011). Even when intentions are good, errors such as a poor choice of intervention period could also hamper a trial even before the first patient is recruited. These basic design factors alongside the more detailed items discussed below create a multitude of problems when it comes to making judgments on the "value" one should place on the results from an individual trial.

However, there is good news in that the quality of RCT conduct and reporting has improved in recent years for a couple of notable reasons. First, there has been a greater emphasis on preregistration of clinical trials and for trial protocols to be published, with some leading general medical journals no longer considering RCT reports unless these conditions are met [8,9]. Second has been the requirement by all major journals that results of RCTs should be reported according to the CONSORT guidelines which were initiated in the 1990s to improve the reporting of RCTs [10].

In the checklist below, we assume the RCT being assessed is a two-arm parallel group trial comparing an intervention of interest with either an active (e.g. an existing intervention) or nonactive (e.g. placebo in a drug trial) comparator. Many of the principles and issues outlined below, however, should readily apply to other types of RCT.

10-Point checklist for evaluating a clinical trial

1 How was the allocation sequence generated, and was this adequately concealed?

A successful randomization scheme is one which guards against selection bias. Use of a computer algorithm and random number tables are both acceptable methods of allocating patients to a treatment. As long as the assignment cannot be predicted in advance then the method is adequate. In practice a prespecified sequence of treatment allocations is generated in advance of the first patient being recruited into the trial. If this sequence becomes known to recruiting clinicians, then the decision to recruit a patient (or level of coercion they apply on participants to consent) could be influenced by the treatment they will receive. As a result patients with better prognosis could be directed to the experimental group and those with worse prognosis the placebo group, or vice versa [7].

A central telephone or internet-based randomization service is often used in multicentre trials or those which are carried out in collaboration with specialist clinical trials units to enable concealment of treatment

allocations. Even smaller trials or those where central randomization is not possible, can ensure allocation concealment by use of sealed, opaque envelopes, or by having the allocation sequence generated by a pharmacy which then administers the treatments in bottles or packets that are identical in weight and appearance. The important thing is that unlike double-blinding, concealment of allocation is always possible [11].

2 Were participants and study personnel blinded?

Blinding (or masking) is easiest to implement in placebo-controlled drug trials where placebo tablets can be manufactured to be identical in appearance to the active treatment. An open trial is where there is no blinding. Single blinding is when either the patient or outcome assessor knows which treatment is being received. Double blinding is when both are unaware. If the treating clinician(s) is also unaware of the treatment and is distinct from the person(s) who assesses outcome, this is sometimes termed triple blinding, although the terminology used here is not consistent [12]. Even when a trial report states that the study was double blind this is not a guarantee that blinding has been maintained. Where treatments have common and well-recognized side effects, blinding could be removed for a considerable proportion of participants. If blinding was not successfully maintained or where blinding is not possible such as in trials of surgical or lifestyle interventions, results are open to both performance and detection bias [7] (see point 7) and the implications of this should be discussed.

3 How many participants were lost to follow-up?

Failure to obtain follow-up data on randomized participants (attrition bias) is a concern for RCTs. Missing outcome data for more than 20 % of randomized patients is considered to be high enough to cause a considerable risk of bias [13]. Drop-out in a trial is likely to relate both to the intervention being received and factors related to outcome and as such the benefits of randomization in alleviating confounded treatment group comparisons will be eradicated [14,15]. Even if overall drop-out rates do not differ between intervention arms, reasons for withdrawal could still differ.

It is important therefore that the number of patients lost to follow-up in each intervention arm is clearly reported, along with how they are dealt with in the statistical analysis. Under the principal of intention to treat, all patients should be analyzed in the groups to which they were assigned, and furthermore, all patients should contribute data to the analysis of the primary outcome. If patients with missing outcomes were excluded from the trial analysis, then an "available case" rather than full "intention to treat" analysis has been carried out, and the implications for attrition bias should be discussed [16].

4 Was the study population well defined?

Having a clearly defined study group and limiting the number of exclusion criteria employed will positively impact on the external validity of the trial. In pharmaceutical drug trials, inclusion and exclusion criteria are carefully instigated to target patients with a particular condition who would benefit most from the proposed intervention and be most likely to comply with study treatment and medication. The nature of the exclusion criteria will be specific to the circumstances of an individual trial. Exclusion criteria should be employed to ensure the trial does not include participants who (i) do not have the disease under investigation using clearly defined criteria (or who have the study endpoint at baseline in the case of primary prevention trials), (ii) have need to take either of the study interventions, (iii) have contraindications to either of the interventions, (iv) are unlikely to comply with either of the interventions. Extensive exclusion criteria which do not fit with the above should be viewed with some skepticism.

5 Was the intervention(s) administered correctly and appropriately?

The trial report should confirm whether intervention(s) were fully delivered to participants as described in the protocol. In the absence of blinding, there is the potential for patients in each group to be treated differently aside from the randomized intervention (performance bias) [7]. For example, if patients in the active intervention arm received more GP or healthcare visits, or a greater number of telephone contacts from members of the research team, or if measures introduced to improve patient compliance were not identical between the study arms.

6 Were intervention groups comparable at baseline?

Given that the point of randomization is to remove any systematic imbalances, we would hope the answer is yes (conditional on the important assumption that there were no problems with the generation and concealment of the allocation sequence). Therefore, this is arguably the one item on this checklist on which critical appraisers lavish too much attention. However, chance imbalances can occur. These should not be assessed by statistical testing to compare

intervention arms with regards to baseline factors, but instead statistical adjustment should be made for variables defined a priori as important prognostic factors and this is both acceptable and encouraged. As well as removing any chance imbalances that exist for these variables, such adjustment has been shown to increase the precision of treatment effects, and this holds equally for binary, continuous, and survival outcomes [17,18].

7 Were outcome measures suitable and valid?

Trial outcomes measures should be clinically meaningful, validated, and easily measureable. In practice, the first of these is the most difficult to achieve due to the abundance of surrogate outcome measures that are often available to trialists in contrast to the challenges of identifying objective evidence of new occurrences of disease. The effectiveness of an osteoporosis drug, for instance, can be established with smaller numbers of patients followed up for shorter periods of time if bone mineral density, rather than the more clinically important, but far rarer, outcome of hip fracture is used. The utility of a surrogate endpoint is considerably enhanced if it can be established that it provides a plausible and consistent dose–response link with the development or progression of disease [19].

The appropriate use of objective versus subjective outcome measures also needs consideration. Objective outcome measures are naturally encouraged as they are less prone to some forms of bias (specifically detection bias). However, this needs to be balanced against the need for outcomes relating to patient satisfaction and quality of life, which by their very nature are more subjective. Finally, were the primary outcome(s) identical to those in the study protocol? It is perhaps not too cynical to believe that the literature contains trials where the "original" primary outcome has been given reduced prominence in a trial report or even omitted from this completely because it failed to reach statistical significance. Such a practice is at best misleading and at worse cheating.

8 Was the sample size adequate and were the data analyzed properly? (What statistics have been used and how do I interpret them?)

All RCT reports should contain a sample size calculation and these should contain the following four components: alpha, beta (or power), control group information (estimated event rate or standard deviation), and the minimum treatment difference deemed to be clinically important. Even in trials published in high impact general medical journals, this key information is frequently missing [20]. Furthermore, discrepancies in sample size estimates and assumptions often exist between the trial report and the original protocol [21]. Sample size calculations are usually based solely on a single primary outcome. If the trial investigators have placed equal emphasis on multiple outcome measures then the sample size calculation should reflect this (e.g. using alpha = 0.01 to account for multiple significance testing).

Careful scrutiny is required of how results are presented in the trial report as this is where investigators have more opportunities to cheat. For example, claiming that a treatment doubles cure rate is not quite so impressive when in fact this represents an increase from 1 % to 2 % (Hywel Williams, personal communication, 2011). Absolute event (or cure) rates should be presented in addition to any measures of relative effect. Finally, make sure that all the great work carried out by a study team over a large number of years has not been let down by basic errors when it comes to data analysis. A few pitfalls to look for when it comes to analyzing data from trials are listed in Table 4.1.

9 CONSORT: have the results been reported following these guidelines?

Improved reporting of RCTs allows assessments of trial reporting and trial conduct to be more synonymous. The CONSORT (CONsolidated Standards of Reporting Trials) statement was first released in 1996 to ensure adequate reporting of RCT results [22] and with updated versions released in 2001 and 2010 [10,23]. All leading general medical journals and many specialty journals now request that all RCTs are reported according to the CONSORT criteria. CONSORT items are chosen to focus exclusively on factors that influence the risk of bias, and as such tally closely with the items discussed in the present chapter. Whilst a CONSORT checklist is an invaluable tool to have at hand when reading a trial report, the equally important role of intuition when assessing trial literature should not be overlooked.

10 Were the conclusions properly drawn based on the results?

By properly drawn we mean do they fully take account of any limitations inherent in the design and conduct of the trial as well as the results. The medical profession and media are naturally more excited about the findings of the trial rather than any key methodological deficiencies, so authors have a responsibility to

Table 4.1 Statistical pitfalls to be avoided in the analysis of RCTs

1. Were participants randomized on the basis of cluster rather than individually?	Where randomization takes place according to cluster (e.g. schools or general practices) but outcomes were obtained for individuals, this should be accounted for in the analysis. Standard errors of treatment comparisons need to be modified to reflect the correlation of responses within cluster.
2. Were model assumptions checked?	In particular for time-to-event outcomes, violation of the proportional hazards assumption should be tested for. A treatment may be effective shortly after intervention but less so long term.
3. Were interim analyses carried out?	These are sometimes are included in a trial protocol so that a trial can be stopped early if the intervention is clearly effective (or harmful). There is, however, a formal requirement to reduce alpha levels in this instance due to the repeated testing of the primary hypothesis.
4. Does the trial make repeated outcome assessments over time?	If so, were appropriate analyses to take account of repeated measures used, or was the analysis only carried out at a single point in time (wasteful of data) or are repeated cross-sectional analyses performed at every time point (multiple significance testing).
5. For a continuous outcome measures (e.g. blood pressure, body mass index) were baseline values adjusted for?	This is necessary to increase the precision of treatment effects and correct for any chance baseline imbalance (see item 6 in check list).

temper any conclusions as appropriate to ensure that the reader is not misled. A key concern here is ensuring that "positive results" do not receive too much attention at the expense of the overall finding (the primary outcome measure). Similar caution should be directed towards subgroup analyses, for example that no significant result was found overall, but the treatment was effective for men aged under 30 born under the star sign of Leo. Subgroup analyses can present all manner of problems with respect to interpretation, due to a toxic mix of type I and type II error. They will cause the introduction of multiple comparisons (thus the need to adjust the alpha level) and almost inevitably pre-study sample size calculations will be based on the entire trial group, leaving analyses carried out on subsets of trial participants woefully underpowered. Subgroup analyses should therefore always be pre-specified in the original trial protocol.

Case study: Critical evaluation example

The Women's Health Study (WHS) was a primary, placebo-controlled, 2 × 2 factorial trial evaluating the effects of low-dose aspirin (100 mg every other day) and vitamin E (600 IU every other day) in a cohort of 39,786 healthy US female healthcare professionals [24]. The primary outcome was the occurrence of any major cardiovascular event; however, total cancer was an important secondary endpoint, and of the specific cancers studied, the a priori evidence for a protective effect was arguably strongest for colorectal cancer (CRC). Only results from the aspirin component of the trial were reported in the paper (i.e. results were pooled over the two vitamin E arms). The average duration of follow-up was 10.1 years. Overall, the number of women developing CRC was very similar among women receiving aspirin (n = 133) and placebo (n = 136) (relative risk = 0.97; 95 % CI 0.77–1.24). The implications of this finding were important considering that observational studies had consistently found that people regularly taking aspirin had a 30–40 % lower risk of developing CRC, whilst secondary prevention RCTs have found that aspirin reduced the risk of recurrent adenomas by an almost similar magnitude [25]. The average follow-up duration appeared to be sufficiently long for aspirin to show an effect, therefore the dose used (50 mg day^{-1}) was touted as the most likely explanation for this negative finding [26].

1 How was the treatment allocation sequence generated, and was this adequately concealed?

There was an absence of details concerning how and by whom the random allocation sequence was

generated, neither were any steps taken to conceal the allocation sequence mentioned. This information was also absent from earlier papers describing methods of the trial [27–31]. The specification that randomization used blocks of size 16, merely provides reassurance that randomization took place and that future treatment assignments could not be foreseen (were the randomization sequence to have been successfully guarded). Given that the trial recruited participants from across 48 states [31], we can be reasonably sure that randomization sequences would have been held centrally, and therefore guarded from those recruiting. However, the fact remains that this information was not available from the trial report.

2 Were participants and study personnel blinded?

The WHS trial was double-blinded. One would assume that participant blinding was maintained if only because it was stated that compliance rates were broadly similar in the two groups. It is unlikely this would have been the case if large numbers of women receiving aspirin and/or placebo had correctly identified their treatment allocation. Even if participant unblinding had occurred, the objective nature of the outcome measure would make detection bias unlikely (providing those assessing pathology reports to confirm any cancers were adequately blinded).

3 How many participants were lost to follow-up?

It was stated that analyses were carried out on the basis of intention to treat and that morbidity follow-up was stated to be 97 % complete. However, greater elaboration on how this high level of morbidity follow-up was obtained was probably needed. Information on cancer diagnoses was obtained via questionnaires sent to participants, initially every 6 months and subsequently every 12 months. The likely level of response to a routine postal questionnaire following a cancer diagnosis was not discussed.

4 Was the study population well defined?

Yes, and the exclusion criteria were not too extensive with exclusions based solely on prior history of outcome (previous history of cardiovascular disease or cancer), need for trial medication (women currently taking aspirin or NSAIDs more than once a week), and unsuitability for randomization to aspirin (use of anticoagulants or corticosteroids, history of adverse effects to aspirin). Also, efforts were made to restrict trial participation to those likely to comply with intervention, by requiring the participants to enter a 3-month run-in phase. A total of 65,169 women entered this phase to identify long-term compliers to pill taking, of whom 25,293 were excluded due to noncompliance, unwillingness, or ineligibility. This approach was rewarded through high compliance among those randomized, with 76 % still taking the trial medication at 5 years and 67 % at 10 years.

5 Was the intervention(s) administered correctly and appropriately?

As trial participants were only required to take one tablet every other day this would not be a concern.

6 Were intervention groups comparable at baseline?

Baseline characteristics were shown to be similar between the aspirin and placebo groups; however, the authors committed the sin of carrying out significance tests for all variables at baseline (thus carrying out a pointless exercise of repeatedly testing a null hypothesis which is known to be true). Despite this, the terms adjusted for in the Cox regression analysis were sensible and seemed to be defined a priori; these were age (an obvious predictor of CRC risk) and vitamin E assignment.

7 Were outcome measures suitable and valid?

The endpoint of CRC was definitive, thus avoiding the surrogate outcome of colorectal adenomas as used by the aforementioned secondary prevention trials of aspirin. All cancers identified through follow-up questionnaires or death certificates were confirmed on the basis of pathology or cytology reports.

8 Was the sample size adequate and was the data analyzed properly? (What statistics have been used and how do I interpret them?)

No sample size calculation was reported in the present paper; however, in an earlier report from the WHS it was expressed that the trial was powered to detect a 25 % reduction in the primary endpoint of any major cardiovascular event [29]. Whilst this does not directly address our outcome of interest (colorectal cancer), in a trial report confidence intervals (CI) give a more accurate indication of the actual power achieved than the information from a pre-trial sample size calculation [32]. The CI for aspirin on CRC incidence (0.77–1.24) confirmed the study was large enough to rule out (with 95 % confidence) a greater than 24 % increase in the risk of CRC and a 23 % decrease in risk. Importantly, this latter figure is outside the range of likely risk reductions (30–40 %) estimated from observational research. The statistical analyses carried out were appropriate, with the Cox proportional hazards regression model

containing an interaction term to check that the effect of aspirin did not vary according to time since randomization.

9 CONSORT: have the results been reported following these guidelines?

Comparing information from the trial report with the CONSORT 2010 checklist yielded some discrepancies. As well as the missing details concerning randomization, there was also no mention of the settings or locations where the data were collected in the present report. Whilst the overall percentage of participants successfully followed up for morbidity and mortality was stated, this was not provided by treatment group, neither were exact compliance figures provided separately for aspirin and placebo participants. Other aspects of the CONSORT diagram were reported well, in particular a comprehensive description of the statistical methods used was provided.

10 Were the conclusions properly drawn based on the results?

The overall conclusion that aspirin does not reduce the risk of colorectal and other cancers was clearly supported by the data. Of all the results reported in the paper the only statistically significant result was a reduced risk of death from lung cancer among aspirin users (and this was not even one of the 21 outcome measures listed in the main results table). The authors were therefore quick to point out this was probably a chance finding. From our eyes, however, it can be argued that type I error was not a serious consideration because we were analyzing the paper from the point of view of a single outcome (colorectal cancer incidence).

Overall, the WHS provided good quality data from a cohort of sufficient size, with no obvious potential for serious bias. The quality of reporting, however, was less impressive at times, which is surprising from a journal that was compliant with the CONSORT statement at the time of publication [33]. By carrying out the study in a group of individuals one would expect to be highly motivated (middle-aged female health professionals), a high level of compliance was achieved. The downside of this would be whether these results could be generalized to other groups. There is no obvious biologic reason why this would not be the case, especially as the only previous primary prevention trial of aspirin and CRC, which also reported a null result, was carried out in male health professionals [34,35].

How to read a randomized controlled trial: 10-point checklist

1 **How was the allocation sequence generated and was this adequately concealed?**
- *Were patients truly randomized and how was this performed?*
- *This is done to prevent selection bias by the investigator. The sequence of treatment allocation should be concealed from the investigator and patient to prevent subversion of the randomization process by knowledge of the next assigned treatment.*

2 **Were participants and study personnel blinded?**
- *Double blinding is when both the participant and outcome assessor are unaware of treatment allocation.*
- *This is used to minimize performance bias (additional intervention provided preferentially to one group), and detection bias (preconceived views of the participant or investigator influence subsequent outcomes assessments).*

3 **How many participants were lost to follow-up?**
- *The benefits of randomization are lost when only participants who complete the study (provide valid outcome data) are included in the primary analysis, or if they were not analyzed in the groups they were originally assigned to.*

4 **Was the study population well defined?**
- *Were the population described in enough detail for you to judge whether they are generalizable to the patients you encounter in your own practice?*
- *Were exclusion criteria reasonable or too extensive?*

5 **Was the intervention(s) administered correctly and appropriately?**
- *If applicable, were interventions fully delivered to participants as specified in the protocol?*

6 **Were intervention groups comparable at baseline?**
- *In order to assess whether the randomization process has worked, the authors should report key participant characteristics in a table – have they and what does it show?*

7 **Were outcome measures suitable and valid?**
- *Were outcome measures easily measureable?*
- *Were surrogate endpoints relevant?*

8 **Was the sample size adequate and were the data analyzed properly? (What statistics have been used and how do I interpret them?)**

- *Did the authors perform a power calculation?*
- *Was choice of statistical analysis appropriate?*

9 **CONSORT: have the results been reported following these guidelines?**

- *See items in CONSORT checklist.*

10 **Were the conclusions properly drawn based on the results?**

- *What does the study result imply for clinical practice?*

Multiple choice questions

1 Rank the following types of study design according to the order they appear in the evidence hierarchy:

A Observational study (cohort or case-control)
B Randomized trial
C Expert opinion
D Meta-analysis of observational studies

2 A clinician responsible for recruiting patients into a trial held a sealed envelope up to the light to determine what treatment a patient would receive if randomized. He repeated this for approximately 20 % of participants, which type of bias has been introduced.

A Selection
B Attrition
C Detection
D Performance

3 Which of the following methods of randomization would be considered to be adequate?

A Tossing a coin and allocating all "heads" to intervention.
B Allocating patients alternately to intervention or control.
C Allocating patients from one clinic to intervention and the second clinic to control.
D None of the above are acceptable

4 What is meant by type II error?

A A statistically significant result is obtained by chance
B The size of the study is too small in order to be able to detect an important difference between two interventions
C Knowledge of the treatment allocation by the outcome assessors leads to a biased result
D High drop-out rates make interpretation of the study findings difficult

5 Which of the following would be an appropriate conclusion from the WHS?

A Aspirin will need to be taken at a higher dose over 10 years to reduce the incidence of CRC
B The trial findings are incompatible with a 30 % reduction in CRC risk following low-dose aspirin use
C A trial is needed which is more generalizable before we can confidently conclude that aspirin does not reduce the risk of CRC
D A high risk of type I error has made results difficult to interpret

References

1 Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609–13.
2 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
3 Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601–5.
4 Medical Research Council. Streptomycin treatment for pulmonary tuberculosis: a Medical Research Council investigation. *BMJ* 1948;2:769–82.
5 Truelove S, Witts L. Cortisone in ulcerative colitis. Final report on a therapeutic trial *BMJ* 1955:1041–8.
6 Stone GW, Pocock SJ. Randomized trials, statistics, and clinical inference. *J Am Coll Cardiol* 2010;55:428–31.
7 Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6.
8 Groves T. Mandatory disclosure of trial results for drugs and devices. *BMJ* 2008;336:170.
9 DeAngelis CD, Drazen JM, Frizelle FA, et al. Clinical Trial Registration. *JAMA* 2004;292:1363–4.
10 Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:834–40.
11 Altman DG, Schulz KF. Statistics notes: Concealing treatment allocation in randomised trials. *BMJ* 2001;323:446–7.
12 Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002;136:254–9.

13 Sackett DL, Strauss SE, Richardson WS, et al. (2000) *Evidence-based Medicine: How to Practice and Teach EBM*, Churchill Livingstone, Edinburgh.
14 Lachin JL. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:526.
15 Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials[ast]. *Clin Phar Ther* 1995;57:6–15.
16 Higgins JPT, Green S (eds.) (2009) *Cochrane Handbook for Systematic Reviews of Interventions*, John Wiley & Sons, Ltd, Chichester.
17 Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002;21:2917–30.
18 Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *Int Stat Rev* 1991;59:227–40.
19 Greenhalgh T. How to read a paper. Papers that report drug trials. *BMJ* 1997;315:480–3.
20 Charles P, Giraudeau B, Dechartres A, et al. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009;338:b1732.
21 Chan A-W, Hrobjartsson A, Jorgensen KJ, et al. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;337:a2299.
22 Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637–9.
23 Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.
24 Cook NR, Lee IM, Gaziano JM, et al. Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. *JAMA* 2005;294:47–55.
25 Cole BF, Logan RF, Halabi S, et al. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. *J Natl Cancer Inst* 2009;101: 256–66.
26 Jacobs EJ, Thun MJ. Low-dose aspirin and vitamin E. *JAMA* 2005;294:105–6.
27 Rexrode KM, Lee IM, Cook NR, et al. Baseline characteristics of participants in the Women's Health Study. *J Women's Health Gend Based Med* 2000;9:19–27.
28 Lee IM, Cook NR, Manson JE, et al. Beta-carotene supplementation and incidence of cancer and cardiovascular disease: the Women's Health Study. *J Natl Cancer Inst* 1999;91:2102–6.
29 Ridker PM, Cook NR, Lee I-M, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *New Engl J Med* 2005;352:1293–304.
30 Buring JE, Hennekens CH. The Women's Health Study: Rationale and background. *J Myocardial Ischaemia* 1992;4:30–40.
31 Buring JE, Hennekens CH. The Women's Health Study: Summary of the study design. *J Myocardial Ischaemia* 1992;4:27–9.
32 Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994;121:200–6.
33 JAMA. Instructions for Authors. *JAMA* 2005;294:119–27.
34 Gann PH, Manson JE, Glynn RJ, et al. Low-dose aspirin and incidence of colorectal tumors in a randomized trial. *J Natl Cancer Inst* 1993;85:1220–4.
35 Sturmer T, Glynn RJ, Lee IM, et al. Aspirin use and colorectal cancer: Post-trial follow-up data from the Physicians' Health Study. *Ann Intern Med* 1998;128:713–20.

Answers to multiple choice questions

1. B, D, A, C
2. A
3. A
4. B
5. B

5 How to read a systematic review and meta-analysis

Alexander C. Ford[1] *& Paul Moayyedi*[2]
[1]Leeds Teaching Hospitals Trust, Leeds, West Yorkshire, UK
[2]Division of Gastroenterology, McMaster University, Hamilton, ON, Canada

Key points

- The volume of published medical literature is increasing rapidly.
- Keeping up to date with advances in knowledge can be difficult.
- Systematic reviews and meta-analyses may aid the busy clinician by synthesizing all available evidence pertinent to a single question, and providing precise estimates of the efficacy of particular interventions.
- However, these types of study have potential biases and limitations that the reader should be alert for.

Introduction

The arrival of the Internet has increased the ease of access to medical information, and the proliferation of medical journals, both in print and online, has resulted in a greater volume of scientific articles being published. For busy clinicians keeping up to date with advances in knowledge can be time-consuming and difficult. To make matters more complicated, published studies often contradict each other. For this reason, studies that set out to report on all the available literature that has previously addressed a specific research question may provide a valuable source of information to guide clinical practice. Such studies are called systematic reviews and differ from narrative reviews, which are often based on expert opinion, and may therefore be subject to inherent biases due to the personal opinions of the author. Instead they use rigorous and reproducible methodology to identify, summarize, and extract data from all the available evidence on a given subject, using pre-specified eligibility criteria, thereby minimizing the risks of bias and hopefully increasing the reliability of the conclusions. Most systematic reviews summarize data from randomized controlled trials (RCTs), but there are some scientific hypotheses that cannot be studied using an RCT design, so methods for conducting systematic reviews have been expanded to include summaries of observational research, such as cohort and case-control studies [1].

A meta-analysis is a statistical technique, used within a systematic review, to combine and summarize the results of all available independent studies addressing the specific question being considered. It is perfectly acceptable to perform a systematic review without a meta-analysis. For instance, there may be insufficient published evidence to conduct a meta-analysis, or published studies may differ so greatly in their underlying methodology, or be at such high risk of bias, that combining them in a meta-analysis would be inappropriate. However, a meta-analysis should never be performed without first conducting a systematic review of the available literature. Unfortunately,

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

this golden rule is often ignored by researchers, and leads to erroneous conclusions being drawn from data that have not been systematically collected and analyzed.

Meta-analyses differ from other types of original research discussed in previous chapters. Rather than the researchers recruiting a group of participants themselves and setting out to study a specific research question, they instead use all the available published evidence from previous studies that address the topic and, if appropriate, combine the data using meta-analysis. This may lead some researchers to dismiss meta-analyses as "secondary" research. Nevertheless, their conduct is one of the cornerstones of evidence-based medicine, and meta-analyses are often able to provide more precise estimates of the effects of health care, which may be of sufficient importance to alter clinical practice.

Biases commonly seen in meta-analyses

The commonest bias encountered in meta-analyses is that positive studies are more likely to be identified and included, for several reasons. Publication bias occurs because studies demonstrating a beneficial effect of an intervention are more likely to be published in peer-reviewed journals, compared with those in which no statistically significant effect is demonstrated [2,3]. This is because negative studies may not be written up, due to a perceived lack of interest in the study results, or they may be deliberately "suppressed" by sponsoring agencies, such as pharmaceutical companies, and finally they are more likely to be rejected by journals.

Even when negative studies are published they are less likely to be identified by a systematic review because studies with a statistically significant result are more likely to be published rapidly (time lag bias) [4], are often published in high impact factor journals (location bias) [5], are usually published in English (language bias) [6], may be published more than once (multiple publication bias) [7], and are more likely to be cited by other authors (citation bias) [8]. These biases will tend to increase the proportion of studies demonstrating a beneficial effect of a health intervention identified in a literature search, and may therefore lead a meta-analysis to overestimate the efficacy of the intervention under study. In addition, multiple publication bias may lead to data from the same study being incorporated more than once into a meta-analysis, sometimes leading authors to claim a benefit of treatment where there was no statistically significant effect [9–11].

A final form of bias occurring in meta-analyses results from the reporting of data in the available studies included. This occurs when data concerning multiple outcomes are collected, but results are reported selectively in the published article, again usually due to the direction of results, and is referred to as outcome reporting bias [12]. The rigorous methodology that underpins a systematic review and meta-analysis is designed to minimize the likelihood of these biases occurring, but the reader should remain alert as to whether or not bias is likely to have influenced the study results.

Other important strengths and limitations of meta-analyses

A well-conducted systematic review is considered the gold-standard approach to summarizing all the available evidence addressing a specific issue. A meta-analysis of these data provides greater precision than that available from individual studies. Meta-analyses also enable data concerned with other secondary outcomes, such as adverse events, to be combined, which individual studies are usually underpowered to assess in any detail. This means that possible harms of therapies can also be estimated.

Weaknesses of meta-analyses often relate to either the quantity or the quality of the data available. If there are only a small number of individual studies addressing the question of interest, containing few participants, then the 95 % confidence intervals (CI) of the effect obtained from the meta-analysis may be wide, and the estimate of effect is likely to be altered substantially by the publication of further, larger, studies. Although a meta-analysis of RCTs is set at the highest level within the hierarchy of research evidence, meta-analyses themselves suffer from the "garbage in, garbage out" phenomenon. In other words, the quality of meta-analyses depends upon the quality of the data that are available for summary within them. If the individual studies that provide data are all poorly designed and subject to bias, then the meta-analysis will also be flawed as a result. In addition, individual

studies can differ so much in terms of their underlying methodology, or the outcomes reported, that performing a meta-analysis may be inappropriate [13]. A final weakness, as with any other type of research, is the potential for investigators to make errors during the conduct of the meta-analysis itself. When a series of eight meta-analyses reporting on the efficacy of various medical therapies for irritable bowel syndrome were deconstructed, there were errors in terms of identifying truly eligible studies in six, inclusion of studies that were ineligible in five, and errors in data extraction in all of the meta-analyses [14]. These led to errors in 15 of the 16 treatment effects reported in the meta-analyses, and a change in the statistical significance of the recalculated treatment effect in four cases.

Important issues to consider when reading a meta-analysis

The archetypal methodology for conducting and analyzing a systematic review and meta-analysis is thoroughly described in the *Cochrane Handbook for Systematic Reviews of Interventions* (JPT Higgins, S Green (eds.)). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011 (available from www.cochrane-handbook.org). The Handbook is to be commended to anybody keen on reading further about the subject. Similarly, the PRISMA statement (available from www.prisma-statement.org/) gives a consensus approach to the clear reporting of such work. The 10-point checklist below outlines a shorthand approach to critically appraising such work.

1 Did the authors set out suitable eligibility criteria a priori, and were these adhered to?

Once the specific question that the meta-analysis has been designed to address is decided upon, the authors need to have defined the eligibility criteria for the types of studies they are going to consider for inclusion in the meta-analysis. This should be done before the literature search is commenced, and cannot be altered at a later date on the basis of the studies that are identified subsequently. These criteria should be reported clearly in the article. In a meta-analysis of treatment trials for the efficacy of a therapy in a particular disorder a commonly used mnemonic for defining study eligibility is PICO: where P is patient group (the disorder under study); I is intervention (the new treatment); C is comparator (the old treatment, or placebo); and O the outcomes under study (usually treatment efficacy and adverse events arising as a result of therapy). These are a guide only, and there may be other eligibility criteria required by the study authors, such as age of participants, minimum duration of therapy or follow-up within the trials, or a specific subgroup of patients within the disease under study, all of which will depend upon the nature of the question the meta-analysis has been designed to answer.

2 How did the authors perform a comprehensive search of the medical literature?

The authors should have designed a suitable search strategy, and applied it to more than two electronic databases. The most commonly used databases are MEDLINE, EMBASE, CINAHL, and the Cochrane central register of controlled trials. A meta-analysis is a synthesis of research, which requires scientific rigor, so the reporting of methodology needs to be in sufficient detail for an independent investigator to replicate the study methods exactly, if they so wished. For this reason the authors need to specify which databases were searched, including the dates between which the search was conducted, the search terms used in the databases, and how these were combined. The latter is usually done with Boolean logic set operators such as AND, OR, and NOT. Electronic databases usually do not include all available evidence, so there should be evidence that the authors conducted an exhaustive search using other methods to minimize the risk that studies have been missed. This includes some, or all, of: a recursive search of the bibliographies of eligible studies to identify other potentially eligible articles that the electronic search may have failed to identify; a search of the "gray" literature, consisting of conference proceedings, in order to find studies published only in abstract form; and contact with experts in the field and/or pharmaceutical companies to try to obtain data from unpublished studies.

3 Did the authors undertake assessment of eligibility in duplicate?

All potentially relevant studies within the citations identified by the literature search should be retrieved and their eligibility judged according to the predefined criteria. In order to minimize the risk of bias when assessing the eligibility of identified studies this should be performed in duplicate, and the fact that this was

done should be stated in the study methodology. The reader needs to be able to assess the validity of this process, and so there should be a report of the degree of agreement between investigators concerning eligibility of individual studies identified. This is usually done using a Kappa statistic, and a value of 0.60 or more indicates a good level of agreement. In addition, the authors should explain how disagreements in study eligibility were dealt with. This is usually achieved either by discussion to reach consensus, or by adjudication from a third person. Ideally, the entire process of dealing with the citations identified in the literature search through these various stages in eligibility assessment should be reported, from the number identified, number retrieved for eligibility assessment, number excluded (with the reasons for exclusion), through to the number of studies finally eligible and included, reported using a flow diagram.

4 How did the authors assess for risk of bias in the studies they included?

There are specific methods for assessing the risk of bias of studies that are included in a meta-analysis, depending on the design of the studies being reviewed. For RCTs, this is well characterized, and includes reporting of the method used to generate the randomization schedule and conceal treatment allocation, presence of blinding, completeness of follow-up, and whether there is evidence of selective reporting of study outcomes [15–17]. For observational studies these criteria are less researched, but include masking of outcome assessors to whether or not the subject has the disease of interest [18,19]. This information should be provided in detail for the reader in a table, so that the risk of bias of all included studies underpinning the meta-analysis can be appraised.

5 How did the authors assess for publication bias?

Publication bias may have an impact on the summary result of a meta-analysis for the reasons discussed earlier. The authors should therefore have assessed whether there is any evidence of this using funnel plots. These are graphical representations of the study effect size on the *x*-axis and a measure of overall study size on the *y*-axis. In theory, large studies should group together closely around the summary effect size, denoted by a vertical line, towards the top of the plot. Smaller studies should be spread more widely either side of the vertical line towards the bottom of the plot, due to their reduced ability to provide a precise estimate of the summary effect size. Therefore the plot should appear symmetrical, as in Figure 5.1. If publication bias exists, however, there will be a dearth of small negative studies in the lower right quadrant of the plot, leading to asymmetry, as in Figure 5.2. Statistical tests can be applied to funnel plots, such as the Egger test [20], to assess whether the degree of funnel plot asymmetry is likely to be statistically significant.

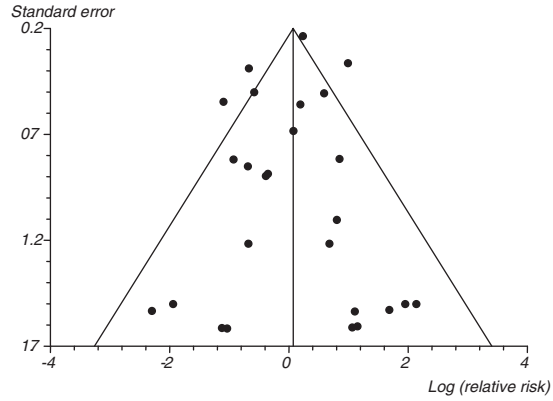


Figure 5.1 Symmetrical funnel plot.

6 How did the authors summarize the characteristics of the individual eligible studies they identified?

Information about the characteristics of all the studies contributing data to the meta-analysis should be reported, ideally in a table of included studies. This is helpful because it provides more precise information about the types of participant included in the studies that contribute data to the meta-analysis, as well as details of the interventions applied, such as the

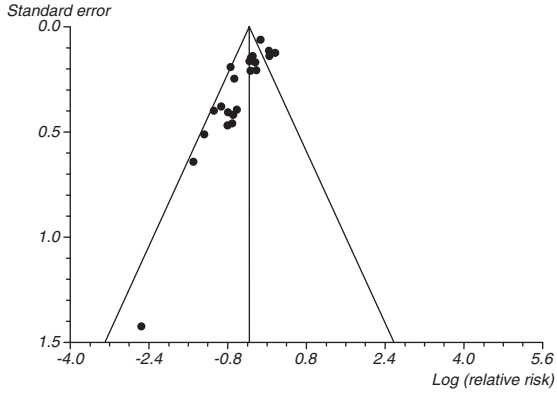


Figure 5.2 Asymmetrical funnel plot.

duration and dose of therapies used, allowing the individual reader to assess whether the findings are likely to be able to be applied to their own routine clinical practice, and whether it was appropriate to perform the meta-analysis.

7 Was it appropriate to perform a meta-analysis based on extractable study data and risk of bias of studies?

Extractable data provided by eligible studies needs to be assessed carefully by the investigators to ensure it is appropriate to combine them. If studies report their results using different outcome measures to define treatment success, use vastly different doses or durations of therapy, or there are large differences in duration of follow-up between studies then combining the results in a single meta-analysis may not be appropriate. It may still be possible to combine data, if there are a sufficient number of eligible studies, and conduct subgroup analyses according to these different characteristics, in order to assess their effect on the result of the meta-analysis. Sometimes the risk of bias of the studies included is so high that performing a meta-analysis to obtain a summary effect size is of little value. At this point, the reader has to question the rationale for conducting one at all [21,22]. It is well worth remembering the proverb that it is not possible to make a silk purse out of a sow's ear, and similarly performing a meta-analysis for its own sake does not compensate for underlying poor methodology among all the included studies.

8 Were the data analyzed correctly?

When analyzing the data, the first step is to decide on the summary statistic to use, such as a relative risk, odds ratio, risk difference, or another ratio, such as a proportion [23]. In a meta-analysis of RCTs it is usual to use a relative risk, whilst if the results of individual case-control studies are being combined it is more appropriate to use an odds ratio. Once this has been decided, the summary statistic should be calculated for each study included in the meta-analysis, and the overall result is obtained using a weighted average of these individual summary statistics.

There are two approaches to combining data in a meta-analysis. One approach is to use the fixed effect model, which assumes that each study is measuring the same underlying effect, and thus any variation between studies is due to chance. In this analysis the weight given to individual studies within the meta-analysis depends on the rate of the event of interest in each study. If there is heterogeneity between studies, then the key assumption for a fixed effect model is incorrect, and it is probably not appropriate to use this. The other method is to use a random effects model, which does not make the assumption that all the studies are measuring the same underlying effect. In this analysis, a constant is added to the weighting of the studies that relates to the between-study variance, making the relative weighting of studies similar, but widening the confidence interval of the summary statistic, therefore giving a more conservative estimate. One problem with the latter approach is that smaller studies, which are often at greater risk of bias than larger studies, are then given more emphasis in the meta-analysis compared to fixed effect models, and this may lead to bias in the overall result. For these reasons, there is no clear consensus on which of the two models should be preferred to synthesize the data [24].

9 Did the authors assess for, and explore potential explanations for, heterogeneity between individual study results?

Individual studies, when combined in a meta-analysis, may give diverse results. This inconsistency between individual study results is termed heterogeneity, and may arise due to clinical or methodological differences between studies. Clinical differences include those related to the study participants, such as age, sex, ethnicity, or how the presence of the condition under study was defined, and those related to the intervention under investigation, for example the use of different drugs within the same class. Methodological differences include differences in study design, risk of bias of the study, definition or recording of the outcome of interest, and differences in how the intervention was applied to the participants. However, heterogeneity may also occur due to chance, so a statistical test is usually applied to study results in order to assess whether the degree of heterogeneity observed is significant.

The authors of a meta-analysis should assess for the presence of heterogeneity, ideally using either the Cochrane Q test, with P values <0.10 indicating statistically significant heterogeneity, or the I^2 value, which ranges from 0 % to 100 %, with 0 % representing no observed heterogeneity, and larger values indicating increasing heterogeneity. A value around 25 % is usually chosen arbitrarily to represent low levels of heterogeneity [25]. Where statistically significant heterogeneity exists, the authors should explore

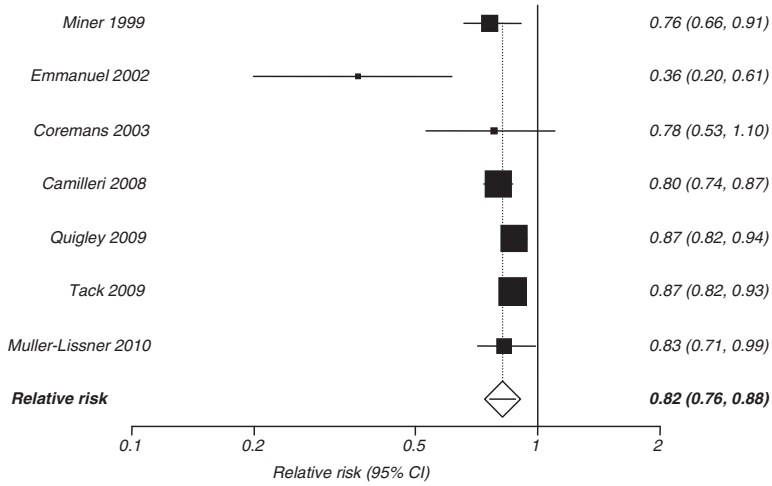


Figure 5.3 Meta-analysis of randomized controlled trials of prucalopride in chronic idiopathic constipation.

(**How to read a Forest plot:** Forest plots are a graphical representation of a group of studies. Each box represents an individual study, with the size denoting the weight given to it in the meta-analysis. The horizontal line through the box is the 95 % confidence interval (CI) of that study result. The dotted vertical line on the plot is the overall result when all studies are combined, and the diamond at the bottom of the plot represents the 95 % CI of this. If the diamond crosses the vertical axis (labeled 1) then the combined study result is not statistically significant. In this example the diamond is to the left of the vertical axis, suggesting that prucalopride is effective in the treatment of chronic idiopathic constipation.)

reasons for this, which can be done using either subgroup analysis or meta-regression.

Subgroup analysis separates studies according to factors that may be important in causing the heterogeneity, allowing the investigator to assess whether heterogeneity is less within each group. For example, in a meta-analysis of RCTs studying the efficacy of prucalopride in chronic idiopathic constipation there was heterogeneity between the seven individual trial results (Figure 5.3), with an I^2 value of 60 % [26]. However, when the effect of the definition of chronic idiopathic constipation used in the studies was examined, six used the Rome II criteria. When only these studies were included in the analysis the degree of heterogeneity observed was much lower (Figure 5.4), with an I^2 value of 13 %.

Meta-regression is a technique that is akin to logistic regression analysis but uses individual studies, rather than individual participants, as the unit in the analysis [27]. Its advantage is that it can be used to adjust for multiple study characteristics in the analysis simultaneously. Unfortunately, because it uses study level data, it evaluates the average of patient characteristics within each trial, and this summary data may misrepresent individual patients within each treatment arm. The technique is therefore vulnerable to giving spurious results, due to the ecological fallacy [28].

10 Were the conclusions drawn valid based on the results?

When reading the discussion of a meta-analysis it is important to ask, as with any research article, whether the conclusions drawn by the authors are supported by the data they present. If there is evidence of one or more of: high risk of bias of included studies; publication bias; or heterogeneity between studies when results are combined then the authors should acknowledge that this may limit the robustness of the overall summary result, and be more guarded in the conclusions that they draw from the meta-analysis. To quote Thompson and Pocock: "Meta-analysis is not an exact statistical science that provides definitive simple answers to complex clinical problems. It is more appropriately viewed as a valuable objective descriptive technique, which often furnishes clear qualitative conclusions about broad treatment policies, but whose quantitative results have to be interpreted cautiously" [24].

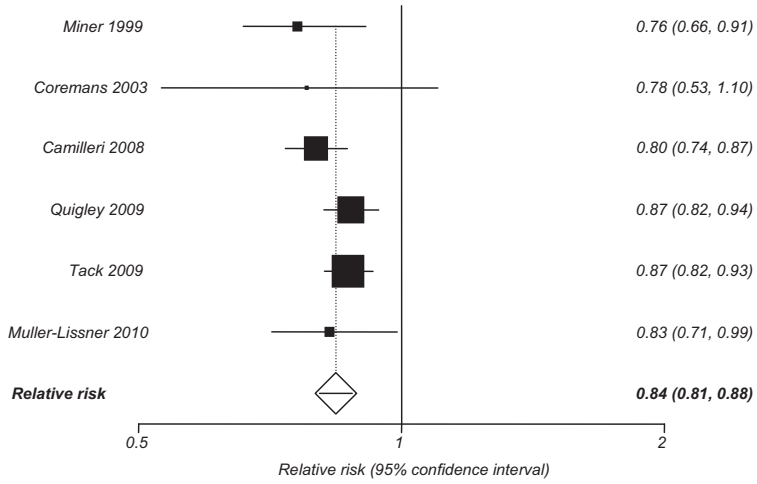


Figure 5.4 Meta-analysis of randomized controlled trials of prucalopride in chronic idiopathic constipation (sensitivity analysis according to definition of chronic idiopathic constipation used).

Case study: Critical evaluation

Aspirin may have a chemo-preventative effect on neoplasia in the colon. As a result, there has been considerable interest in long-term aspirin therapy as a means of reducing the incidence of colorectal adenoma. Several RCTs have been conducted to test this hypothesis, but individual results are conflicting, and studies may not have been designed with sufficient power to examine this endpoint, leading to a failure to detect a significant effect of aspirin in this situation. Cole et al. therefore conducted a systematic review and meta-analysis of available RCTs examining this issue, using individual patient-level data [29].

1 The authors set out their eligibility criteria clearly.

2 They reported their search strategy, which included contact with experts in the field.

3 However, the authors did not explain how study eligibility was assessed.

4 Nor did they undertake a formal assessment of risk of bias of the studies, although they stated in their eligibility criteria that all studies had to be double-blind placebo-controlled RCTs to be included.

5 In terms of publication bias, there was no evidence that the presence of this was assessed, although with only four trials eligible for inclusion, the power to detect this within the meta-analysis would be low.

6 Individual characteristics of both the trials, and the patients included within them, were provided in comprehensive detail.

7 As the authors had obtained individual patient-level data from the authors of the four RCTs they identified, performing a meta-analysis was appropriate, as differences in individual study design were able to be adjusted for by subgroup analyses, which the authors performed according to individual patient characteristics such as sex, age, body mass index, and number of previous adenomas.

8 Data from these four RCTs were combined correctly using a relative risk, and also a random effects model, meaning that the meta-analysis is likely to provide a more conservative estimate of the efficacy of aspirin in preventing colorectal adenomas.

9 Heterogeneity was assessed between studies using the both the Cochrane Q test, with a P value <0.05, and an I^2 value $>50\,\%$. Using this definition, heterogeneity between the studies for the primary analysis was not statistically significant ($P = 0.16$, $I^2 = 41.5\,\%$), although again with only four trials the power to detect this is low, and many researchers would define an I^2 of 41.5 % as indicating moderate heterogeneity. The authors reported a benefit of aspirin in the prevention of colorectal adenoma, and this observation remained stable across the majority of subgroup analyses conducted, although there was significant heterogeneity in some of these. Interestingly, there was no dose-response relationship demonstrated, with higher dose aspiring having no statistically significant effect on adenoma prevention, and the significant benefit of aspirin attenuated with duration of follow-up beyond 3 years. Due to the collection of

individual patient-level data, the authors were able to assess adverse event rates in the studies, which were no higher with aspirin than with placebo.

10 The authors concluded that there was a 17 % reduction in risk of any colorectal adenoma with aspirin therapy, and a 7 % reduction in absolute risk. The lack of a convincing dose–response effect as a potential limitation of their study was noted in the discussion. Their conclusion that aspirin reduced the risk of recurrence of colorectal adenomas, but that these benefits needed to be considered in the context of potentially deleterious effects of aspirin, seems justified.

10-Point checklist: How to read a meta-analysis

1 **Did the authors set out suitable eligibility criteria a priori, and were these adhered to?**
- *For instance PICO: patient; intervention; comparator; outcome*
- *These must not be altered after the search has been conducted.*

2 **How did the authors perform a comprehensive search of the medical literature?**
- *More than two electronic databases should be searched, with dates up to when the search was conducted*
- *Hand-searching of conference proceedings*
- *Recursive search of retrieved articles bibliographies*
- *The search strategy should be reported as part of the methods, and be reproducible.*

3 **Did the authors undertake assessment of eligibility in duplicate?**
- *This must be done in order to reduce the risk of bias in the selection process*
- *The authors should report the degree of agreement in assessing study eligibility using a Kappa statistic.*

4 **How did the authors assess the quality of the studies they included?**
- *Using the Cochrane Handbook criteria for RCTs, the STROBE statement for observational studies, or other published and/or widely accepted criteria*
- *This information should be reported in a table of included studies.*

5 **How did the authors assess for publication bias?**
- *Funnel plot asymmetry should be examined, where there are sufficient studies*
- *This can be done visually, or with a statistical test, the Egger test.*

6 **How did the authors summarize the characteristics of the individual eligible studies they identified?**
- *Study characteristics should be reported in a table of included studies.*

7 **Was it appropriate to perform a meta-analysis based on extractable study data and study quality?**
- *Appropriateness may be objective, but usually depends on study quality, number of studies, and whether there are large differences in underlying study methodology.*

8 **Were the data analyzed correctly?**
- *Was choice of summary statistic appropriate?*
- *Was a fixed effect or random effects model used?*

9 **Did the authors assess for, and explore potential explanations for, heterogeneity between individual study results?**
- *Statistical testing to assess whether the degree of heterogeneity between studies is significant*
- *Sensitivity analyses according to study characteristics to explore reasons for heterogeneity.*

10 **Were the conclusions drawn valid based on the results?**

Multiple choice questions

1 It is not possible to perform a meta-analysis of which of the following types of study:
- **A** Case-control studies
- **B** Cohort studies
- **C** Case reports
- **D** Studies of the accuracy of a diagnostic test
- **E** Randomized controlled trials

2 Which of the following are not types of bias that may be encountered during the conduct of a meta-analysis:
- **A** Publication bias
- **B** Outcome reporting bias
- **C** Citation bias
- **D** Selection bias
- **E** Multiple publication bias

3 Stages in the literature search conducted for a meta-analysis may include which of the following:
- **A** An electronic search of MEDLINE and EMBASE
- **B** Hand-searching of conference proceedings
- **C** Contacting pharmaceutical companies

D A recursive search of retrieved articles bibliographies

E All of the above

4 Which of the following are not used to assess the quality of studies in a meta-analysis:

A Duration of follow-up of participants

B How randomization was performed

C Whether blinding was employed

D Whether there was evidence of selective reporting of outcomes

E How treatment allocation was concealed

5 Which of the following would not account for heterogeneity between studies in a meta-analysis:

A Differences between study participants

B Differences between principal investigators of the studies

C Differences in the way the intervention was applied in the studies

D Differences in how the outcome of interest was defined among studies

E Chance

References

1 Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008–12.

2 Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisted: The effect of the outcome of statistical tests on the decision to publish and vice versa. *Am Stat* 1995;49:108–12.

3 Hopewell S, Loudon K, Clarke MJ, et al. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009;Jan 21;(1):MR000006.

4 Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;279:281–6.

5 Pittler MH, Abbot NC, Harkness EF, et al. Location bias in controlled clinical trials of complementary/alternative therapies. *J Clin Epidemiol* 2000;53:485–9.

6 Egger M, Zellweger Z, Schneider M, et al. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;350:326–9.

7 Gotzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989;36:429–32.

8 Gotzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;295:654–6.

9 Fuccio L, Zagari RM, Eusebi LH, et al. Meta-analysis: Can *Helicobacter pylori* eradication treatment reduce the risk for gastric cancer? *Ann Intern Med* 2009;151:121–8.

10 Ford AC, Moayyedi P. Redundant data in the meta-analysis on *Helicobacter pylori* eradication. *Ann Intern Med* 2009;151:513.

11 Correction: Can *Helicobacter pylori* eradication treatment reduce the risk for gastric cancer? *Ann Intern Med* 2009;151:516.

12 Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: Review of publications and survey of authors. *BMJ* 2005;330:753.

13 Moayyedi P. Meta-analysis: Can we mix apples and oranges? *Am J Gastroenterol* 2004;99:2297–301.

14 Ford AC, Guyatt GH, Talley NJ, et al. Errors in the conduct of systematic reviews of pharmacological interventions for irritable bowel syndrome. *Am J Gastroenterol* 2010;105:280–8.

15 Higgins JPT, Green S. (2011) *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0. Available at: www.cochrane-handbook.org (last accessed May 16, 2013).

16 Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Control Clin Trials* 1996;17:1–12.

17 Juni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6.

18 Newcastle-Ottawa quality assessment scale case control studies (2011). http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed October 11, 2013).

19 von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet* 2007;370:1453–7.

20 Egger M, Davey-Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34.

21 Habermalz B, Sauerland S. Clinical effectiveness of selective granulocyte, monocyte adsorptive apheresis with the adacolumn device in ulcerative colitis. *Dig Dis Sci* 2010;55:1421–8.

22 Thanaraj S, Hamlin PJ, Ford AC. Is the benefit of granulocyte monocyte adsorptive apheresis in ulcerative colitis overstated? *Dig Dis Sci* 2010;55:1803.

23 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;21:1575–600.

24 Thompson SG, Pocock SJ. Can meta-analyses be trusted? *Lancet* 1991;338:1127–30.

25 Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.

26 Ford AC, Suares NC. Effect of laxatives and pharmacological therapies in chronic idipathic constipation: systematic review and meta-analysis. *Gut* 2011;60: 209–18.
27 Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559–73.
28 Lau J, Ioannidis JPA, Schmidt CA. Summing up evidence: one answer is not always enough. *Lancet* 1998;351: 123–7.
29 Cole BF, Logan RF, Halabi S, et al. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. *J Natl Cancer Inst* 2009;101: 256–66.

Answers to multiple choice questions

1. C
2. D
3. E
4. A
5. B

6 How to decide if any of this matters

Kate M. Fleming & Timothy R. Card
Division of Epidemiology and Public Health, Nottingham City Hospital, University of Nottingham, Nottingham, UK

Key points

- Not all associations are causal. Chance, bias, confounding, and reverse causality are important alternative explanations which need to be considered.
- The evidence hierarchy and criteria to judge causality can help in the assessment of literature.
- Even where a risk factor (or protective effect) is shown to be causally related to an outcome, its importance needs to be judged.
- Clinically important associations have large absolute effects on important outcomes and can be altered.
- The number needed to treat (or to harm) can be a useful measure when examining the effect of altering exposures.

Introduction

In this chapter we will concentrate upon a scenario in which you the reader, having read a paper or papers describing a study or studies (perhaps of one of the types described in previous chapters), are trying to decide whether the results matter to you. That is, should they alter your practice? To decide this, you are likely to wish to know at least three things:

1 Is the relationship described in the paper causal?
2 Is the relationship important?
3 Is the relationship something we can affect or influence?

You will not necessarily have the answers to all of the above questions, but knowing some will help you to decide whether to invest time and possibly money in investigating the others. Over the following pages, we will endeavor to give some pointers as to how these questions can be addressed both from the data within a paper and from the wider literature. To do this, we will utilize as specific examples the papers described in the previous chapters to address the question of whether aspirin use can prevent colorectal cancer. Those who are interested in reading further into this subject could do worse than to look at the *User's Guide to the Medical Literature* published by *JAMA* in the 1990s [1].

Is the relationship described in the paper causal?

The attribution of causality in observational research is a difficult subject which alone can occupy far more space than this chapter provides. Those wishing to read further may find the work of Rothman useful in this regard [2]. In epidemiology, one of the ways in which we approach the decision about whether a relationship is causal is by considering the other options. These can generally be simplified to the possibilities

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

that the relationship is due to chance, bias, confounding, or reverse causality. In the reading of epidemiologic studies we can consider the potential role of these options if we know what they are and how they can occur and be detected. *Chance* (the potential for a relationship to have occurred due to random variation) is assessed in the statistical analyses within a paper. Ninety-five percent confidence intervals (CI) tell us in essence that there is a 95 % *chance* that the true value lies within the range that they describe. A *P* value of 0.05 states that when considering whether a difference between observed samples is due to a true difference between populations, the observed or a more extreme result will occur by *chance* on 5 % of occasions. *Confounding* occurs when an apparent association between two factors is influenced by their shared association with a third. A classic example is that the drinking of alcohol may be associated with lung cancer not because alcohol causes lung cancer, but rather because drinkers smoke more than the average and smoking causes lung cancer. To assess the possibility that this is of importance one must first think of what third factors may be present and then examine to what extent the authors have corrected for any effect in their analysis. A discussion of the statistical techniques which can be so employed is beyond the scope of this chapter, but in brief the commonly used options are to conduct stratified analyses (where all subjects in a stratum who are analyzed together are equally exposed to the potential confounder) and/or to conduct multivariate analyses. Further descriptions of such techniques can be found in standard statistical texts. *Bias* is a systematic deviation from the truth in either measurement, selection of supposedly representative samples or other factors influencing the results or inferences from a study. A number of specific opportunities for this to occur have been discussed in preceding chapters and a further discussion can be found in Rothman's textbook [3]. Bias can never be corrected for in analysis and is minimized by optimal methodology. In reading the literature one must always be alive to this possibility since bias can occur in many ways. *Reverse causality* occurs when what is being regarded as an effect is in fact causing what is being regarded as an exposure. As with bias the best the reader can do is to seriously consider this possibility with regard to the relationship studied.

As is clear from the previous paragraph, thinking through all possible noncausal explanations can involve appreciable effort for the reader. When one comes to determining one's own practice the issues become even more complicated as there will be many sources of information to guide decisions and not just one paper. There are, however, some rules of thumb which can help when thinking of causality in other ways. We are going to consider two of these.

The evidence hierarchy

Evidence-based medicine has been much discussed and variously praised or abhorred over recent years. One of the ideas that has been common in this field and can help us as a shortcut in assessing research is the hierarchy of studies. Various versions of this have been proposed, but a fairly expansive version is suggested by the Oxford Centre for Evidence-Based Medicine where level 1 is considered the highest or best form of evidence and level 5 the least (Table 6.1).

As can be easily seen from the descriptions of the levels given, this list does not free the reader from considering the role of bias, although as long as one has a knowledge of the individual study types, it does offer some assistance in recognizing those that will provide the most reliable results. It therefore helps in choosing between pieces of evidence. To elucidate a little further we can consider why the hierarchy is ordered as it is. The best evidence is provided by a well-conducted meta-analysis of good-quality randomized controlled trials (RCTs). The meta-analysis or systematic review is likely to be superior to an individual RCT not primarily because it provides increased power, but because one would not expect identical biases to operate in multiple trials, and so the combination of studies is less likely to be biased in one direction than is any individual study. The great benefit of the RCT is not that it is unbiased (if poorly conducted, it can be very easily), but rather that it has the potential to remove the effects of all confounders both known and unknown. In this it is clearly superior to any of the observational methodologies, which at best allow for correction of suspected or recognized confounders. When descending the hierarchy to the observational studies, the same arguments regarding meta-analyses apply, and hence systematic reviews and meta-analyses of observational studies are similarly above the individual studies from which they are constituted. Applying these ideas to the papers on aspirin and CRC

Table 6.1 Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence

Question	Step 1 (Level 1*)	Step 2 (Level 2*)	Step 3 (Level 3*)	Step 4 (Level 4*)	Step 5 (Level 5)
How common is the problem?	Local and current random sample surveys (or censuses)	Systematic review of surveys that allow matching to local circumstances**	Local non-random sample**	Case-series**	n/a
Is this diagnostic or monitoring test accurate? (Diagnosis)	Systematic review of cross sectional studies with consistently applied reference standard and blinding	Individual cross sectional studies with consistently applied reference standard and blinding	Non-consecutive studies, or studies without consistently applied reference standards**	Case-control studies, or "poor or non-independent reference standard**	Mechanism-based reasoning
What will happen if we do not add a therapy? (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*	Case-series or case-control studies, or poor quality prognostic cohort study**	n/a
Does this intervention help? (Treatment Benefits)	Systematic review of randomized trials or *n*-of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study**	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning
What are the COMMON harms? (Treatment Harms)	Systematic review of randomized trials, systematic review of nested case-control studies, *n*-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect	Individual randomized trial or (exceptionally) observational study with dramatic effect	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning
What are the RARE harms? (Treatment Harms)	Systematic review of randomized trials or *n*-of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect			
Is this (early detection) test worthwhile? (Screening)	Systematic review of randomized trials	Randomized trial	Non -randomized controlled cohort/follow-up study**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning

Source: Reproduced from: OCEBM Levels of Evidence Working Group. Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653
*OCEBM Table of Evidence Working Group = Howick et al.

which have been discussed in previous chapters, one can argue that if they were answering the same question, it is likely that the meta-analysis [4] would be more reliable than the RCT [5], which in turn would be more reliable than the observational studies [6].

Austin Bradford Hill

Another influential and much cited aid to the determination of causality is what are commonly referred to as Austin Bradford Hill's criteria. These come from a presidential address given by Professor Hill to the Royal Society of Medicine [7] and within the same address is contained a clear warning against considering the "viewpoints" he proposes as being either necessary or sufficient for the attribution of causality individually. Nevertheless, this list can, if used with caution, help us in considering causality. When doing so, however, the counterarguments to this list should be considered [2].

Strength

The first consideration given is the strength of association. A strong association is more likely to suggest causality because it is less likely to occur purely through the action of a bias or confounder. (In the example of aspirin and colorectal cancer the association is relatively weak with the protective effect against adenomas in the meta-analysis having a risk ratio of 0.83 and that against colorectal cancer having an odds ration of 0.78 in the case-control study and a hazard ratio of 0.77 in the cohort study.)

Consistency

An association found repeatedly and consistently is less likely to be due to chance bias or confounding. This is particularly true where different forms of study produce like answers as it would be harder for biases or confounders to operate consistently under these circumstances.

Specificity

An association which is specific is more likely to be causal. In the example of aspirin and colorectal cancer, the case would be less convincing if it were claimed that all cancers were reduced by all medications used rather than this one specific drug preventing one specific cancer. However, there are numerous epithelial gastrointestinal malignancies which appear to be reduced in those exposed to aspirin [8] and this should not discourage a belief in causality if there is a clear mechanism to explain this.

Temporality (timing)

An exposure occurring after an outcome cannot cause it. Hence a consideration of temporal relationships can help us in the consideration of whether it is more likely that A causes B or B causes A. (In the aspirin and CRC example the exposure is clearly before the outcome in each of the studies.)

Biological gradient (dose response)

It is generally easier to believe that an exposure causes an associated risk of disease if greater exposure is associated with greater risk (i.e. there is a dose-response relationship). (In the example of aspirin and colorectal cancer the cohort study shows a clear dose relationship with increasing use of NSAIDs giving a lower risk of cancer. The meta-analysis of RCTs for adenoma detection does not show a clear dose-response relationship.) However, the lack of such a gradient is clearly not necessarily a deterrent to belief in causality since alcohol consumption's relationship with cirrhosis appears to have a threshold response rather than a graded one [9], and few gastroenterologists do not believe alcohol causes cirrhosis.

Plausibility

If there is a known biologic or cellular mechanism which could explain how an exposure causes an illness then it will be easier to believe that any association is causal. This does not mean, however, that we should necessarily refuse to believe causal associations for which we cannot comprehend the mechanism. The lack of apparent plausibility may as easily indicate the limitations of biologic scientific knowledge as a lack of causality.

Coherence

Although novel observations should not be discounted, it is clearly easier to believe that an association is causal if this idea does not conflict with

other existing knowledge. (In the example of aspirin and colorectal cancer therefore the combination of data supportive of the relationship from multiple sources is suggestive of causality.)

Is the relationship important?

The concept of importance of an observed association is one that will vary considerably between individuals, it being influenced by many factors.

For example, the association between aspirin use and colorectal cancer may be viewed as wholly unimportant to a population of children to whom colorectal cancer is largely unheard of and in whom aspirin use is relatively rare and mostly contraindicated. However, the same association may be viewed as extremely important to drug manufacturers wishing to create a drug with similar efficacy to aspirin but without its adverse effects. A person, or population of people, considered at higher risk of colorectal cancer may wish to balance the advantageous effects of taking aspirin in reducing the risk of colorectal cancer with its adverse effect of an increase in gastrointestinal bleeding. Likewise the clinician considering prescribing aspirin will balance these risks before deciding on a course of action. Which way the balance tips will depend on many things, including:

- the relative and absolute risks
- the (measured or perceived) severity of disease.

We now approach the consideration of these two points mainly from the perspective of the practicing clinician but never forgetting the views of the patient.

Relative versus absolute risk

The relative risk of a disease is (strictly speaking) the ratio of the risk of disease in one group of people compared with the risk of disease in another group of people. However, it is common for the term "relative risk" to encompass many ratio measures of effect (including odds ratios and rate ratios) sometimes confusing the reader. The absolute risk of disease is the probability of developing a particular disease, ideally over a specified time period. Authors and journalists alike are of course keen to "sell" their papers and the reporting of a number that implies a strong(er) association will almost always appear more attractive at first glance than one with a weak(er) association. One of the most regular ways in which risk is misconstrued is through the reporting of relative risks in the absence of absolute risks.

Consider the paper by Chan et al. discussed in Chapter 2 on "How to read a cohort study". The second line of the results section of the abstract reports "Among women who regularly used aspirin the multivariate relative risk (RR) for colorectal cancer was 0.77 (95 % CI 0.67–0.88)". Until we read the full results section we are not aware that this 23 % decrease in risk could otherwise be presented as an absolute risk of colorectal cancer of 5.56 cases per 10,000 women regularly using aspirin compared with an absolute risk of colorectal cancer of 6.36 cases per 10,000 women who did not regularly use aspirin, that is, an absolute risk reduction of only 0.8 cases per 10,000.

When assessing the importance of an association it is therefore crucial to be cognizant of the fundamental difference between relative risks and absolute risks and to interpret any reported "relative risk" in the context also of the absolute risk. Gigerenzer is a long-term advocate of the reporting of risks through frequency statements, natural frequencies and absolute risks, rather than the preponderance of single-event probabilities and relative risks seen in the medical literature [10] and writes eloquently on how best to present these pieces of information to patients to enable them to make the most informed decision possible. His papers and books on the communication of risk are to be recommended to all physicians involved in direct patient care and decision making.

Severity of disease

Individual diseases often have a "bespoke" system to designate the severity of disease from a clinical standpoint. For example Child-Turcotte-Pugh score for cirrhosis based on measurements of blood indices and presence or absence of certain symptoms and signs [11]. Similarly in colorectal cancer the clinical severity of disease is largely measured by the stage of disease as defined by Dukes' classification [12], taking into account progression of disease to lymph nodes and further spread with latter stages representing more serious disease with a worse overall prognosis.

From a perspective of considering the relative importance of a specific disease compared with

another, crudely one may consider which disease would be most likely to shorten overall life expectancy by the largest length of time. In reality, the severity of a disease is not limited to its influence on mortality but is also related to its effect on morbidity and quality of life. Two frequently used measures in assessing the relative severity of disease are those of the disability-adjusted life year (DALY) and the quality-adjusted life year (QALY). Essentially a DALY is a year of full, healthy life lost to the disease with a QALY taking into account any years of life which may be lived at less than full health (counted therefore as a fraction of one full healthy year). These somewhat subjective measures, being based on the assignation of disability weights and the judgments of patients and professionals, are nonetheless useful as by considering the likely magnitude of DALYs or QALYs gained or lost through acquisition of competing diseases both patients and professionals alike can assess the relative severity of these diseases to inform their choices.

Ultimately, what constitutes an "important association" is of course entirely at the discretion of an individual. We will all have different ideas about the relative values of life and disability, for example. One person may be willing to accept a 10 % risk of stroke for a 10 % chance of increasing their life expectancy for a year, whereas another may not. Doctors' agencies, licensing authorities, and so on, may additionally have their own guidelines as to the acceptability of risks often invoked when considering the likely benefit of introducing a specific treatment or therapeutic intervention, but decision making at the level of the physician–patient dyad should allow for dialogue and clear interpretation of all available information.

Is the relationship something we can affect or influence?

Assuming that we believe an association to be causal in a population, we then have to question if there is anything we can do to either enhance a beneficial association or reduce a harmful association. Essentially the whole of medical practice is based on this fundamental principle even when the evidence to support or refute an association is not as robust as perhaps we may wish it to be.

In the case of aspirin and colorectal cancer, we will want to know who will most benefit from taking aspirin with respect to it potentially reducing the risk of colorectal cancer, whether taking aspirin will lead to such adverse effects as to render its protective ability against colorectal cancer irrelevant, and the cost implications associated with the course of action we decide to take. To do this we can use a variety of methods from simple maths to more complicated health economic judgments.

Number needed to treat (NNT) and numbers needed to harm (NNH)

Both the number needed to treat (NNT) and number needed to harm (NNH) give simple, readily accessible and understandable values to the potential benefit and harm of an intervention. The NNT is defined as "the number of persons needed to be treated, on average, to prevent one more event" [13]. Where the treatment acts independently of other background factors it can be calculated as the reciprocal of the absolute risk reduction. The NNH is defined as "the number of persons needed to be treated, on average, to produce one more adverse event" [13]. Where the treatment acts independently of other background factors it can be calculated as the reciprocal of the absolute risk increase.

There is no magic cut-off value for NNT or NNH, although intuitively low values of NNT and high values of NNH are considered better. Balancing the NNT and NNH will to a large extent be mediated through consideration of the relative importance of the diseases prevented or produced, as discussed earlier.

Assuming the relationships between aspirin use and the outcomes studied to be causal, the paper by Chan et al. suggests that about 12,500 patients (women between the ages of 40–53) needed to be treated to prevent one case of colorectal cancer whereas the meta-analysis by Cole et al. (Chapter 5) reports an absolute risk reduction of 6.7 % equating to only 15 patients (with history of sporadic colorectal adenoma or previous colorectal cancer) who needed to be treated to prevent one case of colorectal adenoma. This difference in NNTs highlights the importance in defining the target population when considering the NNT/NNH of a particular intervention. One may be more inclined to introduce an intervention with a higher NNH if the intervention were to be targeted at a group of patients already at high risk with a lower NNT, for example

those with a recent history of colorectal adenoma as per the Cole paper, as opposed to introducing the treatment to a broader population at any risk, akin to the Chan paper.

Population attributable fraction

This measure is not directly applicable to the relationship between aspirin and colorectal cancer, but is important when considering the removal of harmful exposures. The population attributable fraction is the proportion of all cases in the whole population that may be attributed to the exposure, that is, if the exposure were to be eliminated in its entirety the reduction in the number of cases of the disease in the whole population. This assumes a causal relationship between the exposure and the disease in question. It is therefore entirely possible that an association with a high relative or absolute risk in a specific population will be of minimal importance in the population at large if the exposure is rare. Again, the target population of a given intervention becomes crucial when assessing its relative importance.

Cost implications

Of course our ability to influence health outcomes is often limited not by science, but by finance. Although a discussion of the economics of health care is beyond the scope of this chapter, we would recommend that it be given consideration and a good starting point for this includes the online chapter accompanying this book (see Online Chapter 8 on "Health economics").

Conclusion

It is not possible within a chapter on assessing the meaning of medical literature in general to give a clearly argued case on the literature pertaining to one individual question or group of questions. However, this section of the textbook has focused on papers in one area – the potential for aspirin to prevent colorectal neoplasia. We hope that after reading the chapter and those before it you are able to return to the wider literature on this subject with a new perspective. It should be easy to see that there is a wealth of evidence suggesting that aspirin can indeed prevent colorectal neoplasia in some, but that for most people the chance of receiving this benefit is likely to be small. Whether such use is appropriate for an individual will therefore remain a case of weighing the chance and importance of benefit against the chance and severity of potential harms for that individual.

References

1 Guyatt G, Rennie D. (2002) *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*, American Medical Association, Chicago, IL, p. xxiii.

2 Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005;95(Suppl 1):S144–50.

3 Rothman KJ, Greenland S, Lash TL. (2008) *Modern Epidemiology*, 3rd ed., Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, PA, p. x.

4 Cole BF, Logan RF, Halabi S, et al. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. *J Natl Cancer Inst* 2009;101(4):256–66.

5 Cook NR, Lee IM, Gaziano JM, et al. Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. *JAMA* 2005;294(1):47–55.

6 Chan AT, Giovannucci EL, Meyerhardt JA, et al. Long-term use of aspirin and nonsteroidal anti-inflammatory drugs and risk of colorectal cancer. *JAMA* 2005;294(8):914–23.

7 Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58: 295–300.

8 Langman MJ, Cheng KK, Gilman EA, Lancashire RJ. Effect of anti-inflammatory drugs on overall risk of common cancer: case-control study in general practice research database. *BMJ* 2000;320(7250):1642–6.

9 Kamper-Jorgensen M, Grønbaek M, Tolstrup J, Becker U. Alcohol and cirrhosis: dose–response or threshold effect? *J Hepatol* 2004;41(1):25–30.

10 Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ* 2003;327(7417):741–4.

11 Pugh RN, Murray-Lyon IM, Dawson JL, et al. Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg* 1973;60(8): 646–9.

12 Dukes CE, Bussey HJ. The spread of rectal cancer and its effect on prognosis. *Br J Cancer* 1958;12(3):309–20.

13 Porta MS, Greenland S, Last JM (2008) *A Dictionary of Epidemiology*, 5th ed. (edited for the International Epidemiological Association by M Porta; associate eds. S Greenland, JM Last), Oxford University Press, Oxford, xxiv.

PART THREE

How to Do Clinical Research in GI

7 How to develop and validate a GI questionnaire

Enrique Rey[1] *& G. Richard Locke III*[2]

[1]Functional GI Disorders Unit, Division of Digestive Diseases, Hospital Clinico San Carlos, Universidad Complutense, Madrid, Spain

[2]GI Epidemiology/Outcomes Unit, Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

Key points

- Questionnaires are measurement instruments and should be evaluated prior to use in research or clinic.
- Questionnaires should be assessed for reliability and validity.
- Reliability is about precision (answers are the same in repeated measures) and validity is about accuracy (answers represent the truth).

Why do we need questionnaires?

Medicine was born upon binary endpoints (relief/not relief) but it jumped a big step, especially in the twentieth century, with the incorporation of quantitative and objective endpoints (measurement and monitoring of biologic markers of diseases). Their importance in practice is obvious for all clinicians and researchers; a case of diabetes is defined by his/her glycemia and the disease is perfectly controlled if glycosylated hemoglobin is within the normal range. However, elements of the patient–physician relationship and research information may not be easily quantified with "objective biologic markers"; diabetes may be controlled but it does not necessarily mean that the patient with diabetes feels well. Moreover, the World Health Organization's definition of health emphasizes the relevance of these elements. Questions like "how are you doing?" are common in clinic and research. Despite their utility for the clinical evaluation of a single patient, sometimes, especially in the research scenario, it is hard to know how poorly a patient is doing, whether they are doing better or worse, or whether one subject is doing better or worse than another. In addition, diseases often express themselves as symptoms, which may have key importance in those diseases without an objective marker. For example, the diagnosis of irritable bowel syndrome relies only on symptoms, and therefore depends upon the subject's ability to communicate their symptoms and the physician's (or researcher's) ability to understand them correctly. Moreover, research (and to some extent also clinical practice) requires a homogeneous set of information from subjects in order to allow comparisons, or to investigate potential intervening factors [1]. For example, one may sometimes wonder whether a patient has truly developed a symptom because of an intervention or whether it was present prior to the intervention but was not expressed or not asked about; "unknown" is not the same as "not present".

Subjective elements in clinical practice comprise:

1 Communication of subjective experiences (symptoms, wellbeing, perceptions, cognitions);

2 Comparison of subjective experiences, both within the same subject and between subjects;

3 Homogenization of information collected.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

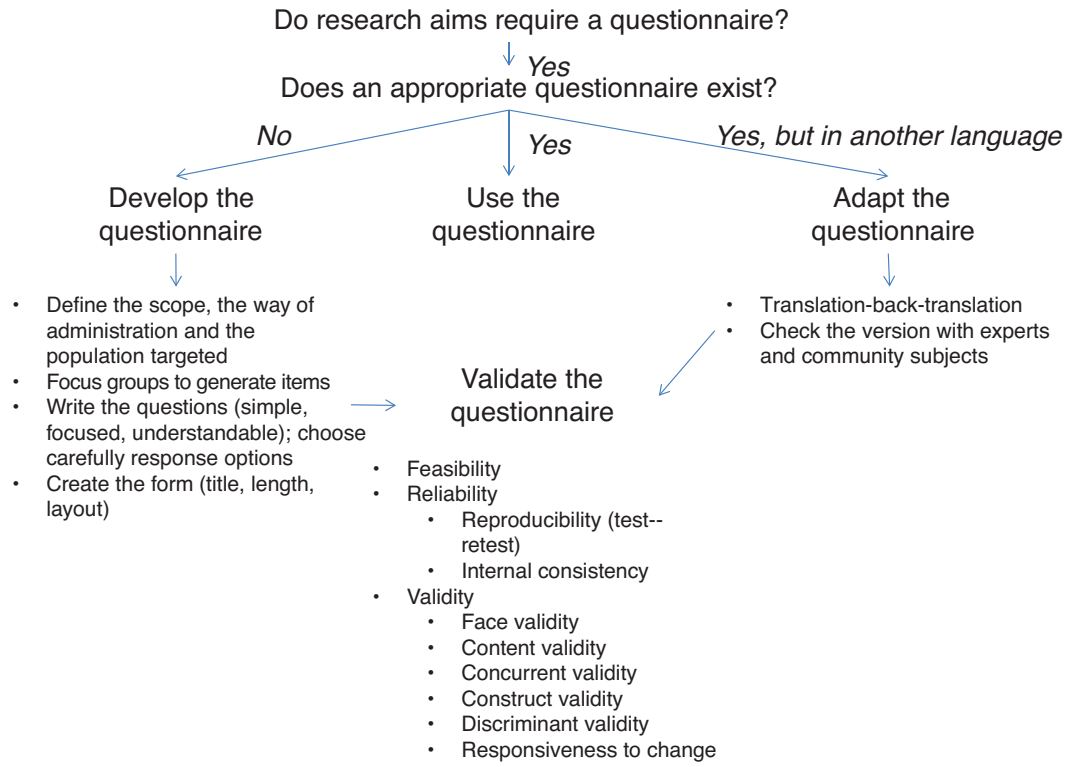


Figure 7.1 Schematic representation of the development and validation processes of a questionnaire.

The only way to disclose this subjective information is direct questioning of the source of information: the subject. However, many factors can interfere with this process: how questions are done, how answers are interpreted, what circumstances surround the communication process (e.g. the mood of the interviewer) can change moment by moment. It is not only important to make the question, but also how it is done; for example, a subject could be asked "How are you doing? or "You are probably doing well, aren't you?"; it is obvious that an affirmative answer is more likely after the second question; similarly, a negative answer to the question "Do you suffer pyrosis?" does not necessarily mean it is true, because the subject may simply answer "No" because they do not understand the word "pyrosis". Questionnaires are intended to measure as objectively as possible these subjective elements and to overcome, as much as possible, these limitations. Questionnaires are measurement instruments of these subjective experiences, and should be viewed as such. Like a glucometer, questionnaires should be "manufactured" and "subjected to quality control"(see Figure 7.1). "Manufacturing" a questionnaire is known as development or, when appropriate, adaptation, while "quality control" is known as validation.

Do we need to develop a questionnaire?

The first question when facing a particular study or a clinical protocol is whether there is a need to develop a new questionnaire. Development of a questionnaire requires specific knowledge, can take months to years to perform, and can be quite expensive. In the last three decades, an incredible number of questionnaires have been developed for specific purposes. Vast libraries of questionnaires are available via the Internet, summarizing information on the features of each questionnaire. In general, it is more efficient to use something already developed, even it is not absolutely perfect, for the study at hand.

Research in non-English-speaking populations represents a particular situation. It may be difficult to find a questionnaire developed in the native language fitting the research aims, since most questionnaires are developed in English. Researchers must then decide between developing a new instrument, or adapting an already available questionnaire. In general, the use of a questionnaire already developed is more efficient, although an adaptation process should be undertaken. This approach possesses several advantages: it is less time-consuming, the performance of the questionnaire is already known, and transcultural comparisons of the results are then possible.

Development of a questionnaire

If a new questionnaire is deemed necessary for a particular study, a sequential process of development should be undertaken, including determining the scope, deciding the survey method, and writing the questions [2].

Determining the scope

The first step in developing a questionnaire is quite obvious, but at the same time, very important: what needs to be measured and what are the appropriate topics? To answer these questions, a deep review of the literature helps one to identify what is already known. However, little information is found in many instances and other sources should be utilized. Expert opinion is always advisable in all steps of a questionnaire's development, but specifically for this purpose, focus groups are also very useful [3]. Focus groups are small groups of persons (5–15) belonging to the population intended to be studied who are individually interviewed to collect in-depth responses to general questions on the subject of the questionnaire [4]. One can then delineate which topics should be covered by a questionnaire from the analysis of their answers.

Survey method

The mode of administration of a questionnaire is a crucial step in its development. There are many ways to administer a questionnaire, each with their advantages and disadvantages, but, in any case, the questionnaire should fit the mode of administration [5]. For example, face-to-face interviews are good for asking open-ended questions and following up with additional questions based on specific responses. However, when human beings interact, they usually present their most positive light, so socially undesirable answers are less likely [6]; for example, one may anticipate that acknowledgement of problems with fecal incontinence would be less likely in a face-to-face interview than in an anonymized self-administered questionnaire. In addition, interviewers should be trained to administer the questionnaire in a structured manner. All these issues could lead to a higher cost. For these reasons, the majority of medical studies use self-administered questionnaires; these questionnaires can be circulated by mail, telephone or online, but they require a careful design of the questions and the possible answers.

Writing the questions

Although writing a question seems very easy, it is a challenge, and a great part of the success of the research will depend on it. Each questionnaire item should be simple (address only one item at a time), focused (intended to obtain the desired information), and clear (understandable by the interviewed person) [7]. For example, a question like "do you suffer heartburn or acid regurgitation?" is a bad question, because it includes two items at the same time (heartburn and acid regurgitation), it is not focused (it is unclear whether the response is "either/or" or "both"), and it may be not understandable (the terms heartburn and acid regurgitation are not understood by everyone) [8]. Usually, a fourth grade reading level is recommended and this should be checked with an expert or by using dedicated software.

The response options are also important. The first decision is whether the questions should be open or closed. Open-ended questions are easier to write, permit subjects to answer in their own words, and may provide richer information. However, analysis is extremely difficult in most situations, owing to a time-consuming content analysis, limiting their utility. In general, most questionnaires use closed-ended questions because response choices can be anticipated and it is easier to analyze and report the results.

Closed-ended questions are more difficult to write because the success in obtaining the relevant information depends on the options provided. Sometimes, nonquantifiable information is required; in this case,

the choices are a binary response (yes/no; agree/disagree; male/female . . .), and categorical nonordered response (married/single/widowed/divorced), although the second case is just a summary of several yes/no questions. When the desired information is quantitative, ordinal scales (Likert scales) or visual-analog scales can be used. Data-entry is easy in the first case and time-consuming (and therefore expensive) in the second case. Nonetheless, the level of detail pursued is the key issue in selecting and designing quantitative responses, specifically in designing ordinal scales.

There are several general rules that should be kept in mind when constructing the response options. First, positive and negative options should be balanced. For example, if when asking about general wellbeing (how do you feel?), results would not be the same if the options are "very bad/bad/slightly bad/good" or "bad/slightly bad/slightly good/good". Secondly, vague or subjective categories should be avoided. For example, when asking about alcohol intake (how much do you drink?), a list of options like "a little bit/not too much/like any other/more that I should" is completely dependent on the subjective concept of how much alcohol intake is usual or appropriate. Third, responses categories should be mutually exclusive. For example, if when asking about the location of abdominal pain, the individual can be provided with the options of upper (above the navel) or lower (below the navel), but persons with upper and lower abdominal pain may be confused about how they should answer. One solution for these questions is using "check all that apply" options, but the authors should then consider that it will force an analysis considering each response option as a yes/no question.

The final step in writing the questionnaire is the creation of the form, which is especially relevant in questionnaires intended as postal surveys. The order of the questions is important for the success of a questionnaire; as a general rule, key questions should be placed first and general questions (e.g. demographics) at the end of the questionnaire. Also, questions directed to activities, habits or any issue that persons may be reluctant to divulge (income, sexual activity, use of recreational drugs) should be managed carefully, and placed in the last part of the questionnaire, where persons may feel more comfortable answering them. The time spent on other details of the questionnaire, such as title and lay-out, is usually well-invested [9]. For example, it has been shown that the title of the questionnaire may lead to a difference in response rate of up to 10 % [10].

Adapting a questionnaire

Adaptation of an already developed questionnaire is usually the best option in epidemiologic research in non-English-speaking populations. Although adaptation of a questionnaire is easier than developing a new one, the process is more complicated than just translating it [11]. Nonetheless, translation of the questionnaire is the initial step, and it is usually done by the translation and back-translation process. First, the questionnaire is translated into the target language by one or two natives with a fluent knowledge of the original language (usually English); translators should focus on the usual wording and sentence construction of the target language and adapt it to the cultural environment (when appropriate) whilst trying to keep the meaning of the original questions, rather than producing an exact translation [12]. The translation is then back-translated to the original language by a native. After this process, all participants meet to evaluate if the back-translated questionnaire has kept the original meaning and intent of the questionnaire. Sometimes, this preliminary version of the questionnaire is presented to an expert (physician, linguistic expert, expert in education) in order to obtain feedback as to how it could be improved. If this process is successful, questions should then be presented to a small group of subjects, representative of the population, to assess if questions are understandable and whether the answers are appropriate. As with a newly developed questionnaire, it is advisable to undertake a validation process with the adapted version of the questionnaire.

Validation of a questionnaire

Whether the questionnaire has been newly developed or adapted it should be validated, which is a way of measuring how much it can be trusted. The two key concepts are reliability and validity, but a prior and necessary step is feasibility [13–15].

Feasibility

The first condition for a questionnaire to work properly is whether or not it is feasible to use. Usually,

it is easy to test the feasibility of a questionnaire. It should be administered to a small group to be sure they can actually complete it without confusion. The time to completion is measured in order to assess the responder burden as this will affect response rates. The questionnaire should be reviewed to assess for missed questions or blank answers [16]. The person can be interviewed to assess for misunderstood questions. Often questionnaires have "go to" directions that allow people to skip questions. The issue is whether those instructions are easily understood.

Several additional aspects should be kept in mind while testing feasibility. First, feasibility testing should fit the method of administration; for example, some questions may be difficult to understand in self-administered questionnaires making them unfeasible, but the same questions are feasible in face-to-face interviews, when a well-trained interviewer can resolve doubts. Secondly, the population targeted should be considered; for example, a questionnaire intended for use with the very elderly should not be tested on the general population; cognitive and physical abilities are not the same and even issues like font size may alter what is feasible.

The goal is to make questionnaires measure patient-reported outcomes with the accuracy and precision that are expected of any other measure in medicine. In psychometrics, the terms validity and reliability are used in place of accuracy and precision. Once it is known that the questionnaire is feasible, whether or not it is reliable and valid should be tested.

Reliability

A questionnaire is reliable if it provides the same result on repeated measurement under stable conditions (precision). There are several types of reliability, but test–retest reliability is commonly evaluated. Test–retest reliability can be measured by administering the questionnaire on two separate occasions close together [17]. The goal is to make the interval long enough so that the respondents cannot remember their answers, but short enough that no change in their condition will have occurred. These two answers can then be compared. However, one must remember that high levels of agreement can occur by random chance if a question measures an issue for which just 5 % of people say "yes" then 95 % will say "no." This is like flipping a coin that has only a 5 % chance of heads. If flipped a second time, the coin will give the same answer 90 % of the time [$(0.05 \times 0.05) + (0.95 \times 0.95)$]. A high level of agreement has occurred just by random chance. Therefore a different standard is needed to decide if this question is reliable. The kappa statistic is a chance-corrected measure of agreement that is used to assess such dichotomous (e.g. yes/no type) answers. A weighted kappa statistic can be calculated for multilevel responses. This statistic is calculated as:

$$(\text{Observed} - \text{Expected Agreement})/(1 - \text{Expected Agreement})$$

A value of 0.4 or above is satisfactory. Kappa statistics will be low for infrequent answers; questions that have closer to 50–50 responses will have better results.

The kappa statistic is useful to test reliability of questions with nominal answers, but it is not adequate to test consistency in numerical answers. In that case, the responder to the questions is considered to be an "observer", and one should test as though estimating a numerical measurement. For example, a question that refers to the severity of a symptom with seven options ordered from no symptoms to unbearable can be viewed as a numerical estimation. In this case, the appropriate statistic for evaluating test–retest reliability is the intraclass correlation coefficient, which takes into account not only random error (like the ordinary correlation coefficient) but also systematic error (e.g. tendency to score higher than others).

The other type of reliability is internal consistency. The agreement among different items that measure the same thing within one questionnaire can be tested. This approach is most useful in longer questionnaires; otherwise the respondents will clearly notice that they are being asked the same thing twice. The correlation between the answers to these similar questions can then be assessed, using Cronbach's alpha statistic.

Validity

A questionnaire is valid if it measures what it intends to measure (accuracy). Validity has several forms and is best thought of as a continuum – that is, a measure is not simply valid or invalid, but rather has a degree of validation. One should not ask whether the questionnaire is valid, but rather, how valid is it? Face

validity, content validity, criterion validity, discriminant validity, construct validity, and responsiveness are all features that an instrument can have, and these issues need to be addressed as the instrument is developed and tested.

Face validity represents the degree to which a questionnaire appears to be measuring what it is supposed to measure. It is a rather simple concept: does the questionnaire look valid to an expert? Often this is the only level of validity that a questionnaire has; people just make up questions to suit the purpose of the study.

The next level is content validity; content validity refers to the degree to which the items of the questionnaire are representative of the characteristics being investigated. In other words: does the questionnaire measure the appropriate issues? For example, after wording a questionnaire intended to measure digestive symptoms, one should think: Does a reflux questionnaire measure heartburn and acid regurgitation? Does it measure chest pain, dysphagia, dyspepsia and respiratory symptoms? Does a bowel questionnaire measure all the elements of diarrhea and constipation? Content validity is also typically determined by panels of experts.

Therefore, in a practical setting, face and content validity are usually done as the first step after constructing the questionnaire, no matter how expert a research group is in developing questionnaires. The feedback from a panel of experts may warrant: (i) an external evaluation of the work that was done, sometimes providing improvements; (ii) an evaluation before undergoing further validation studies.

The next level is concurrent validity, also called criterion validity, which is the ability of the questionnaire to identify a characteristic known to be associated with the characteristic intended to be measured. In essence, it represents how well a new questionnaire compares with the gold standard. The gold standard in this case should not be confused with the diagnostic gold standard in clinic. For many symptom surveys, the gold standard is a face-to-face physician interview. In some situations the gold standard may be a long survey from which a newer shorter version has been derived. For example, the sickness impact profile (SIP) has over 100 questions and was later modified to create the Short Form-36 (SF-36), which in turn was further modified to give the Short Form-12 (SF-12). However, often no gold standard exists. The new questionnaire is designed to be better than existing instruments. This makes assessing concurrent validity much more difficult. What is the truth? In this case multiple measures can be used to develop a consensus definition of the truth and then the new questionnaire can be compared to that standard.

Perhaps the most difficult concept to understand is construct validity. In this case the designer develops a construct of how the questionnaire should perform. Based on this construct some patients should have worse scores than others; and then the survey is administered to see if this holds true. The questionnaire needs to perform in a predictable fashion.

Discriminant validity refers to the concept that a questionnaire should be able to identify distinct groups. Can the survey responses distinguish one group from another? Responsiveness is yet another attribute a questionnaire should have; especially one that will be used in a clinical trial. For instance, is the questionnaire sensitive to a change in symptoms? The challenge to testing responsiveness is that there must be an intervention or something that leads to a change in the person's condition. Typically, responsiveness is assessed in the setting of a clinical trial. Most responsive questionnaires must be developed in one clinical trial before they can be used to assess the effects of therapy in another.

Not all questionnaires need to be tested and shown to fulfill all these forms of validity. In some situations a discriminant questionnaire is needed; for others the questionnaire needs to be responsive. The questions needed to address these two tasks may be quite different. At a minimum, questionnaires must have face and content validity. However, an assessment of concurrent validity is preferred.

Using questionnaires in research

Developing a questionnaire is a hard task, as detailed in earlier sections. However, one should realize that questionnaire development is just a path to a research objective. The questionnaire is only the instrument. For example, it may be hypothesized that sympathetic/parasympathetic imbalance is the reason behind dizziness in patients with IBS; a machine may be developed that is able to monitor these parameters during a one-week period but which requires subcutaneous implantation. Its accuracy and precision could be tested in some patients; however, it is unlikely that

it could be used in a long clinical study because most patients will not accept its implantation just for a study; and therefore even though this may be the best instrument, it will be useless for this purpose. Similarly, a good questionnaire requires a well-designed development and validation process but it should also fit the research needs. For example, a long 16-page validated questionnaire designed for community postal surveys will be useless if most subjects in a specific population will not respond to it; however, the same questionnaire in the same population may be useful if it is administered by telephone.

How the questionnaire will be used is an important issue that should guide the process of development and validation, always keeping in mind the targeted population and the way it is going to be used.

Conclusions

Many physicians and researchers conducting clinical studies do not need to know all these details of questionnaire development. The key messages of this chapter are (i) a questionnaire is a measurement instrument, and like any other, should provide reliable results; (ii) making up the questions is not enough – when using a questionnaire, researchers need to be sure that the instrument they are using was developed rigorously.

Fortunately, well-validated measures exist to assess most GI symptoms and diseases. Additional measures exist to assess quality of life, physical functioning, work productivity, and other outcomes. A catalog of these is beyond the scope of this chapter but such instruments can easily be found. Sometimes the actual questionnaires are published as appendices to a journal article. This places them in the public domain and available for use. Most often the authors need to be contacted for permission. Some questionnaires are proprietary and can only be used for a fee.

The goal of this chapter was to familiarize the reader with the steps required to develop and test a GI questionnaire rigorously. Alternatively, one can and should search the literature for existing measures and write to the authors to obtain permission for their use. By using well-tested measures the physician and researcher can be assured that they know what is being measured when they ask people "How are you doing?"

Multiple choice questions

1 A questionnaire that provides the same result when it is administered repeatedly over a short period of time is said to have high: (choose one)

A Reliability
B Face validity
C Content validity
D Responsiveness

2 The concepts of reliability and validity are similar to the concepts of accuracy and precision, which one of these two statements is correct?

A Reliability is similar in concept to accuracy; validity is similar in concept to precision.
B Reliability is similar in concept to precision; validity is similar in concept to accuracy.

3 Disease-specific quality of life measures offer what advantage over generic measures? (choose one)

A Ability to compare across conditions
B Rigorous testing in thousands of patients over time
C Sensitivity to change in a patient's condition
D Ability to compare to community norms

4 If you want to measure change over time in a clinical trial, above all else you need a questionnaire that is: (choose one)

A Reliable
B Valid
C Responsive
D Discriminative

5 Many studies have used the medical outcomes study Short Form-36. What type of instrument is this? (choose one)

A A disease-specific health utility measure
B A disease-specific health status measure
C A generic health utility measure
D A generic health status measure

General reference

Dillman DA (2007) *Mail and Internet Surveys*, 2nd edn (ed. DA Dillman), John Wiley & Sons, Inc., New York.

References

1 Testa MA, Simonson DC. Assesment of quality-of-life outcomes. *New Engl J Med* 1996;334(13):835–40.

2 Stone DH. Design a questionnaire. *BMJ* 1993;307(6914): 1264–6.
3 Juniper EF, Guyatt GH, Streiner DL, et al. Clinical impact versus factor analysis for quality of life questionnaire construction. *J Clin Epidemiol* 1997;50(3):233–8.
4 Powell RA, Single HM. Focus groups. *Int J Qual Health Care* 1996;8(5):499–504.
5 Edwards P, Roberts I, Clarke M, et al. Methods to increase response rates to postal questionnaires. *Cochrane Database Syst Rev* 2007;(2):MR000008.
6 Okamoto K, Ohsuka K, Shiraishi T, et al. Comparability of epidemiological information between self- and interviewer-administered questionnaires. *J Clin Epidemiol* 2002;55(5):505–11.
7 Niederhauser VP. Seven strategies for successful surveys. *J Pediatr Health Care* 2006;20(3):210–13.
8 Locke GR, Talley NJ, Weaver AL, et al. A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc* 1994;69(6):539–47.
9 Beebe TJ, Rey E, Ziegenfuss JY, et al. Shortening a survey and using alternative forms of prenotification: impact on response rate and quality. *BMC Med Res Methodol* 2010;10:50.
10 Lund E, Gram IT. Response rate according to title and length of questionnaire. *Scand J Soc Med* 1998;26(2):154–60.
11 Moreno Elola-Olaso C, Rey E, Rodriguez-Artalejo F, et al. Adaptation and validation of a gastroesophageal reflux questionnaire for use on a Spanish population. *Rev Esp Enferm Dig* 2002;94(12):745–58.
12 Almansa C, Garcia-Sanchez R, Barcelo M, et al. Translation, cultural adaptation and validation of a Spanish version of the Irritable Bowel Syndrome Severity Score. *Rev Esp Enferm Dig* 2012;103(12):612–18.
13 Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008;65(23):2276–84.
14 Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11(3):193–205.
15 Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118(8): 622–9.
16 Smeeth L, Fletcher AE, Stirling S, et al. Randomised comparison of three methods of administering a screening questionnaire to elderly people: findings from the MRC trial of the assessment and management of older people in the community. *BMJ* 2001;323(7326): 1403–7.
17 Rey E, Locke GR, III, Jung HK, et al. Measurement of abdominal symptoms by validated questionnaire: a 3-month recall timeframe as recommended by Rome III is not superior to a 1-year recall timeframe. *Aliment Pharmacol Ther* 2010;31(11):1237–47.

Answers to multiple choice questions

1. A
2. B
3. C
4. C
5. D

8 How to do population-based studies and survey research

Sanjiv Mahadeva[1] *& Hematram Yadav*[2]

[1]Division of Gastroenterology, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

[2]Department of Community Medicine, Faculty of Medicine, International Medical University, Kuala Lumpur, Malaysia

Key points

- Population-based studies can be either observational (descriptive) or analytical (cohort, case-control, cross-sectional, or longitudinal). Identifying the correct community/residential area and using an appropriate sampling method is necessary.
- Data collection is usually made through a structured questionnaire.
- The survey method used to collect data should be tailored to the study population, with established methods such as face-to-face interview, postal survey, telephone interview, or via electronic mail.

Introduction

Population-based studies, as opposed to hospital-based studies, offer the only means of estimating the true prevalence or incidence, and determining the epidemiology, of gastrointestinal (GI) or any other chronic diseases that affect a particular community. Information obtained from population-based studies has wide implications. For example, knowledge regarding the prevalence or incidence of a particular disease will enable healthcare providers to budget and plan for service provision at both primary and secondary care levels. Furthermore, increased recognition of the burden of a particular illness to both employers and healthcare providers can facilitate enhanced resources in tackling such a condition in order to improve the overall management of this condition.

Several methods of conducting population-based medical research are available, ranging from questionnaire-based surveys to detailed physical examinations and simple laboratory investigations. As the aim of population-based research is to obtain as much information from as many individuals as possible, surveys have become a popular means of obtaining this data.

Population-based surveys

Large-scale population-based surveys can be expensive to undertake. Hence, prior planning is an essential component to any such study. Some of the basic issues that need to be addressed before embarking on a population-based survey are:

- a clearly defined research question
- type of study – e.g. observational (descriptive) or analytical (see later)
- study area (location and type of community to be studied)
- a clearly defined sampling frame
- sample size calculation – this will depend on whether a disease is rare or common

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

- type of instrument used to collect data
- timing of survey
- logistics to be calculated
  - amount of staff required to collect data
  - funding available to conduct the study.

Research question and aim of the study

One of the major pitfalls in population-based surveys is that either the research question or the aim of the study is not clearly defined. Stating these early on will enable the researcher to be clear in selecting the type of methodology (study design), as well as anticipating the types of statistics that need to be applied to the data collected, so that the research question can be answered.

Location of survey and engaging the population of study

Theoretically, population-based samples can be obtained from various places where large sections of the community reside, such as places of employment, schools, community centers, or even shopping malls. However, data obtained from such "convenience" samples are usually biased in their selection. Unless the aim of the survey is to target a particular section of the community, for example individuals of younger age in schools or adults in employment in offices, specific locations such as these are not desirable due to the obvious bias of the type of individuals who will be residing in these locations. To obtain data from a more representative sample of the population in general, the household is usually utilized as the commonest sampling unit.

Depending on the region and part of the world where the study is conducted, it is usually necessary to seek and obtain permission from an appropriate administrative authority responsible for the community before commencing the survey. In a city or urban area, this will normally be the local town or city council, but this may vary in rural areas in various parts of the globe.

Apart from administrative authorities, it is generally advisable to engage the community leaders or members of the residential association. Spending some time and effort in gaining the trust of such members of the community will greatly enhance cooperation for the study when it commences. In addition to seeking approval, as some surveys can be time-consuming and inconvenience people in their daily routines, it may be advisable to reward individuals for their participation in the survey. Depending on the study's budget, the reward can range from anything as simple as a token (e.g. towel, stationery, and so on), to offering a free screening check-up at the nearest health center. This can be invaluable in maximizing participation in a community survey, particularly in urban areas, where most individuals lead extremely busy lifestyles.

Sampling

It is more economical to study a sample of, rather than the whole, population but this sample should be representative of the entire population, at least from a sociodemographic perspective. Sampling is a technique where a group of subjects from a larger population are selected. In sampling, each person or household should have an equal chance of being selected from the sample. Once an ideal community that is representative of the entire population has been selected, a sampling frame has to be chosen. The sampling frame refers to a listing of the members of the population from which the sample for the study is to be drawn. For example, if one wanted to study the incidence of dyspepsia amongst adults in a population, then the sampling frame would be all the adults in the community under study. There are several accepted sampling methods that are commonly used [1]:

A *Simple random sample*

In this method, a number is usually assigned to each household in the sampling frame. Using a table of random numbers, households are randomly selected for inclusion in the study. This method ensures appropriate randomization and provides the greatest number of possible samples. Also, each household has an equal chance of being selected in the sample, thus minimizing bias in the study.

B *Systematic random sample*

Using this method, numbered households are selected at regular intervals. For example, a randomly numbered house from 1 to 10 is chosen, and every fifth household after that particular numbered household is selected for the survey. This method enables each unit within a sampling frame to be selected. However, the total number of samples will be less than the random sampling method in (A), as only a fixed number of houses can be selected in a sequential order.

C *Stratified random sample*

This method of sampling is usually used when trying to capture data from different groups within the community. For example, in a multiethnic community, the ethnic groups may not be equally distributed within a particular residential area. Hence, based on available data (perhaps from the residents association or the local council) a sample is deliberately drawn to capture equal areas of representation of the various ethnic groups within a particular residential area. For example, if the population has a ratio of Chinese to Indians of 4:1, then the population can be stratified into two groups (Chinese and Indians) and individuals from the two ethnicities can be selected randomly in that ratio.

Whilst a sample taken from any population is thought to best represent the entire population, sampling error is inevitable, due to possible differences between different samples from the same population. From a practical perspective, the only means of reducing sample error would be by enrolling as large a sample size as possible in the study.

Data collection in a survey

There are several methods by which data can be obtained in a population-based study. Most investigators use a structured questionnaire.

Questionnaires in surveys

When designing a questionnaire that will be applied in the community, a balance has to be obtained between obtaining as much data as possible, and not making the questionnaire too long, such that potential participants in the survey become either bored or irritated. Details of how to develop such questionnaires are covered in Chapter 7. When using a questionnaire that has been developed in a language foreign to the population of study (e.g. using an English questionnaire in a rural population in Asia), it is vital that a proper process of translation and validation is performed before conducting the study. Prior to using the questionnaire in the population, a *pilot study* among a convenient sample of subjects, using the various language versions of the questionnaire should be conducted. In this pilot study, the investigator has to determine how easily the questionnaire is understood, and the average duration taken to complete the entire questionnaire. The latter information is vital when planning the details and total duration of study when the survey commences in the community. Any modifications and subsequent re-testing of the questionnaire should be conducted at this stage, before applying the questionnaire in the community to avoid major problems later.

Methods of data collection

Several methods can be employed to administer a questionnaire in a survey. In general, the method utilized depends on the level of development of a particular society and its communication network. All methods have their pros and cons.

Direct interview. A direct, or face-to-face, interview method is usually employed in less developed countries. This method is labor-intensive, requiring a set of paid data collectors who have to be trained to ensure familiarity with the questionnaire, and to ensure that all interviews are conducted in a standardized manner.

Advantages of this method are:

- any queries about the survey questionnaire items can be resolved immediately
- individuals with lower educational levels are more likely to complete the questionnaire appropriately
- there is generally a greater participation rate among respondents, due to a sense of obligation to, or trust gained by evidence of official documentation from, data collectors.

Disadvantages of this method are:

- the total sample size is usually smaller, as data collectors require funding, which may be limited
- there is a lack of anonymity among respondents, which may affect the accuracy of data collected.

Postal survey. Mailed questionnaires are usually utilized in a country with an established and extensive postal network.

Advantages of this method are:

- it is less labor-intensive, and cheaper
- a larger number of the population can be reached.

Disadvantages of this method are:

- a lot of motivation from participants is required to complete and mail back questionnaires. Hence, the yield tends to be lower using this method;
- queries about questionnaire items cannot be resolved, and may lead to inaccurate data;
- a higher level of literacy is required from respondents. Hence, data from individuals with lower education levels may be missed in this survey method.

Telephone interview. In countries with an extensive telecommunications network, this method of interview may be employed. Once again, trained data collectors will need to be employed to facilitate this type of study.

Advantages of this method are:

- a larger number of the population can be reached
- any queries about the survey questionnaire items can be resolved immediately
- data collectors do not have to physically go out into the field to collect data
- repeat calls for individuals missed earlier can be made without difficulty.

Disadvantages of this method are:

- participation rate may be variable as the trust gained from respondents may be less.

E-mail/Internet survey. In the current technological era, with increased availability of personal electronic mails (e-mails) for many individuals in various parts of the world, communication via the Internet has become another method of conducting surveys.

Advantages of this method are:

- rapid dispersion of questionnaires can be performed within a short duration of time
- no need for paper and stationery
- data from questionnaires completed can be stored without difficulty
- data analysis from electronic forms may be performed faster
- minimal funding is required provided the Internet is readily available.

Disadvantages of this method are:

- limited to populations that have readily available Internet access
- e-mail addresses within a given population may not be readily available
- motivation from participants is required to complete and e-mail back questionnaires, although this may be easier than via normal post.

Examples from the GI literature

In Malaysia, Mahadeva et al. [2] conducted a population survey in a rural community to determine the prevalence of dyspepsia using a *face-to-face direct interview* method. A total of 2260 adults from 1642 households, identified by a systematic random sampling method, were approached by trained data collectors and invited to participate in the study. Two thousand (88.5 %) individuals completed the survey in this study.

In Hong Kong, Hu et al. undertook a population-based survey to investigate the prevalence of GI symptoms, and their association with healthcare-seeking behavior, using a *telephone interview method* [3]. Random telephone numbers were generated by computer and dialed automatically. Only numbers corresponding to ethnic Chinese households were used in the study. Office numbers, facsimile machines, and non-Chinese households were excluded. The telephone interview was conducted by a team of 15 trained telephone interviewers. A total of 2640 individuals were contacted, and 1649 (62 % response rate) were able to complete the survey.

In the United Kingdom, Ford et al. conducted a follow-up *postal survey* to examine the effect of quality of life on subsequent development of dyspepsia in the community [4]. From an original cohort of 8407 individuals, 6416 could be traced 10 years later and were mailed a set of GI and demographic questionnaires. Of the 8407 individuals originally included, 4003 (48 %) responded by completing and mailing back the questionnaires.

In Japan, Hongo et al. conducted a population-based survey of GI symptoms using an *electronic internet survey* [5]. Registered members of a survey organization were selected as the sample, and a stratified random sampling method was applied to obtain equal distributions of age groups, gender, and residential area. In the survey, 6000 questionnaires were e-mailed out, and 2125 (35 % response rate) individuals returned a completed questionnaire.

Design of population-based studies

Thus far, we have described the "nuts and bolts" of undertaking a population-based study. One of the most important steps in getting the study done will be deciding on the research strategy or the study design of the population survey. The study design of most population surveys can be broadly categorized as either descriptive or analytical [6].

Descriptive design

This type of study design is aimed at generating hypotheses, rather than testing them. Typical examples include clinical or demographic features of a

particular disease in the community or community-based disease registries.

Analytical designs

These types of study design are usually aimed at testing hypotheses. There are several types of analytical designs and we will elaborate on these further:

- Cohort study – prospective and retrospective
- Case-control study
- Cross-sectional study
- Follow-up studies – longitudinal or cross-sectional at repeated intervals.

Cohort studies

In the cohort design, groups of both exposed and non-exposed individuals in the population are recruited, and then followed up prospectively for a period of time. The development of the disease of interest, also known as the incidence, is then compared between exposed and nonexposed individuals (Figure 8.1). The selection of a study population can be on the basis of exposure to a particular risk factor, or by selecting a defined population before individuals are exposed. It is generally expected that the group with exposure to the risk factor will have a higher incidence of the disease compared with the group with no exposure.

Example from the GI literature

In Japan, Uemura et al. conducted a prospective cohort study to investigate whether *Helicobacter pylori* had a direct causal link with gastric cancer [7]. Among 1526 patients with various benign upper GI conditions diagnosed by endoscopy, 1246 had *H. pylori* infection and 280 did not. After a mean duration of follow-up of 7.8 years, 36 (2.9 %) patients with *H. pylori* infection developed gastric cancer, whereas no cancers were found in the non-infected patients. This prospective cohort study has provided one of the strongest causal links between *H. pylori* and gastric cancer.

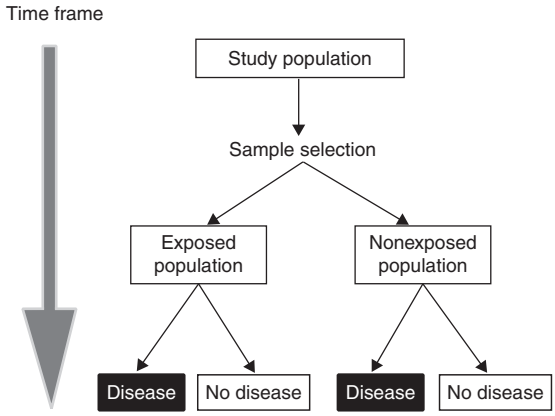


Figure 8.1 Design of a population-based cohort study.

An alternate approach is to begin the study with a pre-existing population and use historical data from the past to determine individuals who are, and those who are not, exposed to the particular risk factor. This is known as a retrospective cohort study. The design of both a prospective and retrospective cohort study is essentially the same: that is, a comparison between exposed and nonexposed individuals.

Example from the GI literature

In Taiwan, Chang et al. investigated the link between mass hepatitis B vaccination, which had been introduced in 1984, and development of hepatocellular carcinoma in later life [8]. Using a well-developed national disease registry over a 20-year follow-up period, investigators were able to observe that 64 hepatocellular cancers developed among 37,709,304 person-years in the vaccinated birth cohorts compared with 444 cancers in 78,496,406 person-years among the unvaccinated birth cohort. This retrospective cohort study provided further evidence for the link between hepatitis B infection and development of hepatocellular carcinoma in a population.

Advantages of a cohort study:

- the time sequence in the cohort study strengthens the inference between exposure and outcome
- they enable a more precise measurement of prognostic or risk factors (no recall bias)
- fewer confounding factors.

Disadvantages of a cohort study:

- period of study can be long in diseases that are slow to develop following exposure to risk factors
- large cohorts with prolonged follow-up are required in diseases with a low incidence rate
- identifying a population cohort with exposed and nonexposed individuals to a particular risk factor can be difficult, even when such evidence is available.

In view of some of these difficulties, other study designs may be employed to demonstrate an association between a particular factor and a specific disease.

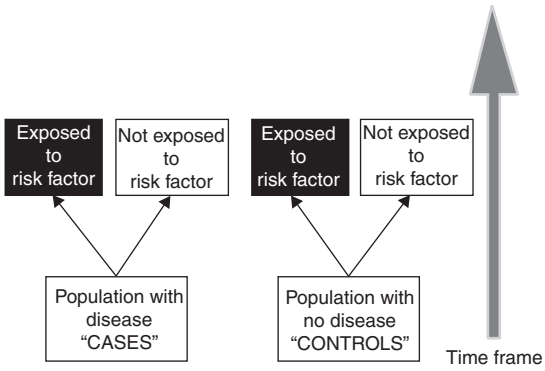


Figure 8.2 Design of a population-based case-control study.

Case-control studies

In case-control studies the outcome (disease) is identified first and the exposure is determined retrospectively (Figure 8.2). Therefore in these types of studies the population with a particular disease (cases) is first identified, and then compared with a group without the disease (controls). Ideally, the controls should be from the same population, and should be "matched" to the cases by basic sociodemographic parameters (e.g. age, gender or educational level), provided these demographic factors are not the risk factor being studied. The ratio of cases to controls can be similar, that is, 1:1, or multiple, that is, 1:2 or 1:4, and so on. By using multiple controls per case the power of the study is increased. This type of study design may be confused with that of a retrospective cohort study. However, in the latter a group of individuals with something in common from the same population (cohort) are retrospectively assessed for their exposure and nonexposure to a particular risk factor and then observed for disease development (as in the example given earlier). A case-control study begins with a diseased population and their controls, who are not necessarily from the same population, (not a cohort), with exposure assessed retrospectively in both groups.

Example from the GI literature

In Sweden, using a comprehensive hospital registry, Lagergren et al. captured all new cases of esophageal and gastric cardia adenocarcinoma (i.e. cases) between 1994 and 1997 [9]. At the same time, age- and sex-matched adults from the population identified from the Swedish population register, and individuals with newly diagnosed squamous cell esophageal cancers served as controls. Weekly GERD symptoms were found to be higher in cases (60 %) compared with controls (16 %). The investigators further calculated that adults with persistent GERD symptoms were seven times more likely to develop esophageal adenocarcinoma later in life compared with those without any symptoms. This study provided a strong, although not necessarily causal, association between GERD symptoms and esophageal adenocarcinoma.

Advantages of a case-control study:

- usually less expensive to conduct
- study duration is often quite short.

Disadvantages of a case-control study:

- recall bias in both cases and controls can influence data analysis
- precise temporal relationship between exposure and disease cannot be ascertained.

Cross-sectional studies

One of the simplest study designs in population-based surveys, a cross-sectional study, involves taking a "snapshot" view of a particular population at a point in time. In this type of study, both disease presence (outcome) and exposure to a particular factor are determined simultaneously (Figure 8.3). Hence, the proportion of the population with a particular disease identified in a cross-sectional study represents the "prevalence" of that disease in that population at that moment in time. These are sometimes known as prevalence studies.

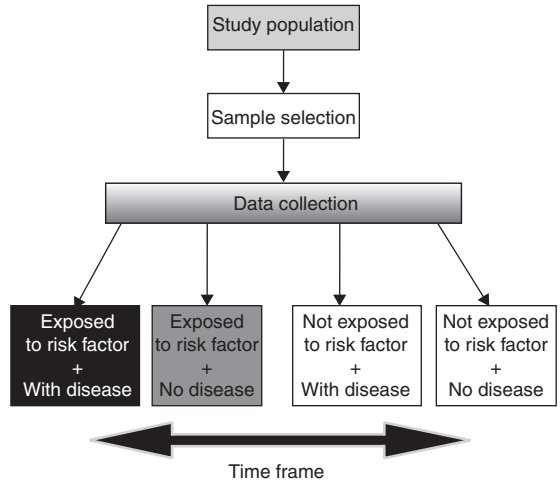


Figure 8.3 Design of a population-based cross-sectional study.

Advantages of cross-sectional studies:

- fairly quick and easy to perform
- low cost, compared with case-control and cohort studies
- useful in determining the prevalence of the disease.

Disadvantages of cross-sectional studies:

- unable to determine the temporal relationship of cause and effect
- unsuitable for studying severe or rapidly fatal disease
- problems with recall bias in some cases.

Example from GI literature

In Malaysia, Mahadeva et al. conducted a cross-sectional survey in a rural Asian population to determine the prevalence and epidemiology of dyspepsia [2]. A systematic random sampling method was used to survey a total of 2000 adults in the community. The prevalence of dyspepsia in the population was found to be 14.6 %, and was strongly associated with higher socioeconomic factors, use of regular analgesia, and the presence of other chronic illnesses.

Follow-up/longitudinal studies

A longitudinal study involves a prolonged observation of a particular population over a period of time, in order to study disease development. This form of observation can be conducted continuously (or ongoing), or it can be done by repeated cross-sectional surveys at different points in time. In medicine, this type of study is useful for determining the natural history or survival patterns of a particular disease.

Example from GI literature

In the United States, a community-based research project in Olmsted County, Minnesota, identified 426 patients with nonalcoholic fatty liver disease (NAFLD) in the population between 1980 and 2000 [10]. This study population was followed up for a mean of 7.6 years, and the natural progression to cirrhosis, liver cancer, and other liver-related complications was 5 %, 0.4 %, and 3.1 % respectively. This study demonstrated that NAFLD was not an entirely benign disease, and that progression to serious sequelae could occur within less than a decade.

Conclusion

Population-based studies, in the form of surveys, are able to provide invaluable data on the epidemiology of various GI diseases. However, detailed and careful planning is required prior to the conduct of these studies. Various methods of data collection are currently available, and investigators are advised to tailor the method to the population being studied. Whilst the cohort design provides the most precise link between exposure and disease development, case-control or cross-sectional studies are more often conducted for economic reasons.

Multiple choice questions

1 Population-based surveys are usually conducted in residential areas for the following reason:

A To capture data from individuals who are unemployed

B To capture data from pre-schooling or retired individuals

C To avoid bias among hospitalized individuals

D To obtain a representative sample of the population in general

E It is easier than collecting data in public locations such as community centers

2 In a multiethnic population, the best sampling method to obtain representative data from all major ethnic groups would be:

A Simple random method

B Stratified random method

C Systematic random method

D Direct sampling method

E Complex random method

3 A major disadvantage of the cohort population-based study design includes one of the following:

A It provides a temporal relationship between an exposed risk factor and the development of disease in individuals

B Precise measurement of exposure is possible

C Incidence rates of disease can be calculated

D The duration of study can be prolonged

E There must be equal numbers of persons in both exposed and nonexposed study groups

4 Case-control studies are often preferred to cohort studies by investigators for the following reason:

A Incidence rates can be calculated precisely

B A greater proportion of exposed to nonexposed individuals can be studied

C Recall bias is usually not a problem

D They are better for studying diseases with a low occurrence rate

E The study groups start with diseased and nondiseased persons, as opposed to exposed and nonexposed persons.

References

1 Park K. (2005) *Park's Textbook of Preventive and Social Medicine*, Banarsidas Bhanot, Jabalpur.

2 Mahadeva S, Yadav H, Rampal S, Goh KL. Risk factors associated with dyspepsia in a rural Asian population and its impact on quality of life. *Am J Gastroenterol* 2010;105(4):904–12.

3 Hu WH, Wong WM, Lam CL, et al. Anxiety but not depression determines health care-seeking behaviour in Chinese patients with dyspepsia and irritable bowel syndrome: a population-based study. *Aliment Pharmacol Ther* 2002;16(12):2081–8.

4 Ford AC, Forman D, Bailey AG, et al. Initial poor quality of life and new onset of dyspepsia: results from a longitudinal 10-year follow-up study. *Gut* 2007;56(3):321–7.

5 Hongo M. Epidemiology of FGID symptoms in Japanese general population with reference to life style. *J Gastroenterol Hepatol* 2011;26(Suppl 3):19–22.

6 Gordis L. (2009) *Epidemiology*, 4th edn, Saunders/Elsevier, Philadelphia, PA.

7 Uemura N, Okamoto S, Yamamoto S, et al. *Helicobacter pylori* infection and the development of gastric cancer. *New Engl J Med* 2001;345(11):784–9.

8 Chang MH, You SL, Chen CJ, et al. Decreased incidence of hepatocellular carcinoma in hepatitis B vaccinees: a 20-year follow-up study. *J Natl Cancer Inst* 2009; 101(19):1348–55.

9 Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *New Engl J Med* 1999; 340(11):825–31.

10 Adams LA, Lymp JF, St Sauver J, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. *Gastroenterology* 2005;129(1): 113–21.

Answers to multiple choice questions

1. D
2. B
3. D
4. D

9 How to find and apply large databases for epidemiologic research

Jonas F. Ludvigsson, Joe West, Jessica A. Davila, Timothy R. Card, & Hashem B. El-Serag

Key points

- Large databases can be a powerful source of information to examine the clinical epidemiology and outcomes of digestive and liver disorders.
- Research using large databases requires the same essential skills needed to conduct research studies using other data sources. These include a rigorous study design, expertise in analytic methods, and relevant research questions.
- The completeness and accuracy of information contained in each particular database must be assessed. Methods for improving the quality and completeness of this information should be considered.
- Despite similarities among large databases, gaining insight and experience into the structure and content of each database is essential.
- Examples of commonly used large databases are presented with a synopsis of information contained in the database, as well as strengths and limitations of using the database for research.

Introduction

Although a simple Excel spreadsheet containing information on a few subjects is technically a database, this discussion is restricted to large databases with thousands (or millions) of records. These databases may be collected primarily for research purposes (e.g. disease registries or large health surveys), but might have other primary purposes such as administrative purposes (e.g. healthcare claims), or indeed be a collection of electronic clinical notes. The primary purpose for which data is input will, to a large extent, determine the nature and quality of the data collected.

"Database study" and "data mining" are terms often used to describe research that utilizes large datasets. We feel these terms inaccurately describe many studies that utilize large databases, and underestimate the complexity and rigor of the methods used to conduct these studies. We recommend a systematic approach to utilizing large databases to address research questions, which includes:

1 developing specific research questions and determining the best possible study design to answer the question;

2 evaluating all potential data sources, which may include a pre-existing database, cross-sectional survey, or medical record review;

3 selecting the most appropriate data source based on the study question and design.

Several types of studies have been performed using large databases, including the evaluation of temporal (secular) trends, geographic variations, economic burden of disease, outcomes of disease management,

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 9.1 Databases and their web links

Data source	Website*
World Health Organization	http://www.who.int/healthinfo/statistics/mortality/en/index.html
Cancer incidence and mortality worldwide in 2008 (GLOBOCAN)	http://globocan.iarc.fr/
Surveillance, Epidemiology and End Results Program (SEER)	www.seer.cancer.gov
Medicare	www.cms.hhs.gov
Department of Veterans Administration (VA)	www.virec.research.va.gov
SEER-Medicare	www.healthservices.cancer.gov/seermedicare
American Medical Association (AMA)	www.ama-assn.org
Healthcare Cost and Utilization Project (HCUP)	www.hcup-us.ahrq.gov
National Hospital Discharge Survey (NHDS)	www.cdc.gov/nchs
Behavioral Risk Factor Surveillance System (BRFSS)	www.cdc.gov/brfss
Medical Expenditure Panel Survey (MEPS)	www.ahrq.gov
United Network for Organ Sharing	www.unos.org
United Kingdom Clinical Practice Research Datalink	www.cprd.com
Swedish National Registers	http://www.socialstyrelsen.se/statistics
National Health and Nutrition Examination Survey (NHANES)	www.cdc.gov/nchs/nhanes.htm
Canadian Institute for Health Information	http://www.cihi.ca/CIHI-ext-portal/internet/EN/ApplicationIndex/applicationindex/applications_index_main
Medicaid	www.cms.hhs.gov

* All websites accessed May 2013.

resource utilization, determinants of disease, and pharmacoepidemiologic studies. The most commonly used study designs include cross-sectional, cohort, case-control, and ecologic studies.

Commonly used databases

For each database, we will provide a brief description of the contents, highlight strengths and weaknesses, and provide links for more detailed information (see Table 9.1).

Global death and cancer registries

World Health Organization mortality database

The data available on the WHO mortality database website comprise deaths registered in national civil registration systems, with underlying cause of death as coded by the relevant national authority. Underlying cause of death is defined as "the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury" in accordance with the rules of the International Classification of Diseases.

Death registration coverage and cross-national differences in coding practices, particularly in the use of codes for ill-defined and unknown causes, must be taken into account to validly compare mortality rates for specific causes across countries. Additionally, where coverage is less than 100 %, the cause of death distribution for the uncovered population may differ from that of the covered population.

GLOBOCAN

For the last 30 years, the International Agency for Research on Cancer (IARC) has published regular estimates of the incidence of, and mortality from cancer worldwide in broad areas of the world and more recently at the country level through its GLOBOCAN series [1]. The most recent set of estimates have now been updated to 2008 using new sources of data and improved methods of estimation. Facilities for the tabulation and visual description analysis of the full dataset of 182 countries and 30 world regions

by sex can be accessed via the IARC home page (www.iarc.fr).

Incidence data derive from population-based cancer registries. These may cover entire national populations but more often cover smaller, subnational areas and, particularly in developing countries, only major cities. For example, only 21 % of the world population is covered by cancer registries and for data of good quality it is even lower: only 8 % of the world population is covered by cancer registries that meet international standards. While the information from most of the developing countries may not meet a specific criteria for quality set, this information is still of unique importance as it often remains the only relatively unbiased source of information available on the profile of cancer.

Population-based cancer registries can also produce survival statistics by following up their vital status of cancer patients. Survival probabilities can be used to estimate mortality from incidence in the absence of mortality data. Mortality data are generated from the WHO mortality database described earlier and national population estimates for 2008 were extracted from the United Nations population division.

Inevitably, therefore, there is error in the estimation of cancer occurrence using these data due to the inherent problems of accurate collection. However, in terms of gaining the best estimates of worldwide cancer incidence over time, by age, sex, and geographical distribution, these data are second to none.

United States of America databases

SEER program

The Surveillance, Epidemiology and End Results (SEER) program is an important source of population-based cancer incidence and survival in the United States. It currently covers approximately 25 % of the US population. The SEER program of the National Cancer Institute (NCI) provides support for population-based tumor registries in seven metropolitan areas (San Francisco/Oakland, Detroit, Atlanta, Seattle, Los Angeles County, San Jose–Monterey Counties, and the Greater California area) and eight states (Connecticut, Iowa, New Mexico, Utah, Hawaii, Kentucky, New Jersey, and Louisiana).

The SEER database contains information on more than 2.5 million cancer cases, and approximately 160,000 new cases are accessioned each year. Routinely collected information includes patient demographics, primary tumor site, tumor morphology, stage at diagnosis, first course of treatment, and follow-up for vital status.

SEER registries hold the highest level of certification of data quality [2–4], including completeness of case ascertainment, accuracy of data recording, and reliability of data abstraction. The SEER program's standard for the completeness of case ascertainment is 98 % [5]. SEER public use data can be accessed at no cost through the SEER website (http://seer.cancer.gov). In addition, reports on cancer statistics are available from the SEER website. SEER also offers three free software programs (SEER*Prep, SEER*Stat, and Health Disparities Calculator) which can be used to analyze SEER public-use datasets.

Several studies have examined digestive and liver malignancies using the public-use SEER database. For example, we examined temporal trends in the incidence and survival of hepatocellular carcinoma [6], cholangiocarcinoma [7], esophageal adenocarcinoma [8], and malignant gastrointestinal tumors [9–11].

Medicare claims files

The Medicare Claims Data System collects information on all services provided to Medicare beneficiaries under its hospital (Part A) and supplemental (Part B) insurance plans. All Medicare beneficiaries receive Part A benefits and 95 % of beneficiaries subscribe to Part B coverage [3,4]. The former covers inpatient hospitalizations and care in skilled nursing homes, whereas the latter covers physicians' services, hospital outpatient services, durable medical equipment, home health services, and other outpatient medical services such as diagnostic X-rays and laboratory tests.

Several individual files are included as part of the Medicare database. Denominator files contain data on enrollment information, demographics (date of birth, race, zip code of residence), month-by-month eligibility information, HMO membership, and date of death.

The Medicare Provider Analysis and Review (MedPAR) File contains inpatient hospital and skilled nursing facility stay records. Information contained in this file includes dates of admission and discharge, up to

10 diagnosis codes (International Classification of Diseases (ICD)-9-CM), and up to 10 procedure codes.

The Physician/Supplier File consists of claim records and includes some beneficiary demographic information, dates of service, procedure provided (such as office visit, surgical procedure, administration of chemotherapy), place of service (e.g. office, home, outpatient hospital, skilled nursing facility, emergency room), and diagnosis codes in ICD-9-CM format.

The Outpatient Standard Analytic File (SAF) includes dates of outpatient hospital service, revenue center codes, and up to 10 fields for diagnoses (ICD-9-CM) and current procedure terminology (CPT) codes.

CMS routinely monitors and reports the accuracy of Medicare claims and payments. Public reports about data accuracy and quality are available at www.CERTprovider.org.

Medicare claims files have been utilized to examine issues in digestive disease. For example, a study was done to examine the polyp detection rate of colonoscopy using Medicare claims files [12].

Department of Veterans Affairs (VA) administrative databases

VA Patient Treatment File

Since 1970, the Patient Treatment File (PTF) has captured information about inpatient hospitalizations at approximately 127 VA facilities across the United States. The PTF contains medical diagnoses as well as inpatient medical and surgical procedures. Diagnostic (ICD-9-CM) and CPT codes are based on information contained in the medical record, such as healthcare provider progress notes, imaging studies, and laboratory reports. The PTF does not contain information about pharmacy, pathology or laboratory results.

VA Outpatient Care File

In 1996, the Outpatient Care File (OPC) was established to track visits to VA outpatient clinics. This file contains information on clinic specialty, date of visit, provider type, ICD-9-CM diagnosis codes, and CPT codes [13].

The Beneficiary Identification and Records Locator Subsystem (BIRLS) Death File

The mini-Beneficiary Identification and Records Locator Subsystem Death File has dates of death reported by the VA, the Social Security Administration, the Department of Veterans Affairs cemetery system, and funeral directors. Information on cause of death is generally not available [13]. Due to various incentives, up to 90–95 % of deaths among veterans are captured by the BIRLS file as compared with the National Death Index [14–16].

VA-Medicare linked database

This database contains VA administrative data and Medicare claims files for all Medicare-enrolled veterans who use the VA system. Data are currently available for calendar years 1999–2003. Researchers with institutional review board (i.e. an independent ethics committee) -approved protocols can request VA-Medicare linked data through the Veterans Administration Information Resource Center (VIReC; www.virec.research.va.gov).

VA databases have been used to examine the temporal trends of cases of hospitalization as a result of gastroesophageal malignancies [17] and colorectal cancer [18], and to examine the outcomes of fundoplication [17,19,20].

SEER-Medicare linked database

Data from the SEER tumor registries for cancer cases diagnosed from 1973 through 2002 have been linked with Medicare claims data from 1986 through 2003. The SEER-Medicare linkage is updated every three years.

Several studies have been conducted using the SEER-Medicare database, including ones examining risk factors for hepatocellular carcinoma [21,22] and cholangiocarcinoma [23], as well as the extent, patterns and therapeutic outcomes of hepatocellular carcinoma [24]. Other studies have examined the use of upper endoscopy prior to esophageal adenocarcinoma [25]. A complete list of published studies can be found on the National Cancer Institute website (http://healthservices.cancer.gov/seermedicare/overview/publications.html).

SEER-Medicare data are not public-use files, and therefore investigators must obtain approval prior to requesting the datasets. Research protocols and data requests can be submitted to the SEER-Medicare contact, listed on the National Cancer Institute website (http://healthservices. cancer.gov/seermedicare).

Healthcare Cost and Utilization Project (HCUP)

The Healthcare Cost and Utilization Project is a collection of healthcare databases supported by the Agency for Healthcare Research and Quality. Data are collected by state data organizations, hospital associations, private data organizations, and the Federal Government to create a resource for patient-level healthcare data.

National HCUP databases include the Nationwide Inpatient Sample (NIS) and the Kids Inpatient Database (KID). The NIS contains inpatient data from a national stratified sample of over 1000 hospitals (20 % of all US community hospitals) and is currently available for the period 1988 to 2009. It contains data from approximately eight million hospital stays on all patients, regardless of payer. Data elements in the NIS include primary and secondary diagnoses, procedures, admission and discharge status, patient demographics, expected payment source, total charges, length of stay, and hospital characteristics.

State-specific HCUP databases are also available for those states that have agreed to participate. These include the State Inpatient Databases, State Ambulatory Surgery Databases, and State Emergency Department Databases (http://www.hcup-us.ahrq.gov/sidoverview.jsp). HCUP databases are available for purchase through the HCUP Central Distributor. An online application form is available at www.hcup-us.ahrq.gov.

The NIS database has been utilized previously for gastrointestinal (GI) research. One recently published study examined differences in risk factors between black people and white people for hepatocellular carcinoma [26].

National Hospital Discharge Survey (NHDS)

The National Hospital Discharge Survey is a national probability survey, conducted annually since 1965 and designed to collect information on inpatients discharged from non-Federal, short-stay hospitals in the United States. The NHDS collects data from a sample of approximately 370,000 hospital discharges acquired from a national sample of approximately more than 400 hospitals.

Two data collection procedures are used. One is a manual abstraction of data from the medical records performed by hospital or National Center for Health Statistics (NCHS) staff. The other is an automated system in which medical record data are purchased from commercial organizations, state data systems, hospitals or hospital associations. Patient characteristics contained in the database include age, sex, race, ethnicity, marital status, and expected source of payment. Information about dates of inpatient admission and discharge, and discharge status, as well as diagnoses and procedure codes are also available. Quality control procedures and edit checks are used to maintain data quality. A detailed review is also conducted for most variables for each hospital.

Several studies have been conducted using NHDS data. For example, published studies using NHDS data examined examined trends in hemorrhoids and constipation [27,28].

Data from NHDS are released annually and can be obtained free of charge at the NCHS website (www.cdc.gov/nchs). Data files are available on public-use data tapes, or can be downloaded from the ftp (file transfer protocol) server.

Medical Expenditure Panel Survey (MEPS)

The MEPS program conducts three separate but related surveys, including the Household Component Survey, the Medical Provider Survey, and the Insurance Component Survey. The Household Component Survey collects information at the person and household level on health conditions, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment. The Medical Provider Component Survey supplements and validates information on medical care events by contacting medical providers and pharmacies identified by household respondents. The Insurance Component Survey collects data on health insurance plans obtained through private and public-sector employees. MEPS public-use data are available for download directly from the MEPS website or can be ordered on diskette or CD-ROM from the Agency for Healthcare Research and Quality (AHRQ). MEPSnet is an interactive statistical analysis program for MEPS data and is available from the MEPS website (http://meps.ahrq.gov/mepsweb/).

Other US databases

Other databases that have been used extensively for research purposes include the United

Network for Organ Sharing (www.unos.org/), National Health and Nutrition Examination Survey (http://www.cdc.gov/nchs/nhanes.html), and Medicaid (http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/).

Swedish national registers

The structure of national registers in the Nordic countries (Sweden, Finland, Denmark, and Norway) are similar and we have chosen to present Swedish registers as examples. Nordic national registers are usually organized around different health events such as death, cancer, pregnancy and birth, and other hospital care. These registers have virtually 100 % coverage, since private health care only represents a small part of Nordic health care and since each individual in the Nordic countries is assigned a unique personal identity number (PIN) [29]. The PIN is the key for all linkage between registers, and when the PIN is not stored (as in the Swedish register of congenital metabolic disorders (e.g. the phenylketonuria register)), large-scale linkages are not possible. The following databases have all been used extensively to investigate the epidemiology of digestive disease [30–39].

The Total Population Register

This register [40] includes information on the PIN [29], area of residence, sex, age, civil status, and dates of emigration of all Swedish residents. The register was computerized in 1967, and data are collected and updated continuously by the local tax offices.

Swedish Conscription Register

A similarly population-based register is the Swedish Conscription Register. This register began in 1901 and contains information about young men examined for military service. Roughly 95 % of all conscripts are aged 18–19 years. Until recently, conscription has been mandatory and stipulated by law (exemptions were only made for men with severe handicaps or congenital malformations). The conscription register includes data on IQ test, vision, hearing, and also physical examination including blood pressure measurements [41], body mass index (height and weight) [42], resting heart rate, maximum muscle strength tests, physical endurance tests, and also visual acuity [43]. The coverage of the Swedish Conscription Register has varied over time but was probably around 90 % until the late 1980s when it started to decline. Since 2010 (but informally since the late 1990s) conscription is voluntary and therefore the register only covers a proportion of young Swedish men.

Cause of Death Register

The Cause of Death Register [44,45] has existed since 1749 when a nationwide report system was initiated, and annual reports have been published throughout the twentieth century. In 1995, the National Board of Health and Welfare received death certificates on 99.7 % of all deaths. Since 1997, the Cause of Death Register is also matched with the Total Population Register to ensure that all deaths are recorded. In some 0.5 % of the deaths, no underlying cause of death is reported to the board, and the patient is then assigned the ICD-code R99.9. Of note, the Swedish Cause of Death Register uses the ICD system and not the Swedish translation of ICD. This register has been used to examine mortality in, for example, celiac disease [32] and hereditary hemochromatosis [46].

Cancer Register

The Cancer Register was established in 1958 [47]. Ninety-nine per cent of all cancers are morphologically verified, and almost 100 % of all malignancies are reported to the Cancer Register each year. More than 50,000 cases of cancer were reported to the Cancer Register in the year 2005 alone. Physicians reporting to the Cancer Register today report the ICD-10 code, the type and the location of the malignancy in plain text, as well as data on morphology according to the ICD for Oncology, 3rd Edition, from the pathologist. The Cancer Register then centrally supplements these data with the corresponding ICD-7 code.

A landmark study using the Cancer Register was the *New England Journal of Medicine* paper on ulcerative colitis and colorectal cancer [48]. This was later followed up by a paper (also taking advantage of the cancer register) showing that first-degree relatives to patients with inflammatory bowel disease were at no increased risk of colorectal cancer, suggesting that the two diseases do not have a common cause.

Hospital Discharge Register (HDR) and Outpatient Register

The HDR contains individual data from hospital-based discharges in selected parts of Sweden since 1964 (for a review of the HDR see [49]). The register has covered all of Sweden since 1987. Some 99 % of all hospital discharge records contain at least one diagnosis. When 900 diagnoses in the HDR were evaluated in 1994, 36/36 (100 %) of recorded myocardial infarctions, and 20/21 (95 %) hip fractures were correct [50].

The Outpatient Register started in 2001 (a day-surgery module started in 1997) and includes data on all outpatient visits to hospitals but not in primary care. The Outpatient Register allows healthcare personnel and researchers to identify patients with diseases that do not necessitate hospital admission. During 2001–2006, 74 % of outpatient visits in ambulatory care were reported to this register.

The Hospital Discharge Register and the Outpatient Register together make up the national Patient Register in Sweden. The Patient Register is often used in comorbidity studies and to identify cohorts of patients with a certain disorder. For instance, Andersson et al. used it to examine whether appendectomy protects against ulcerative colitis [37]; and whether gastric ulcers are related to stomach cancer [38], while a Swedish-Danish study used it when examining the relationship between celiac disease and primary biliary cirrhosis [51].

Medical Birth Register

The Medical Birth Register contains antenatal and perinatal data on >98 % of all births in Sweden since 1973 [52]. Data collection starts at registration for antenatal care, which occurs by the 12th week of gestation in more than 90 % of the pregnancies [53]. During a first interview, the pregnant woman is asked about demographic information, smoking habits, and previous medical, obstetric and gynecologic history. Thereafter, a pregnant woman generally visits the antenatal care unit around 12 times before delivery. From 1973 to 1982, specific "Medical Birth Reports" were constructed and utilized for register data. Since 1982 data are collected prospectively on standardized forms starting at the first prenatal health visit, and variables include smoking. A recent validation of the Medical Birth Register found that the majority of the variables in the register are of high quality [53].

As part of the Medical Birth Register, all congenital malformations (1964 to June 30, 2008: from gestational week 28 and above; since July 1, 2008: from gestational week 22 and above) must be reported to the Swedish Surveillance for Congenital Anomalies. This later registry is formally a part of the Medical Birth Register. Coverage of anomaly surveillance varies between counties (this is not the case with the main Medical Birth Register). In 2009, 17.5/1000 newborns were reported to have a congenital anomaly or chromosomal defect (http://www.socialstyrelsen.se/publikationer2010/2010-11-16).

The Medical Birth Register contains data on various measurements antepartum. Stephansson et al. used this register to examine the relationship between hemoglobin levels in pregnant woman and risk of stillbirth [54]. It has also been used to investigate the relationship between gestational duration and cardia adenocarcinoma [55].

Multigeneration Register

The Multigeneration Register contains information on the parents and children of all individuals in Sweden born from 1932 onwards and surviving until 1961. Adoption or other nonbiologic relations are flagged in the register. This register allows researchers to explore not only fertility but also mortality and morbidity in first-degree relatives.

The Multigeneration Register is mostly used to study familial associations [34,56,57].

Other registers

Other national registers that may be of importance to GI epidemiologists include the Educational Register (as a measure of socioeconomic position), and registers that will allow for correct calculations of follow-up time and number of person-years such as the Register of Emigrations and Immigrations.

There are also a large number of Swedish National Quality Registers (n = 89 in 2011) (www.kvalitetsregister.se, accessed May 30, 2011). These are maintained by local counties (although some of these quality registers are nationwide). Among the oldest and most explored of these registers is the Register of Information and Knowledge about Swedish

Heart Intensive Care Admissions (RIKS-HIA). It contains data on individuals with myocardial infarction, but also data on hypertension, smoking, and medication in individuals with myocardial infarction. This register started in 1991. In 1995, 21 hospitals had joined the register, and in 2005, 73 of 76 hospitals that managed individuals with acute ischemic heart disease reported to the RIKS-HIA. Of special interest to GI epidemiologists are the national quality registers SWIBREG (Swedish inflammatory bowel disease registry); Gall-RIKS (gallbladder disease registry); National abdominal hernia register; National hernia register, and SoReg – the Scandinavian Obesity Surgery Register.

UK databases

UK primary care databases

Various electronic primary care databases exist in the United Kingdom and they are in essence similar. In the UK National Health Service (NHS) primary care provides both a gate-keeping and referral role for healthcare utilization. Since the early 1990s the majority of primary healthcare contacts and written prescriptions have been recorded and stored electronically. This development was driven by the ease of use and security provided by electronic systems. Alongside this practical development an opportunity for research utilizing these data was taken and since that time the use of primary care electronic healthcare data for research purposes has grown exponentially. One example is the Clinical Practice Research Datalink: www.cprd.com/ (formerly known as the General Practice Research Database (GPRD)), which was one of the first such databases to be developed [58–60]. This database contains diagnostic and prescription data for over 13 million people of the general population in the United Kingdom as recorded in primary care. The database consists of observations, diagnoses made by, and therapies prescribed by GPs plus information sent to them from hospitals such as pathology and radiology reports, and discharge letters. With the development of direct linkages to pathology laboratories results of blood tests are often directly linked to the practices. This database has been used extensively to investigate the epidemiology of digestive disease [61–68].

Hospital Episode Statistics (HES)

HES data contain details of all admissions to NHS hospitals in England comprising demographic data along with information about discharge diagnoses and procedures. There are similarly available data for Wales and, in a slightly different guise, Scotland (Scottish Morbidity Records). HES diagnostic and procedural data are coded using ICD-10 and the Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures, 4th Revision (OPCS-4), respectively. For each episode and admission, therefore, a reasonably complete record of the illnesses that occurred and the therapeutic or investigative procedures undergone during any given hospital admission are available. Additional information is available about the NHS provider and the patient. For example, which consultant or specialty the patient was admitted under and treated by, which region the provider resides in, and the socioeconomic status of the patient's residence from which they were admitted. An NHS cost is assigned to discharge information based on Health Care Resource groupings. Abstracts of the data are freely available at www.hesonline.nhs.uk, and sets of individual records can be obtained by application. This database has been used extensively to study the epidemiology of digestive disease [69–74].

Office for National Statistics for England and Wales (ONS)

The ONS uses a combination of census, surveys, and administrative sources to produce its data. Teams of statisticians, researchers and analysts produce series of data to inform users in their particular area of expertise. Information that is available from ONS for epidemiologic research includes data from cancer and death registration giving absolute numbers of events and populations, and therefore the ability to calculate incidence and mortality rates. There are also details of cancer diagnosis date, type, site, histology, and stage of disease. In many of these aspects ONS data is similar to that provided by Statistics Sweden described earlier. Data can be accessed in aggregated forms at www.statistics.gov.uk.

Socioeconomic status information

Data linked to the hospital admissions data and primary care through postcodes contains information

from the 1991 and 2001 census. This includes derived indices of socioeconomic status such as the Index of Multiple Deprivation, Townsend score, and other geographically based data at the level of Super Output area which represents a small area of around 400 homes.

Linked UK electronic data

In 2010 linkages were made between the GPRD, HES, ONS and socioeconomic data. For the first time, primary and secondary care electronic data were linked alongside that from ONS relating to cancer incidence and cause and fact of death. This gives a full healthcare record for about 5 % of the population in England and Wales from the year 2000 onwards. Added to that the recent linkage between HES and ONS gives death records linked to the entirety of hospital admissions for England from 1998 onwards. The linkage of these datasets is an ongoing process and it is therefore to be expected that data will become increasingly rich with the passage of time.

Other UK data

Numerous large bespoke cohort studies and cross-sectional surveys carried out in the twentieth and twenty-first centuries in various parts of the United Kingdom are available at the UK Data Archive (www.data-archive.ac.uk/).

Personal identity numbers

PINs, or National Identification Numbers, make up suitable linkage keys when matching large datasets. PINs typically consist of an individual's date of birth and a unique birth number that makes it possible to differentiate between two individuals born on the same day. The Swedish PIN consists of year-month-day and a three-digit birth number that is odd for men (e.g. 999) and even for women (e.g. 998) and a control digit (that is calculated based on the date of birth, and the birth number) [29]. The control number reduces the risk of errors, and entry of the wrong person number into digital systems.

The other Nordic countries also have PINs, although both Finnish and Norwegian PINs are made up of 11 characters (not only figures). Many countries also have systems to identify immigrants, and people with transitional work in a country. In the Nordic countries these extra identity numbers are often called coordination numbers.

Many countries, including France and New Zealand, have social insurance numbers or a national health index number; that is, a national identification number used for taxation purposes, employment, health care or social security. In some countries, only adolescents/adults (e.g. Hong Kong) or adults (China) receive an identity card. It should be noted that in some countries the use of personal identity numbers is seen as a threat to personal freedom. The Hungarian constitutional court has forbidden the use of national PINs. In the United States the social security number has become a de facto PIN (http://www.ssa.gov/pubs/10002.html), and is used for employment, to collect social security benefits, and to receive certain government services.

Other databases

There are a myriad of other databases worldwide that it is beyond the scope of this chapter to describe in detail. We have concentrated on those about which we have acquired specific knowledge, and which are accessible widely, at least within our nations. A number of other data sources have had some prominence in recent gastrointestinal research, however, and we will briefly mention some of them. A number of administrative/claims databases have been used among which the datasets from the Kaiser Permanente and the Manitoba Health insurance plan have been prominent. Canadian health data is in fact collected in almost the entire population, and some of it is linked and available through the Canadian Institute for Health Information (www.cihi.ca). In Denmark, data which are constructed under very similar conditions to the Swedish data described earlier has, in addition, linked prescription information, thereby making it similar in some respects to the primary care data from the United Kingdom, and this has also been published on repeatedly.

In countries outside North America and Europe fewer databases of the types discussed have been used for epidemiologic research. Data which could be used in this manner are likely to exist, however, in many other nations. One example in Asia is the Taiwan

National Health Insurance claims database. This database covers 99 % of Taiwan's population and contains records of inpatient and outpatient care, as well as costs, demographic data and all prescriptions issued.

Web links for many of the data sources discussed are provided in Table 9.1.

Recommendations for the use of large databases for research studies

Most caveats described in this section are extensions of sound design and analysis of clinical or epidemiologic research studies, irrespective of the data source (Box 9.1).

Because the information collected in most administrative databases has not been collected with a specific research question in mind, the *completeness* and *accuracy* of information for exposures and outcomes of interest, as well as potential confounders and effect modifiers, should be evaluated. The included data and its quality will almost inevitably reflect the motivation for the existence of the data in the first place. Often this is for the purposes of purchasing and providing health care on a country-wide basis, counting persons alive, dead, born each year, and so on. Understanding why the data exist, and how they came to do so, is crucial to their correct use for epidemiologic research, and cannot be underestimated as a use of time.

Completeness of the database

The investigator has to ask the question: does this data source capture all patient encounters? For example, in the United States patients enrolled in a health maintenance organization (HMO) are likely to receive all, or most of, their care within the constraints of their HMO, as long as they are enrolled, and therefore the majority of their healthcare utilization is likely to be captured. Similarly, the great majority of individuals aged 65 and over will have their healthcare claims recorded in Medicare, and once enrolled, most persons remain in Medicare. Conversely, Medicaid is a less stable engagement, where persons qualify based on income-related criteria and are reviewed periodically. They may therefore lose their Medicaid coverage from time to time and more frequently. In the United Kingdom, by contrast, nearly all of the population is registered with a general practitioner, who provides a gate-keeper role for healthcare utilization. In this context, contact with both primary and secondary care is recorded within the primary care data, with the purpose of having as complete a record as possible. Clearly, individuals can move from practice to practice, so some loss of continuity is therefore inevitable. Another example of possible incomplete data ascertainment is for death. While death may intuitively be considered a robust outcome for any study, some populations at great risk of death will inevitably be excluded from many healthcare databases, such as the homeless. This can lead to an underestimate of the true population mortality rate.

Box 9.1 Basis for successful conduct of studies using large databases

- Advanced knowledge of study design and analysis
- Detailed knowledge of the content research area (e.g., the clinical and epidemiologic aspects of the disorder)
- Knowledge of the database structure and its contents
- Computer programming skills

Representativeness of the database

The investigator will also often need to consider of which population a database is representative. A national, and essentially universal, health system as found in Canada, Taiwan, or the United Kingdom will provide data likely to be generalizable to the whole population. In a more fragmented system, as in the United States, it will be necessary to consider whether the findings can be generalized beyond the population using a particular provider.

Accuracy of information

Many administrative databases use the ICD, in some form or another, to code diagnoses. Many other coding structures exist such as the Read and Oxford Medical Information Systems (OXMIS) codes in the NHS, Systematized Nomenclature of Medicine (SNOMED)

in many other parts of the world, and so on. Equally, procedures may be coded with a variety of methods. These codes are selected based on information contained in the medical record. The presence and accuracy of codes that are specific for the condition of interest is a potential limiting factor of studies that use administrative databases, and therefore should be evaluated prior to using these codes for research. Not all conditions have specific codes; for example the codes for pancreatitis (acute and chronic) may not distinguish between alcoholic and biliary causes. Moreover, the accuracy of these codes can vary depending on the disease, as well as the database. Positive and negative predictive values can be calculated for each code to determine its accuracy. The positive predictive value refers to the presence of disease when the code is present, while negative predictive value refers to the absence of disease in the absence of the code. The accuracy of codes varies depending on the condition, even within the same database, and therefore has to be dealt with individually, one disease or procedure at a time. For instance, rapidly symptomatic and easily diagnosed conditions (e.g. esophageal cancer) are unlikely to remain undiagnosed and therefore the negative predictive value of these conditions is likely to be high. Conversely, the positive predictive value, although intuitively high, may not be specific enough to distinguish between esophageal adenocarcinoma, esophageal squamous cell carcinoma, and other gastro-esophageal junction cancers.

To evaluate the accuracy of diagnostic and procedure codes, an advisable approach is to conduct a survey or chart validation study of subjects nested within the study cohort that was identified in the database. For example, in a study of esophageal peptic strictures, one would identify a randomly selected group of individuals in the database with and without the ICD-9-CM code 150.3 (esophageal stricture) [75]. The medical records for these subjects are then manually or electronically reviewed for the presence (or absence) of esophageal peptic strictures. Agreement between the medical record "gold standard" and the databases can then be evaluated and estimates of accuracy reflecting both positive and negative predictive values for ICD-9-CM code 150.3 can be calculated. The investigator may then decide not to pursue the study question any further due to poor accuracy of the crucial codes in the database. Alternatively, if accuracy is very high, the study can be conducted with great confidence. A likely scenario is that the accuracy is intermediate; in which case, algorithms can be constructed to improve the accuracy of those codes. For example, while codes for upper GI bleeding might be low if only these codes are examined, an algorithm using a logistic regression model that incorporates the presence of hospitalization, an upper GI endoscopy, and blood transfusion into the definition is likely to increase the accuracy of the original codes (under the condition that such procedures are recorded in the database). Such an algorithm also allows the investigator to conduct sensitivity analyses that account for possible miscoding.

It is the responsibility of the investigators to develop a comprehensive, accurate and updated list of codes to indicate a disease condition or a medical/surgical procedure because these codes change over time, with new codes appearing and old codes disappearing. The number of available fields per record in which diagnoses/procedures can be entered should also be considered. For example, a spurious increase in the rate of a disease condition (especially conditions that are unlikely to be the primary reason for the encounter with the healthcare system) may be seen as a result of increasing the number of fields per encounter in which diagnoses can be recorded.

Use of publicly available statistical calculators

For several databases described above (e.g. SEER), there are publicly available calculators to perform statistical computations (e.g. SEER*STAT, which is connected to a Health Disparities Calculator). Although these calculators are convenient to use, it is important to verify that data are being inputted properly into the software program, and that calculations are being correctly performed. We advise investigators to emulate the calculation of previously known figures/rates, even if they do not pertain to the question of interest, to ensure that the program is being used correctly.

Determining patient comorbidity

Patient comorbidity can be captured and adjusted for by calculating one of several disease comorbidity indices [76–83] stemming from the seminal work of Charlson [83]. Ideally, we recommend the use of an index that includes conditions recorded in both inpatient and outpatient files if possible. Older indices

have relied on inpatient diagnoses, but as hospitalizations for most conditions have declined steeply, the amount of comorbidity that can be captured through hospitalization records is relatively limited. Primary care data therefore provides an excellent opportunity to measure comorbidity. Nevertheless, often, a minority of patients in studies that use a comorbidity index have no recorded comorbidity at all. Therefore residual unmeasured comorbidity may still be present, and may confound the observed associations. Diagnosis-based measures, however, are subject to the many known limitations of administrative claims data, such as incomplete or inaccurate coding [84,85]. A growing body of literature has examined the use of pharmacy prescription-dispensing information to create comorbidity measures, where the use of drugs indicates the disease condition of the patient [86,87]. The rationale for their use is that pharmacy prescription records may not have the same weaknesses as diagnostic information. For example, pharmacy measures are based on the actual fill record, and are not subject to variations of coding diagnoses. However, this approach is subject to the availability of complete pharmacy data.

Robustness of findings

Given the various reasons for misclassification and incomplete recording of exposure and outcomes of exposures, outcomes and confounders, one has to be convinced that the results are consistent or robust. Therefore, we recommend performing sensitivity analyses to test the robustness of findings, given different assumptions for accuracy and completeness of disease outcome and exposures. The source for assumptions included in the sensitivity analyses can be derived from the chart validation studies described in the previous section. Given that it is highly unlikely that any code or combination of codes will yield 100 % accuracy, one can define the variables using the worst- and best-case scenario for accuracy and then repeat or rerun the analysis using both assumptions. If the findings are consistent with the main analyses then confidence is given to the findings, otherwise the results should be interpreted cautiously.

Power and sample size considerations

An advantage of using administrative databases is the ability to examine a very large number of subjects using one data source. This large sample size enables the detection of small differences in rare outcomes. The potential disadvantage is that statistically significant differences can be detected that may not be clinically meaningful. One should not confuse large sample size with the number of outcomes of interest. For example, a study with a sample size of one million subjects that has only 30 outcome events (i.e. a rare cancer) is still underpowered. The ability to adjust for potential confounders and effect modifiers is dependent on the number of outcome events, not on the entire underlying sample size. As a rule of thumb, 10 outcome events are required to adjust adequately for one predictor variable.

In conclusion, large databases represent a potentially valuable source of information for research studies that examine the epidemiology and outcomes of a variety of digestive and liver disorders. Regardless of the data source, it is important to begin with an important research question, and a study design that properly addresses that question. If the research question and study design lend themselves to utilizing a particular database as the data source, it is the responsibility of the investigator to consider the strengths (e.g. large sample size, long duration of follow-up, relative low cost and short time required to conduct the study) versus the weakness (issues related to accuracy and completeness of information, and the availability of appropriate expertise in the particular database and advanced computer programming skills) in deciding whether to use the database for their research.

References

1 Ferlay J, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127(12):2893–917.

2 Edwards BK, et al. Annual report to the nation on the status of cancer, 1975–2002, featuring population-based trends in cancer treatment. *J Natl Cancer Inst* 2005;97(19):1407–27.

3 Warren JL, et al. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care* 2002;40(Suppl 8):IV-3-18.

4 Potosky AL, et al. Potential for cancer related health services research using a linked Medicare-tumor registry database. *Med Care* 1993;31(8):732–48.

5 Zippin C, Lum D, Hankey BF. Completeness of hospital cancer case reporting from the SEER Program of the

National Cancer Institute. *Cancer* 1995;76(11):2343–50.

6 El-Serag HB, et al. The continuing increase in the incidence of hepatocellular carcinoma in the United States: an update. *Ann Intern Med* 2003;139(10):817–23.

7 Shaib YH, et al. Rising incidence of intrahepatic cholangiocarcinoma in the United States: a true increase? *J Hepatol* 2004;40(3):472–7.

8 El-Serag HB, et al. Epidemiological differences between adenocarcinoma of the oesophagus and adenocarcinoma of the gastric cardia in the USA. *Gut* 2002;50(3):368–72.

9 Tran T, Davila JA, El-Serag HB. The epidemiology of malignant gastrointestinal stromal tumors: an analysis of 1,458 cases from 1992 to 2000. *Am J Gastroenterol* 2005;100(1):162–8.

10 Nordenstedt H, White DL, El-Serag HB. The changing pattern of epidemiology in hepatocellular carcinoma. *Dig Liver Dis* 2011;42(Suppl 3): S206–14.

11 Siegel AB, et al. Racial disparities in utilization of liver transplantation for hepatocellular carcinoma in the United States, 1998–2002. *Am J Gastroenterol* 2008;103(1):120–7.

12 Cooper GS, Chak A, Koroukian S. The polyp detection rate of colonoscopy: a national study of Medicare beneficiaries. *Am J Med* 2005;118(12):1413.

13 Boyko EJ, et al. US Department of Veterans Affairs medical care system as a resource to epidemiologists. *Am J Epidemiol* 2000;151(3):307–14.

14 Fisher SG, et al. Mortality ascertainment in the veteran population: alternatives to the National Death Index. *Am J Epidemiol* 1995;141(3):242–50.

15 Page WF, Braun MM, Caporaso NE. Ascertainment of mortality in the U.S. veteran population: World War II veteran twins. *Mil Med* 1995;160(7):351–5.

16 Page WF, Mahan CM, Kang HK. Vital status ascertainment through the files of the Department of Veterans Affairs and the Social Security Administration. *Ann Epidemiol* 1996;6(2):102–9.

17 El-Serag HB, Sonnenberg A. Opposing time trends of peptic ulcer and reflux disease. *Gut* 1998;43(3):327–33.

18 Rabeneck L, et al. Surgical volume and long-term survival following surgery for colorectal cancer in the Veterans Affairs Health-Care System. *Am J Gastroenterol* 2004;99(4): 668–75.

19 Dominitz JA, et al. Complications and antireflux medication use after antireflux surgery. *Clin Gastroenterol Hepatol* 2006;4(3):299–305.

20 Tran T, et al. Fundoplication and the risk of esophageal cancer in gastroesophageal reflux disease: a Veterans Affairs cohort study. *Am J Gastroenterol* 2005;100(5):1002–8.

21 Davila JA, et al. Hepatitis C infection and the increasing incidence of hepatocellular carcinoma: a population-based study. *Gastroenterology* 2004;127(5): 1372–80.

22 Davila JA, et al. Diabetes increases the risk of hepatocellular carcinoma in the United States: a population-based case-control study. *Gut* 2005;54(4):533–9.

23 Shaib YH, et al. Risk factors of intrahepatic cholangiocarcinoma in the United States: a case-control study. *Gastroenterology* 2005;128(3): 620–6.

24 El-Serag HB, et al. Treatment and outcomes of treating of hepatocellular carcinoma among Medicare recipients in the United States: a population-based study. *J Hepatol* 2006;44(1):158–66.

25 Cooper GS, Kou TD, Chak A. Receipt of previous diagnoses and endoscopy and outcome from esophageal adenocarcinoma: a population-based study with temporal trends. *Am J Gastroenterol* 2009;104(6):1356–62.

26 Yu L, et al. Risk factors for primary hepatocellular carcinoma in black and white Americans in 2000. *Clin Gastroenterol Hepatol* 2006;4(3):355–60.

27 Johanson JF, Sonnenberg A. The prevalence of hemorrhoids and chronic constipation. An epidemiologic study. *Gastroenterology* 1990;98(2):380–6.

28 Johanson JF, Sonnenberg A. Temporal changes in the occurrence of hemorrhoids in the United States and England. *Dis Colon Rectum* 1991;34(7):585–91; discussion 591–3.

29 Ludvigsson JF, et al. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009;24(11):659–67.

30 Elfstrom P, et al. Risk of lymphoproliferative malignancy in relation to small intestinal histopathology among patients with celiac disease. *J Natl Cancer Inst* 2011;103(5):436–44.

31 Ahsberg K, et al. Hospitalisation of and mortality from bleeding peptic ulcer in Sweden: a nationwide time-trend analysis. *Aliment Pharmacol Ther* 2011;33(5):578–84.

32 Ludvigsson JF, et al. Small-intestinal histopathology and mortality risk in celiac disease. *JAMA* 2009;302(11):1171–8.

33 Ekstrom Smedby K, et al. Autoimmune disorders and risk of non-Hodgkin lymphoma subtypes: a pooled analysis within the InterLymph Consortium. *Blood* 2008;111(8):4029–38.

34 Elmberg M, et al. Cancer risk in patients with hereditary hemochromatosis and in their first-degree relatives. *Gastroenterology* 2003;125(6):1733–41.

35 Andersson RE, et al. Appendectomy is followed by increased risk of Crohn's disease. *Gastroenterology* 2003;124(1):40–6.

36 Askling J, et al. Cancer incidence in a population-based cohort of individuals hospitalized with celiac disease or dermatitis herpetiformis. *Gastroenterology* 2002;123(5):1428–35.

37 Andersson RE, et al. Appendectomy and protection against ulcerative colitis. *New Engl J Med* 2001;344(11): 808–14.
38 Hansson LE, et al. The risk of stomach cancer in patients with gastric or duodenal ulcer disease. *New Engl J Med* 1996;335(4):242–9.
39 Ekbom A, et al. Risk of extrahepatic bileduct cancer after cholecystectomy. *Lancet* 1993;342(8882):1262–5.
40 Johannesson I. (2002) *The Total Population Register of Statistics Sweden. New Possibilities and Better Quality*, Statistics Sweden, Örebro.
41 Johansson S, et al. Risk of high blood pressure among young men increases with the degree of immaturity at birth. *Circulation* 2005;112(22):3430–6.
42 Olen O, et al. Coeliac disease and body mass index: A study of two Swedish general population-based registers. *Scand J Gastroenterol* 2009;44(10):1198–206.
43 Mollazadegan K, Ludvigsson JF. Coeliac disease does not affect visual acuity: A study of young men in the Swedish national conscripts register. *Scand J Gastroenterol* 2009:1–6.
44 Johansson LA, Westerling R. Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics. *Int J Epidemiol* 2000;29(3):495–502.
45 de Faire U, et al. A validation of cause-of-death certification in 1,156 deaths. *Acta Med Scand* 1976;200(3): 223–8.
46 Elmberg M, et al. Increased mortality risk in patients with phenotypic hereditary hemochromatosis but not in their first-degree relatives. *Gastroenterology* 2009;137(4):1301–9.
47 Mattsson B, Wallgren A. Completeness of the Swedish Cancer Register. Non-notified cancer cases recorded on death certificates in 1978. *Acta Radiol Oncol* 1984;23(5):305–13.
48 Ekbom A, et al. Ulcerative colitis and colorectal cancer. A population-based study. *New Engl J Med* 1990;323(18):1228–33.
49 Ludvigsson JF, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.
50 Nilsson AC, et al. [Reliability of the hospital registry. The diagnostic data are better than their reputation]. *Lakartidningen* 1994;91(7):598, 603–5.
51 Sorensen HT, et al. Risk of primary biliary liver cirrhosis in patients with coeliac disease: Danish and Swedish cohort data. *Gut* 1999;44(5):736–8.
52 Cnattingius S, et al. A quality study of a medical birth registry. *Scand J Soc Med* 1990;18(2):143–8.
53 The Swedish Medical Birth Register: A summary of content and quality (2003). Available from www.sos.se; (full text pdf at /fulltext/112/2003-112-3) (last accessed May 21, 2013).
54 Stephansson O, et al. Maternal hemoglobin concentration during pregnancy and risk of stillbirth. *JAMA* 2000;284(20):2611–17.
55 Akre O, et al. Perinatal risk factors for cancer of the esophagus and gastric cardia: a nested case-control study. *Cancer Epidemiol Biomarkers Prev* 2006;15(5): 867–71.
56 Hemminki K, et al. Familial association of inflammatory bowel diseases with other autoimmune and related diseases. *Am J Gastroenterol* 2010;105(1):139–47.
57 Hemminki K, et al. Effect of autoimmune diseases on mortality and survival in subsequent digestive tract cancers. *Ann Oncol* 2012; doi: 10.1093/annonc/mdr590 (first published online: January 6, 2012).
58 Hollowell J. The General Practice Research Database: quality of morbidity data. *Popul Trends* 1997(87): 36–40.
59 Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997;350(9084):1097–9.
60 Garcia Rodriguez LA, Perez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol* 1998;45(5): 419–25.
61 Armstrong RG, West J, Card TR. Risk of cancer in inflammatory bowel disease treated with azathioprine: a UK population-based case-control study. *Am J Gastroenterol* 2010;105(7):1604–9.
62 Card T, et al. Hip fractures in patients with inflammatory bowel disease and their relationship to corticosteroid use: a population-based cohort study. *Gut* 2004;53(2): 251–5.
63 Card TR, et al. Is an internal comparison better than using national data when estimating mortality in longitudinal studies? *J Epidemiol Community Health* 2006;60(9):819–21.
64 Crooks CJ, et al. The epidemiology of haemochromatosis – a population based study. *Aliment Pharmacol Ther* 2009;29:183–92.
65 Grainge MJ, West J, Card TR. Venous thromboembolism during active disease and remission in inflammatory bowel disease: a cohort study. *Lancet* 2010;375(9715):657–63.
66 Solaymani Dodaran M, et al. Risk of oesophageal cancer in Barrett's oesophagus and gastro-oesophageal reflux. *Gut* 2004;53:1070–4.
67 West J, et al. Fracture risk in people with celiac disease: a population-based cohort study. *Gastroenterology* 2003;125(2):429–36.
68 West J, et al. Malignancy and mortality in people with coeliac disease: population-based cohort study. *BMJ* 2004;329(7468):716–19.

69 West J, Card TR. Reduced mortality rates following elective, percutaneous liver biopsies. *Gastroenterology* 2010;139(4):1230–7.
70 Jeyarajah S, et al. Diverticular disease hospital admissions are increasing, with poor outcomes in the elderly and emergency admissions. *Aliment Pharmacol Ther* 2009;30(11–12):1171–82.
71 Thomson SJ, et al. Chronic liver disease – An increasing problem: A study of hospital admission and mortality rates in England, 1979–2005, with particular reference to alcoholic liver disease. *Alcohol & Alcoholism* 2008;43(4):416–22.
72 Faiz O, et al. Traditional and laparoscopic appendectomy in adults: outcomes in English NHS hospitals between 1996 and 2006. *Ann Surg* 2008;248(5):800–6.
73 Kang JY, et al. Diverticular disease of the colon – on the rise: a study of hospital admissions in England between 1989/1990 and 1999/2000. *Aliment Pharmacol Ther* 2003;17(9):1189–95.
74 Tinto A, et al. Acute and chronic pancreatitis – diseases on the rise: a study of hospital admissions in England 1989/90–1999/2000. *Aliment Pharmacol Ther* 2002;16(12):2097–105.
75 El-Serag HB. Temporal trends in new and recurrent esophageal strictures in Department of Veterans Affairs. *Am J Gastroenterol* 2006;101(8): 1727–33.
76 Quan H, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol* 2011;173(6):676–82.
77 Khan NF, et al. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Fam Pract* 2010;11:1.
78 Yan Y, et al. Comorbidity indices to predict mortality from Medicare data: results from the national registry of atrial fibrillation. *Med Care* 2005;43(11):1073–7.
79 Quan H, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43(11):1130–9.
80 Sundararajan V, et al. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004;57(12):1288–94.
81 Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. *Med Care* 2002;40(Suppl 8):IV-26-35.
82 Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol* 2000;29(5):891–8.
83 Charlson ME, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40(5):373–83.
84 Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997;127(8 Pt 2):666–74.
85 Clark DO, et al. A chronic disease score with empirically derived weights. *Med Care* 1995;33(8):783–95.
86 Johnson RE, Hornbrook MC, Nichols GA. Replicating the chronic disease score (CDS) from automated pharmacy data. *J Clin Epidemiol* 1994;47(10):1191–9.
87 Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. *J Clin Epidemiol* 1992;45(2):197–203.

10 How to do genetic and molecular epidemiologic research

Yuri A. Saito
Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

Key points

- Genetic and molecular epidemiology are fields that take laboratory-based discoveries and translate their utility, through epidemiologic methods, to the population.
- In contrast to traditional epidemiology studies most, although not all, genetic and molecular epidemiology studies require collection of biospecimens to obtain genotype or protein expression as markers of exposure or disease.
- Careful study design and application of epidemiologic principles are required in genetic and molecular epidemiology studies.
- Study designs can incorporate families or can use unrelated cases or controls.

Genetic epidemiology (GE)

The field of genetics has evolved significantly over the past several decades. Previously, diseases were thought to be either exclusively genetic or environmental in origin. "Genetic diseases" were typically caused by one or a few highly penetrant mutations in select genes that presented at birth or early in childhood. Other "environmental" diseases were the result of environmental exposures such as infection or chemical exposures. Increasingly, the medical and scientific community is realizing that there are a number of common diseases and disorders with both genetic and environmental contributors, called "complex genetic diseases" (Figure 10.1). Complex genetic diseases are frequently observed in the general population, are thought to explain many chronic diseases, and may present clinically in quite heterogeneous ways due to the presence or absence of various environmental and genetic risk factors and their gene–environment interactions. In contrast to the older "Mendelian" diseases that were caused by rare but significant mutations, complex genetic diseases are thought to be due to minor variations in the DNA sequence, such as single nucleotide polymorphisms (SNPs), which may result only in more subtle protein function change, or a subtle change in encoded protein quantity that affects overall biologic processes. Because these genetic effects are more subtle, and because there are many genetic variants present in the genome, discovering the genetic susceptibility loci for common disorders is challenging, but not futile. Sound research methods are vital in elucidating the potential role that these genetic variants may play in disease presentation. Laboratory-based discovery of these genetic variants is becoming easier due to improved genotyping platforms. The blend of genetics and epidemiology is important to clarify the role of these variants in people. Thus, the relevance of GE as a field has increased with our growing knowledge and understanding of the human genome, as well as due to advances in sequencing technology.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Genetic diseases	Complex genetic diseases	Environmental diseases
Examples:	*Examples:*	*Examples:*
Familial polyposis syndromes	Crohn's disease	Radiation enteritis
Hirschprung's disease	Colon cancer	Viral hepatitis

Figure 10.1 Spectrum of disease etiology.

Molecular epidemiology (ME)

Molecular epidemiology also represents a blend between laboratory science and epidemiology. This area incorporates molecular and biochemical concepts and techniques with epidemiology designs and methods. Although much of the field has focused predominantly on carcinogenesis, toxicology, or infections, its principles can be applied to any disease or disorder. Measured biologic substrates include DNA, proteins, hormones, molecules, chemicals, as well as food derivatives. As such, ME frequently utilizes biologic measurements, often referred to as "biomarkers". These biomarkers may be used to measure exposure, identify mechanisms of disease itself, or even characteristics of individuals that may make them more or less vulnerable to exposures or pathogens. Similar to GE, ME also attempts to identify the interactions between environmental exposures and host factors that result in disease development. In addition, ME evaluates the contribution of genetic and/or molecular factors with environmental risk factors, even those that are identified at a molecular or biochemical level, to determine how they affect disease etiology, development, or distribution of disease in populations and families. Biomarkers identified in ME studies can then be used clinically for disease diagnosis, staging, prognosis, or predicting drug response.

Epidemiology principles

GE and ME share many features in common with traditional epidemiology. Through a combination of descriptive and analytical methods, identifying risk factors for disease development and disease progression are at the heart of the three types of epidemiology studies (Figure 10.2). However, where traditional epidemiology has evaluated only environmental risk factors and their interactions leading to disease in hosts, GE and ME focus on the evaluation of genetic risk factors, and their interactions with environmental risk factors in the development of disease in susceptible individuals. Thus, the research questions are framed slightly differently between the two disciplines, in contrast to conventional epidemiology, which asks questions such as "Who gets the disease?", "Why does an

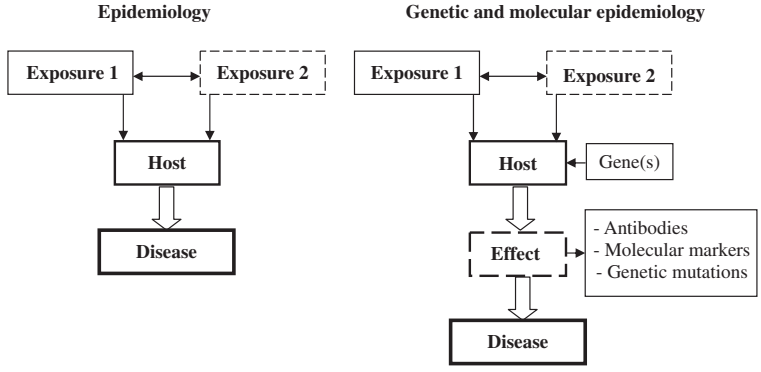


Figure 10.2 A comparison of traditional epidemiology and genetic and molecular epidemiology.

individual get the disease?" or "How does one prevent getting the disease or complications of the disease?" With GE and ME, the question may be "What are the genetic factors that lead to the development of disease?", "What is the risk of the disease in families with genetic susceptibility loci?", "How does that genetic marker predict risk of disease in the general population?", or "Does molecular protein expression predict disease prognosis or natural history?"

There are many other additional inherent differences between classic epidemiology and GE and ME as well. In general, epidemiology focuses on the study of diseases or epidemics in the general population, or a clustering of disease in a community. Although GE and ME studies may ultimately be conducted in a population-based setting, early GE studies may focus more on families, or ethnic or geographic clusters, which share similar genetic backgrounds. It follows that factors such as age of onset, penetrance and carrier state, and expressivity are terms and concepts that may be of greater interest in genetic studies rather than traditional epidemiology. Nonetheless, neither GE nor ME studies need be family-based, and as with epidemiologic studies, they may be clinic-based or population-based.

An important difference between GE and ME with traditional epidemiology is the greater emphasis on biospecimen collection and genetic, laboratory-based testing. Clearly, epidemiology is not laboratory-based. Epidemiologic studies typically require interviews, questionnaires, or databases to collect clinical data about risk factors and disease presentation. Similar data collection methods may also be used in GE and ME studies, but these usually require collection of biospecimens such as buccal swabs, blood, serum, or even tissue in order to measure genetic and/or molecular markers. As such, GE techniques can be used to answer specific questions such as identifying where a gene is in the genome map, how much this gene can explain disease in the population, and providing the causal link between the gene and the disease. As a consequence, a greater understanding of the biology of the disease, of genetics, and of genotyping technologies is required for GE studies, and a team consisting of a clinical researcher, genetic epidemiologist, statistical geneticist, and laboratory-based geneticist are usually required in order to conduct these types of studies. Similarly, ME studies also require multidisciplinary partnerships between clinical researchers, molecular geneticists, molecular biologists, and statisticians. This multidisciplinary approach can be very important as the relationship between genetic and molecular variants may not be a direct one, and is likely to involve a complex interaction between several variants and environment.

Error in GE and ME studies

As with any epidemiologic study, careful study design is an important consideration for GE and ME studies. Inappropriate study design, participant selection, or measurement error can affect the true relationship between an exposure and disease of interest (Figure 10.3). Random error, as well as systematic error including bias and confounding, remain important considerations, as in traditional epidemiologic studies. Random error (chance) is typically due to sampling variability. It can occur during data collection, data transfer, or analysis. Random error can be difficult to quantitate and hence, correction for random error is not feasible. The two main ways to reduce random error are to increase sample size and to select precise measuring instruments.

Systematic errors are consistent errors that affect the true value of the relationship between exposure and disease in the observed sample compared to the "truth". If recognized, the error may be quantifiable and generally should be reproducible, although the error may not be correctable. Measurement error that results from inaccurate measure of exposure and/or disease is an example of a systematic error that could affect the estimate of association between these two

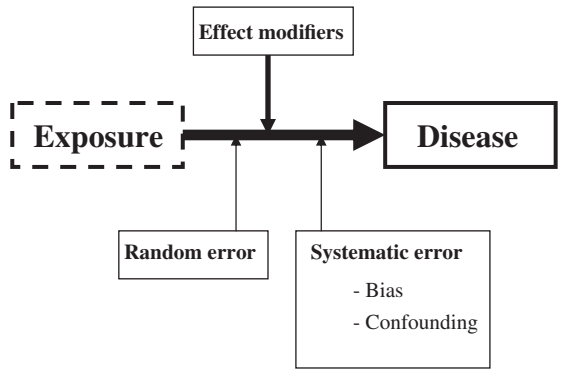


Figure 10.3 Factors affecting the association/relationship between exposure and disease.

factors. Selection and participation bias are also forms of systematic errors that may affect study outcomes. Confounding variables correlate with both the exposure and disease outcome, and can result in spurious associations if not taken into consideration in the study design and analysis.

In addition, for ME studies, validation of biomarkers is an important step. Biomarkers can be markers of exposure, biologic effect, susceptibility, or disease. Biomarkers offer an objective means of qualitatively or quantitatively measuring exposure or disease that overcome some limitations of traditional epidemiologic studies. As biomarkers are a measure of exposure or disease, the validity, precision, reproducibility, and stability of these biomarkers are important. Biomarkers can include genetic polymorphisms and mutations, chromosomal aberrations, viral DNA, proteins, hormones, or metabolites. Sources of imprecision include interindividual variation, intraindividual variation, measurement error from sampling (e.g. circumstances, timing, methods, etc.), variation from processing (e.g. storage, time to processing, etc.), laboratory technique, or interlaboratory variability. Biomarker inaccuracy or misclassification can lead to either overestimation or underestimation of risk. Thus, careful selection of biomarkers as well as careful lab measurements and assays are needed for ME studies.

Study designs and approaches

Collection of biospecimens (e.g. blood, buccal swab, or serum), followed by collection of detailed clinical information, are important first steps in the conduct of a GE or ME study. Nonetheless, the types of study designs vary considerably, and many options are available. Similar to traditional epidemiology studies, GE and ME studies can utilize case-control designs or prospective cohorts. However, GE studies do differ from traditional epidemiology and ME studies in that historically, family studies were often used. A summary of study designs that are commonly used in GE are summarized in Table 10.1.

Molecular epidemiology study designs

Study design in molecular epidemiology is no different from that in traditional epidemiology. The primary difference is that biomarkers are used to characterize exposure, susceptibility, or disease status. The markers used in ME studies may be genes, viruses, antibodies, genetic mutations or variants, or simply protein expression. More efficient, less expensive high-throughput genotyping is now allowing for greater characterization of an individual's genetic susceptibility for disease. Phenotypic assays may include measures of DNA repair capacity. Epigenetic studies that evaluate how environmental exposures and conditions can lead to DNA changes, including methylation or histone deacetylation, that subsequently affect DNA transcription and protein expression can also be incorporated into ME studies to track susceptibility as well as disease progression and/or natural history.

Table 10.1 Genetic epidemiology study designs

- **Clinical and historical studies**
  - Age of onset, race, geography, migration studies
- **Descriptive family studies**
  - Familial aggregation studies
  - Twin studies
  - Adoption studies
- **Segregation analyses**
- **Genome-wide studies**
  - Linkage analysis
  - Genome-wide association studies
- **Candidate-gene association studies**

Genetic epidemiology study designs

Clinical and historical studies

Clinical and traditional epidemiologic studies may be required prior to proceeding with other lines of inquiry. Clinical studies may indicate features of a disease that are suggestive of a genetic basis for the disease of interest. These features include association with other known genetic diseases, racial differences in predisposition, or a family history of disease. Association with other diseases that have a known genetic basis certainly provides one form of evidence for any genetic basis of that disease. If the disease of interest appears race-specific, additional migration or admixture studies may be helpful to determine whether the predisposition is due to genetics or cultural factors. Similarly, a positive family history of disease may

indicate the need for further family-based studies, to determine whether the clustering is due to genetics, shared environmental exposures, or lifestyle. Positive results from these descriptive studies are encouraging, but certainly are not conclusive for an underlying genetic basis for the disease of interest.

Family studies

Three main types of epidemiologic family studies exist: family history studies, family case-control studies, and twin studies. Family studies may simply consist of comparing the frequency of cases with a positive family history with the proportion of controls with a positive family history of the disease of interest. The most basic type of study collects family history data from affected cases, and compares the proportion with a positive family history with either the population prevalence rate or, more ideally, with age-, gender-, and race-matched controls. This study design is no different from a conventional case-control study and is also referred to as the family history approach. However, collecting family medical data from cases and controls can be fraught with error, particularly if the disease of interest is not visible or easily recognizable by lay people, has many causes, is in an early or mild stage, has a stigma associated with it, or has an onset late in life when the individual is not sharing a household with siblings or children. For example, relatives are more likely to be aware of a diagnosis of cancer or cirrhosis in the family, but may be less knowledgeable about whether the cancer is primary or metastatic, or whether the liver disease is due to primary biliary cirrhosis or nonalcoholic fatty liver disease. Family members may also not feel comfortable discussing their bowel habits with others, and thus, irritable bowel syndrome may be under-recognized by cases and controls alike. Thus, there is a danger of misclassification of the relative's affected/unaffected status based on proband report alone.

Rather than collecting the relatives' medical data from probands, a better alternative approach for familial aggregation studies is the family case-control design or family study approach. This consists of direct survey of the relatives themselves, and may involve review of their medical records, or even clinical evaluation and diagnostic testing. Therefore, misclassification of "affected" and "unaffected" status of relatives can be minimized. The disadvantage of this approach is that not all relatives may wish to participate in the study, not all relatives may be alive to participate in the study, and not all relatives may be able to participate in the study due to the presence of other medical comorbidities such as dementia. These factors may lead to much "missing" information. However, missing data can be minimized by asking next-of-kin or those with power of attorney for deceased individuals or individuals unable to participate to provide consent and release of medical information for that relative.

With the data collected from family members, various analyses may be performed. First, pedigree construction (manually or using software such as Progeny [1] or Pedigree-Draw [2]) may be illustrative to identify interesting families worthy of greater study. Second, comparison of proportion of positive family history of disease between cases and controls may also be illustrative of the risk that family members of cases have for the disease of interest, given that there is an affected family member (i.e. recurrence risk). Third, because different relative types may have different patterns of risk, risk or heritability for specific relationship may be calculated. Fourth, evaluation and detection of relevant gene–environment interactions may be performed using this study design. Thus, confirmation of familial aggregation represents an important line of investigation to determine whether additional genetic studies are warranted in searching for disease etiology.

Twin studies represent a specific type of family study, and represent a classic genetic versus environmental approach whereby a genetic basis for disease is assumed if there is greater concordance of disease in monozygotic (identical) twins than in dizygotic (nonidentical) twins. Environment is thought to play a greater role if there is greater or equal concordance of disease in dizygotic twins than monozygotic twins. If there is evidence for a genetic basis for disease, these studies can provide a quantitative estimate of general liability for disease by calculating the difference in concordance between monozygotic twins and dizygotic twins. However, twin studies have several limitations, including the inability to adjust for prenatal differences between twins (position *in utero*, manner of delivery, shared or nonshared placental circulation), postnatal differences in upbringing, and ultimately further study is warranted to identify disease gene loci. Nonetheless, twin studies often

provide additional evidence supporting a genetic basis for disease and provide justification for additional investigation.

Segregation analyses

Segregation analysis is another family-based study, but at a relationship level. Pedigrees are constructed and family members are assigned an affected or unaffected status. Segregation analysis compares the observed distribution of affected and unaffected individuals in a series of families under a specific genetic hypothesis (e.g. autosomal recessive model) to the distribution that would be expected under specific genetic or nongenetic models. For example, for a recessive disease, if the father is unaffected, the mother is affected, and some of the children were affected, one would expect that the father is heterozygous for the disease gene, the mother is homozygous, and thus, 50 % of offspring would have disease. Comparison between actual rates of disease prevalence among offspring in the accrued families and expected rates of disease would provide an estimate of how well the postulated recessive model fits actual data. Segregation analysis tests multiple genetic models, and determines which model best fits the observed data. Segregation analysis studies thus answer questions such as whether or not a major gene contributes to disease expression, whether multiple genes with small effects can result, or whether nongenetic factors contribute to disease etiology. Segregation studies are helpful not only in providing a genetic model for disease transmission, but also in estimating the penetrance and attributable frequency required for parametric analysis in follow-up linkage studies.

However, special considerations regarding segregation analysis must be made. This method identifies the best-fitting model of the ones tested, but does not necessarily reject an incorrect model (type II error). In addition, it is not immune to type I errors (i.e. incorrectly rejecting the correct model). Furthermore, segregation analysis is particularly susceptible to a form of selection bias called ascertainment bias; that is, the families studied may not be representative of those in the general population. In this situation, while adjustments for ascertainment can be incorporated in the analysis, the conclusions drawn may be distorted. Confounding by other factors, such as environmental exposures, must also be taken into account. Many of these points apply to any observational epidemiologic study, but this is particularly important because segregation analysis often provides the parameters and assumptions needed for linkage studies.

Linkage analysis

Linkage studies are performed if no gene has been firmly tied to disease causation and the investigator is trying to narrow the region in the genome map where the disease gene may lie. Approximately 500–6000 genetic markers with known genomic location are selected spanning the human genome, then these markers are genotyped in family members. Parametric or model-based linkage uses set penetrance estimates and an inheritance model (that may be derived from segregation analyses), and the observed transmission of each marker with disease status is compared to determine which genetic marker appears to be the most closely linked to the disease of interest. Alternatively, linkage analysis can be performed without such estimates (model-free linkage or nonparametric linkage analysis). It is hoped that a genetic marker will identify a region up to 10 Mb (megabases) in size, and additional markers can then be used to narrow the region of interest even further. The concept behind linkage is that genetic recombination is more likely between loci that are far apart, thus separating them during meiosis, and by looking at the recombination rate between family members (who should share large regions of the genome inherited from the same recent ancestor), the position of the disease gene relative to the marker can be estimated. Linkage studies do require the participation of families that may be extended multigenerational families, nuclear families, or specific subunits, such as affected sibling pairs. Genetic markers of study include restriction fragment length polymorphisms, variable number of tandem repeats, microsatellites, and single nucleotide polymorphisms.

The main limitation of parametric linkage methods is that the genetic model must be specified, and if incorrect, may result in false positives as well as false negatives. Any error in the model leads to inconsistent parameter estimates and lack of power. For these reasons, nonparametric, model-free methods are also utilized. In addition, because of the low rate of recombination events within most families, linkage

analyses may not be able to narrow the genomic region of interest below several Mbs. Furthermore, although linkage studies are a tried and trusted method, which has been immensely successful in identifying disease susceptibility loci for Mendelian diseases, these studies are less powerful when studying a complex, non-Mendelian, genetic disease caused by multiple genes of modest genetic effect. In this situation, association studies may be a better alternative to identifying a specific disease-causing gene or genes.

Genome-wide association studies

Genome-wide association studies (also referred to as whole-genome association studies) offer investigators the opportunity to physically localize areas of interest on the genome by analyzing genotyping data from genetic markers spanning the genome. Association studies typically use classic epidemiologic case-control designs to compare the allele frequency of a genetic marker in disease cases and unrelated healthy controls with the goal of identifying markers with allele frequencies that differ between the two groups. Association studies are thought to be advantageous when studying common alleles with modest disease risk, and do not require the assembly of pedigrees and collection of DNA from family members, as it is typically easier to collect DNA from a large series of unrelated cases and controls than it is to collect DNA from family members, some of whom may be deceased or unwilling to participate. However, because the number of shared markers between unrelated individuals will be fewer, more (thousands to possibly hundreds of thousands) high-density markers will need to be genotyped among thousands of study subjects. Although high-throughput genotyping technology advances have resulted in decreased costs, this method is still resource-intensive, and therefore, remains expensive to conduct, limiting the number of these studies that can be performed. Furthermore, genome-wide association studies, due to multiple testing issues, do result in many positive results, many of which may be false positives. As such, interpretation of positive findings must be somewhat cautious taking into account the *P*-value (the smaller the value, the better) and whether the findings were reproduced in an independent sample.

Candidate gene-disease association studies

Similar to genome-wide association studies, candidate gene-disease association studies compare allele frequencies of a given polymorphic genetic marker between cases and controls, and if an allele is found to be more common in cases, this finding suggests that this polymorphism may be involved in the development of disease. These studies may involve studying from only one to a few polymorphisms, using chi-square analysis or Fisher's exact test, and thus, are quite simple to perform and analyze. However, they require *a priori* knowledge of a putative candidate gene, and selecting the right polymorphism may not be easy with an estimated 20,000–30,000 genes in the genome, and over 25 polymorphisms in the form of single nucleotide polymorphisms, restriction fragment length polymorphisms, variable number tandem repeats, and insertion and deletions in an average 27 kb (kilobase) gene. Nonetheless, this method represents a direct test of association, which may be quite powerful if the polymorphic marker is carefully selected on the basis of biologic plausibility and potentially linkage studies. Ideally, candidate gene association studies should also include a replication cohort to provide an additional level of evidence that the positive findings were real and not due to chance.

Conclusion

Based on rapid growth in the fields of molecular biology and genetics, there has been a necessary "marriage" between these disciplines and epidemiology. The stringent methodology of traditional epidemiology is necessary to ensure that the voluminous data being generated in labs is not misrepresented due to improper study design, analysis, or interpretation. Furthermore, it is clear that many diseases or physiologic traits are based not only on inherent genetic coding, but also on environmental factors that attenuate disease presentation. Teasing out the genetic and environmental contributors to these complex diseases is not easy, but unraveling the respective genetic etiologies for each disease can be performed using sound genetic and molecular epidemiology methods and techniques. Those interested in pursuing this line of investigation are encouraged to obtain additional training, and collaborate with colleagues in a

multidisciplinary team that includes genetic and molecular epidemiologists, laboratory-based scientists, and statisticians with expertise in interpreting genetic and molecular data. Clearly, these fields will continue to evolve as new discoveries are made and new technologies are developed.

Multiple choice questions

1 Which of the following factors can affect the true relationship between an exposure and disease?

A Measurement error
B Selection or participation bias
C Study design
D All of the above

2 In molecular epidemiology studies, biomarkers can be used to measure exposure, biologic effect, susceptibility, or disease. Which of the following is not an example of a biomarker?

A Genetic polymorphisms
B Viral DNA
C Metabolites
D Smoking history

3 Of the following genetic epidemiology study designs, which requires collection of DNA?

A Twin studies
B Genome-wide association studies
C Family history studies
D Segregation analysis

4 Which of the following does *NOT* apply to segregation analysis studies?

A They identify the gene and its location on the genome
B They can provide evidence for a major gene contributing to a disease
C These studies yield mode of inheritance (e.g. autosomal dominant)
D Yield estimates of penetrance of the disease gene

5 Which of the following study designs can *NOT* help identify environment and genetic contributors for a given disease?

A Linkage analysis
B Migration studies
C Familial aggregation studies
D Twin studies

References

1 www.progeny2000.com (last accessed May 20, 2013).
2 www.pedigree-draw.com (last accessed May 20, 2013).

Answers to multiple choice questions

1. D
2. D
3. B
4. A
5. A

11 Diagnostic studies

Paul Moayyedi
Division of Gastroenterology, McMaster University, Hamilton, ON, Canada

Key points

- The diagnostic utility of a test should be expressed in terms of sensitivity and specificity, positive and negative predictive values, and/or positive and negative likelihood ratios.
- Single measures of the diagnostic utility of a test such as accuracy, diagnostic odds ratio, or area under the receiver operator curve are usually not recommended.
- The positive and negative predictive values of a test vary with the prevalence of a disease in the population under study.
- Selecting a reference standard to compare with a new diagnostic investigation can be a problem when there is no test that is sufficiently accurate. In this setting using a combination of tests or evaluating data using latent class or Bayesian analysis may be helpful.
- As with almost all study designs there are a number of biases that can be introduced into diagnostic studies and threaten their validity.

Introduction

Making an accurate diagnosis is a critical part of medical practice. An accurate diagnosis allows the patient to understand the nature of their illness and also the likely prognosis. The clinician can then suggest treatment that is most likely to benefit the patient according to evidence-based medicine principles [1]. We are often taught in medical school that 90 % of all diagnoses are made through taking a careful history. That may have been true 50 years ago but modern medicine relies heavily on tests to inform the diagnostic process.

A gastroenterologist relies on a diverse menu of investigations from blood tests, breath tests, pH and manometry studies, radiological tests, and endoscopic procedures. Unfortunately there is no test that is 100 % accurate so it is important that clinicians understand the implications of a positive or negative test result. Furthermore the advertised accuracy of a test will be based on diagnostic studies and these can have limitations that can either reduce or inflate (usually the latter) the apparent utility of a test [2]. This chapter will outline measures used to describe test performance as well as outlining how the choice of population, gold-standard test, and study design can influence accuracy.

Measures of test accuracy

Any test performance can be described using a 2×2 table that compares a positive or negative test with whether the disease is truly present or absent (Box 11.1). It is astonishing that such a small table can give rise to so many measures of test performance

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Box 11.1 Depiction of a 2 x 2 table used for a diagnostic test

		Disease	
		Present	Absent
Test result	Positive	a	b
	Negative	c	d

and how these are calculated is given in Table 11.1. There are advantages and disadvantages to each measure.

Sensitivity and specificity

The accuracy of a test in detecting or excluding a given disease is usually expressed in terms of sensitivity and specificity. Sensitivity tells us the proportion of false negatives we should expect of those who truly have the disorder, whereas specificity defines the proportion of false positives among those without the disease (Table 11.1). It can be useful to graphically represent how these metrics relate to each other in a summary receiver operator curve (sROC), which plots the sensitivity against 1-specificity (Table 11.1).

These terms are widely understood, but do not express the information that is most helpful in making clinical decisions [3]. In the clinical setting, what we really want to know are the chances that the patient actually has the disease if the test is positive. Similarly, if the test is negative, what are the chances that the patient does not have the disease? A test that has a sensitivity and specificity of 90 % would seem to be reasonably accurate, but a positive test would not be diagnostically useful if the prevalence of the disease was very low. If the probability of having a disease is 1 %, then a positive test only increases the probability of having the disorder to 8.3 % (Table 11.2) [3].

Table 11.1 Descriptions of terms used to describe the utility of a diagnostic test

Measure	Word definition	Mathematical definition
Sensitivity	Proportion of people with the disease that the test correctly identifies as positive	a/(a + c)
Specificity	Proportion of people without the disease that the test correctly identifies as negative	d/(b + d)
True positive (TP)	Correct positive result	a
True negative (TN)	Correct negative result	d
False positive (FP)	Incorrect positive result	b
False negative (FN)	Incorrect negative result	c
Positive predictive value	Proportion of people with a positive test that have the disease	a/(a + b) or TP/(TP + FP)
Negative predictive value	Proportion of people with a negative test that do not have the disease	d/(c + d) or TN/(TN + FN)
Positive likelihood ratio (LR +)	Describes the odds of a person having the disease if the test is positive	(a/(a + c))/(b/b + d) or sensitivity/(1-specificity)
Negative likelihood ratio (LR−)	Describes the odds of a person not having the disease if the test is negative	(c/(a + c))/(d/(b + d)) or (1-sensitivity)/specificity
Test accuracy	The proportion of correct results that the test gives	(a + d)/(a + b + c + d)
Diagnostic odds ratio (DOR)	Overall measure of diagnostic test accuracy. Describes the ratio of the odds that a person has the disease if the test is positive versus the odds a person does not have the disease if the test is negative	ad/bc or LR + /LR−
Summary Receiver Operator Curve (sROC)	A plot of the relationship between the sensitivity and 1-specificity of the test	Log(DOR)

Table 11.2 Variations in positive predictive value and negative predictive value with prevalence of disease for a test that is 90 % sensitive and specific

Prevalence	Positive predictive value	Negative predictive value
99 %	99.9 %	8.3 %
95 %	99.4 %	32.1 %
90 %	98.8 %	50 %
80 %	97.3 %	69.2 %
50 %	90 %	90 %
20 %	69.2 %	97.3 %
10 %	50 %	98.8 %
5 %	32.1 %	99.4 %
1 %	8.3 %	99.9 %

Source: Moayyedi and Axon 1999 [3]. Reproduced with permission of Nature Publishing Group.

Similarly, if the disease is very common, a negative test with a sensitivity and specificity of 90 % is not clinically helpful (Table 11.2) [3].

Positive and negative predictive values

To overcome this problem, positive and negative predictive values (PPV and NPV) are often quoted to describe the accuracy of a test. This is clinically more useful but has the disadvantage that these will vary with the prevalence of the disease. In the example cited earlier, the test has a PPV of 97.3 % and a NPV of 69.2 % when the prevalence of the disease under study is 80 %, whereas these change to 50 % and 98.8 %, respectively, if the prevalence falls to 10 % (Table 11.2) [3].

Likelihood ratios

What is needed is a test characteristic that expresses results in terms of the probability of patients having a disease, and one that does not vary dramatically with the prevalence in the population. The likelihood ratio [4] fulfills these criteria and can be derived from sensitivity and specificity according to formulae given in Table 11.2. The odds of a disease being present after the test can then be derived from the equation:

$$\text{post-test odds} = \text{pre-test odds} \times \text{LR}$$

Likelihood ratios can be divided in to a positive likelihood ratio, when evaluating the odds of a patient having the disease when the test is positive, and a negative likelihood ratio, when determining the odds of a patient being disease-free when the test is negative. Likelihood ratios are a more clinically relevant method of expressing the accuracy of an investigation, and yet they are not as popular to quote as either sensitivity and specificity or predictive values. The reason for this is that clinicians are used to dealing with probabilities, whereas likelihood ratios express result in terms of odds. The odds are the probability of an event occurring divided by the probability that it will not occur. The odds are similar to the probability when the event is rare but as the event becomes more common it is necessary to switch back and forth between probabilities and odds (see Box 11.2). This can be cumbersome to calculate, which may explain why this metric is not as popular as other approaches. However, there are applications for smart phones that can perform this conversion. Alternatively, for those who are more at home with pieces of paper, a simple nomogram is available that obviates the need to perform any calculations (Figure 11.1) [4]. This inconvenience is worth the effort, as likelihood ratios can provide valuable information on the clinical utility of a diagnostic test.

Single measures of test performance

Single measures of a test's performance are attractive to clinicians because of their simplicity. We are used to thinking in terms of a person being well or ill and it is intuitive to think that a test is either accurate or inaccurate. This can only be achieved with a single measure and the commonest approach to this is accuracy. Previously in this chapter I have used this term colloquially but it does have a precise mathematical definition (Table 11.1) that describes the number of times the test gives the correct result divided by the total of number of people tested.

Box 11.2 Calculations of odds and probabilities

Odds = probability/(1 − probability)
Probability = odds/(1 + odds)

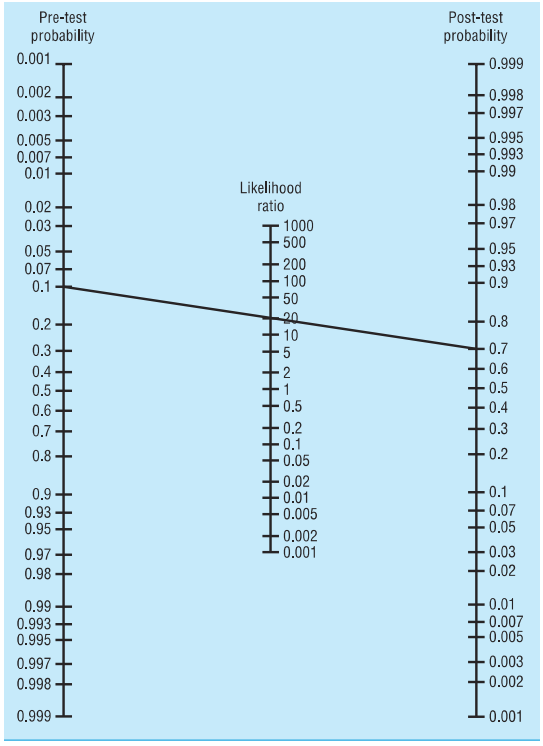


Figure 11.1 Fagan's nomogram for calculating post-test probability of disease when the pretest probability and likelihood ratios are known. Source: Deeks and Altman [4], reprinted with permission of BMJ Publishing Group Ltd.

Whilst single measures of accuracy are intuitively attractive to quote they are less useful than the other measures described earlier. This is because a test can be valuable in identifying the disease when it is present but not useful at excluding it, or vice versa. A single measure will not capture this nuance and is only useful if the test is very accurate at identifying both those with and without the disease. It is therefore preferable to quote sensitivity and specificity, PPV and NPV or positive and negative likelihood ratios, rather than unitary measures. This applies to the diagnostic odds ratios and area under the curve of sROC (Table 11.1) as well as accuracy.

The influence of choice of population on diagnostic test performance

The prevalence of disease in a population predictably influences the PPV and NPV as described earlier [3]. Variations in the proportion of people with disease can also influence sensitivity and specificity, but not as dramatically or as predictably [5]. Layered on this issue is the impact of bias in the sample studied. There are a variety of biases that can influence the results of a diagnostic accuracy study, but the most important of these are spectrum and selection bias [6].

Spectrum bias occurs when a diagnostic test is validated in one population and then applied to another with a different clinical spectrum of disease. Populations with large numbers of well subjects, as seen in screening programs, will have only a few people with disease. Furthermore, usually there will be a higher proportion of early disease compared to a population referred because of abnormal symptoms. Such early disease may be more difficult to detect, and this may lower the sensitivity of the test. An example of this would be screening colonoscopy. Endoscopy is usually extremely accurate at diagnosing colorectal cancer, but when used as a screening tool it is possible that early subtle flat neoplastic lesions may be missed [7]. The impact of spectrum bias on sensitivity and specificity is not usually as notable as the effect of prevalence on positive and negative predictive values, except in extreme circumstances [5].

Selection bias occurs when there is an association between the test result and the probability of being included in the study that is validating the test [6]. For example, colonoscopy is accurate at diagnosing ulcerative colitis but if a study evaluated this procedure in inpatients with diarrhea the proportion of those with ischemic colitis, graft versus host disease, and *Clostridium difficile* colitis would be much higher than seen in the outpatient setting, and may lead to misdiagnosis, lowering the apparent sensitivity and specificity of the test compared with the same study if it had been done in the latter setting. Another example is the specificity of right lower quadrant tenderness in the diagnosis of acute appendicitis, which fell from 89 % in primary care to 16 % in tertiary care [8].

Spectrum and selection biases therefore will impact on the sensitivity and specificity of the test. These are often unavoidable as there is no "perfect" population. Studies should, however, ensure that the disease spectrum that is most commonly seen when the test is used in clinical practice is mirrored by the study design.

The choice of a reference standard

Ideally a new test should be compared with a reference standard test that is 100 % sensitive and specific. This is rarely possible, but often a single test is sufficiently accurate to be used as the "gold standard" to which the new test is applied. A fecal antigen test to diagnose *Helicobacter pylori* infection may use a carbon-13 urea breath test as the gold standard as this has a sensitivity and specificity of 98 %, which many would consider sufficiently accurate [9]. If no single test is accurate enough to use as a "gold standard" a number of different tests can be applied, and a predetermined algorithm used to define the presence or absence of disease [10]. For example, there is no gold-standard test for gastroesophageal reflux disease (GERD) so investigators assessed patients presenting in primary care with upper gastrointestinal symptoms with endoscopy, 48-hour esophageal pH studies, as well as symptomatic response to proton pump inhibitor (PPI) therapy [11]. All patients with esophagitis at endoscopy were considered to have GERD. If they did not have esophagitis but had >5.5 % pH <4 and/or a symptom associated probability of ≥95 % then they were also considered to have GERD. Finally, if the pH data were equivocal (3.5–5.5 % pH <4) but the patient had a dramatic response to PPI therapy then again the patient was considered to have GERD [11].

There are many examples of gastrointestinal diseases where there is no test that can be used to define those with and without the disease. This issue has bedeviled the diagnosis of functional GI diseases and resulted in numerous iterations of expert bodies in defining the disease according to various symptomatic criteria. This is not a unique problem as this issue is faced by all psychiatric diagnoses. Psychiatry researchers have overcome this dilemma by using techniques that avoid the need for comparison with a single accurate test [12]. These techniques can be broadly divided into latent class analysis (LCA) and Bayesian analysis. Both have been used widely in psychiatry as well as other disciplines, but, as yet, have not been widely applied to gastroenterology.

Latent class analysis

Traditional regression techniques describe relations between observed variables. For example, a logistic regression model may suggest a relation between *H. pylori* and stomach cancer that is independent of other variables in the model, such as smoking, sex, or social class. Any variation of the data in this model that is not explained by these observed variables is assumed to occur at random. LCA postulates the existence of an unobserved categorical variable that divides the population of interest into classes (hence the term "latent class") [6]. Members of the population with a set of observed variables will respond differently, depending on the latent class to which they belong.

This technique can be applied to the problems related to diagnostic testing, with the unobserved categorical variable being "disease present" or "disease absent". The observed variables might typically be the results of diagnostic tests, none of them being a gold standard. LCA could then be applied in an attempt to divide the population into "true" positives and negatives. This approach has been shown to require at least three different types of diagnostic test [13]. LCA can then be applied to derive the proportions of patients in each latent class (i.e. estimated to be diseased or free of disease), and the sensitivity and specificity of each diagnostic test. Computer-intensive statistical methods are used to obtain standard errors of estimated parameters and the robustness of the data will, in part, depend on the sample size [14]. This type of analysis has been applied to a number of diagnostic problems but is rarely used in gastroenterology [15].

Bayesian analysis

Standard statistical tests assume that there are no prior expectations of the study results, in adherence with the principle that science is objective. This approach has been questioned because, in most scientific experiments, there is an expectation of what the outcome will be from prior knowledge, which should be incorporated in the analysis [16]. Indeed this is exactly what is done when calculating risk of disease using positive and negative likelihood ratios described earlier. Thomas Bayes proposed a theory over 200 years ago to try and overcome this problem. Bayes' theorem is a formula that describes how our existing beliefs (described as probability distributions) are changed by new study data [6]. Distributions of beliefs before new information becomes available are known as priors; those after the assimilation of new information are

posteriors. Prior distributions can be obtained from existing research evidence, expert opinion, or be set to be "uninformative" (a flat distribution that does not influence the analysis). Posterior distributions are described as means (or proportions) and credible intervals. Credible intervals are the Bayesian equivalent of confidence intervals [17].

The effect of study design on diagnostic test performance

All epidemiologic research has to contend with the design of the study introducing bias into the results. Diagnostic studies are no exception and there are a variety of biases that can influence the estimate of diagnostic test performance [18].

The study design can impact on the spectrum of disease in the population as discussed earlier. One approach to evaluating a diagnostic test is to conduct a case-control study. This is particularly useful where the disease is rare and is a much cheaper approach to analyzing test performance than evaluating a consecutive cohort of subjects that is representative of the population being studied. The problem is that case-control studies will almost certainly overestimate the sensitivity and specificity of the test because the subjects "with disease" are more clearly delineated from those "without disease" [19]. Case-control studies are useful as proof of concept for a new test, but consecutive patients in the correct setting are needed before the test can be used in clinical practice. Thus it may be appropriate to initially test a screening test for hemochromatosis in a tertiary care center, where these patients are overrepresented, but a study must be conducted in the general population before this test is used in a routine screening program.

Verification bias is another example of where the population being studied can be skewed so that patients with more clearly delineated disease are included [20]. This describes a protocol where only certain participants are subjected to the gold-standard test and the rest are assumed not to have the disease. An example of this would be only certain pancreatic lesions being referred for surgery where a definitive diagnosis is made from pathologic assessment of the resected specimen. This approach will only identify those with the most advanced disease and furthermore patients that do not have the gold-standard test will be assumed to be disease-free, whereas a few patients with more subtle indolent disease will be misclassified. There are also some studies that exclude patients as "equivocal" when the test under investigation does not give a clear result. This is an intuitive solution to the problem, but again is artificially structuring the sample to more clearly delineate those with and without the disease.

Lack of blinding is another key problem that can bias results [21]. The investigator applying the gold-standard test should be different from the researcher evaluating the test being studied, and both should be blinded to the results of the other. If this is not achieved then there is the possibility that the researcher could be influenced in their interpretation of the result, usually causing an overestimation of the accuracy of the new test [19,21].

Conclusions

This chapter has focused on objectively establishing the accuracy of a test. This is important as it informs the patient and the clinician of the diagnosis so that appropriate treatment can be instituted and the patient can be given some idea of the likely prognosis. Ultimately, however, a diagnostic test only has real value if it can be shown to improve more direct patient-related outcomes [22]. Establishing that the test leads to greater patient satisfaction, more cost-effective management or better health outcomes must be the true goal of any investigation.

References

1 Schoenfeld P, Guyatt G, Hamilton F, et al. An evidence-based approach to gastroenterology diagnosis. *Gastroenterology* 1999;116:1230–7.
2 Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, et al. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ* 2012; doi: 10.1136/bmj.e686.
3 Moayyedi P, Axon ATR. The usefulness of likelihood ratios in the diagnosis of dyspepsia and gastro-esophageal reflux disease. *Am J Gastroenterol* 1999;94:3122–5.
4 Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329:168–9.
5 Leeflang MMG, Bossuyt PMM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol* 2009;62:5–12.

6 Moayyedi P, Duffy J, Delaney B. New approaches to enhance the accuracy of the diagnosis of reflux disease. *Gut* 2004;53(Suppl IV):iv55–7.
7 Rembacken BJ, Fujii T, Cairns A, et al. Flat and depressed colonic neoplasms: a prospective study of 1000 colonoscopies in the UK. *Lancet* 2000;355:1211–14.
8 Sackett DL, Haynes RB. (2002) The architecture of diagnostic research, in *The Evidence Base of Clinical Diagnosis* (ed. JA Knottnerus), BMJ Books, London, pp. 19–38.
9 Ishihara S, Kaji T, Kawamura A, et al. Diagnostic accuracy of a new non-invasive enzyme immunoassay for detecting *Helicobacter pylori* in stools after eradication therapy. *Aliment Pharmacol Ther* 2000;14: 611–14.
10 Alonza TA, Pepe M. Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Stat Med* 1999;18:2987–3003.
11 Dent J, Vakil N, Jones R, et al. Accuracy of the diagnosis of GORD by questionnaire, physicians and a trial of proton pump inhibitor treatment: the Diamond Study. *Gut* 2010; 59:714–21.
12 Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a "gold standard". *Am J Pyschiatry* 1994;151:650–7.
13 Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. *Stat Med* 1986;5:21–7.
14 Formann A, Kohlmann T. Latent class analysis in medical research. *Stat Methods Med Res* 1996;5:179–211.
15 Christensen AH, Gjorup T, Hilden J, et al. Observer homogeneity in the histologic diagnosis of *Helicobacter pylori*. *Scand J Gastroenterol* 1992;27:933–9.
16 Bland JM, Altman DG. Bayesian and frequentists. *BMJ* 1998;317:1151.
17 Spiegelhalter DJ, Myles JP, Jones DR, et al. An introduction to Bayesian methods in health technology assessment. *BMJ* 1999;319:508–12.
18 Whiting P, Rutjes AWS, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004;8(25).
19 Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.
20 Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987;6:411–23.
21 Detrano R, Lyons KP, Marcondes G, et al. Methodologic problems in exercise testing research. Are we solving them? *Arch Intern Med* 1988;148:1289–95.
22 Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA* 1994;271:389–91.

12 Randomized controlled trials

Paul Moayyedi & Richard H. Hunt
Division of Gastroenterology, McMaster University, Hamilton, ON, Canada

Key points

- Randomized controlled trials (RCTs) minimize confounding and bias provided they are properly randomized, allocation of randomization sequence is concealed, all those involved in the trial are blinded, and all are followed to the completion of the trial.
- All these criteria cannot always be met and so whilst the RCT is the most rigorous study design individual trials may still be open to bias.
- Drug trials are divided into four phases depending on the stage of drug development.
- All trials must be run according to good clinical practice guidelines.
- There are many designs of RCTs including parallel group, factorial, crossover and cluster randomized trials.

Introduction

Previous chapters have described case-control and cohort studies. These are termed observational studies as subjects are simply "observed" and the researcher does not determine any intervention they receive. If the researcher decides which intervention a participant receives this is termed an experimental design [1]. In theory, a researcher could make a conscious decision as to what treatment a patient receives and this would be very similar to an observational study. For this reason this approach is rarely taken and usually the intervention is administered in a random fashion. This is the most powerful experimental design, as all confounding factors will be roughly equally distributed between groups if the sample size is sufficient and the intervention is truly random. A randomized design can also eliminate bias if the participants and researchers are masked as to the intervention received [2].

This type of design is not appropriate for all clinical questions [1]. There are numerous factors that impact on health and disease that simply cannot be randomized; for example, one cannot randomize someone to be old or young in order to study the effect of age on risk of colon cancer. Similarly, there are other interventions that could theoretically be randomized, but to do so would be unethical or inappropriate. For example, whilst one could randomize subjects to take 10 units of alcohol per day or abstain and measure the effect of this intervention on risk of cirrhosis, such a trial would not pass any ethical review. Nevertheless, there are many healthcare interventions where an RCT is the most robust form of evidence, and for this reason the RCT is considered the "gold-standard' experimental design [3].

The concept of an RCT is relatively straightforward but in practice the design and conduct is highly complex [4]. There are a number of books that describe the nuances of RCT design and it is not possible to cover

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

all of these in one chapter. We will therefore focus on key aspects of randomized control design, phases of trial development, and different trial designs.

The key building blocks of randomized controlled design

The key feature of an RCT is highlighted in the first word of the methodology; the interventions are *randomized*. This seems straightforward enough, but is actually more complex than many clinicians appreciate. Randomization has two key components: sequence generation and concealment of allocation [5]. Sequence generation refers to how the randomization list is created [6]. This list must be truly random and this can be achieved by using random number tables or computer-generated randomization sequences. The toss of a coin or drawing certain cards from a well-shuffled pack are in theory random but are rarely used in modern trials with easy access to computer-generated random lists. Methods such as first letter of the last name or date of birth are not random and such approaches are not recommended.

This approach will ensure the interventions are given in a random sequence but this on its own will not ensure that the groups are truly random. If the randomization list is placed on the clinic wall, for example, the investigator will know what treatment the next patient is going to be given. This may be in the form of drug A or drug B, so initially this will not help the researcher. However, as experience accumulates, the clinician can influence the trial by entering patients in a nonrandom fashion. For example, he or she may realize that most people getting drug A get better, whereas those receiving drug B do not. If he knows that the next treatment a patient entered into the study will receive is drug B, he or she may consciously or subconsciously decide that a patient with more severe disease is not eligible or should wait before being entered into the trial, and will wait until a patient with milder disease presents and then place them in the study. This approach will lead to unbalanced groups over time. To avoid this happening there must be concealment of allocation [7], so that the investigator does not realize what the next treatment is going to be. This is achieved by contacting an independent person or trial center that is not involved in any other way with the study for the next randomization code. If this is not possible then sequenced, sealed, and opaque envelopes containing the code can be used.

Correct sequence generation and concealment allocation will lead to roughly similar patients in each group, but only if the sample size is sufficiently large (a sample size of 100 for each group is usually considered sufficient). For smaller trials it is advisable to use a restricted randomization design to ensure that the groups are balanced for key characteristics [8]. Permuted-block stratified randomization is the standard restriction method. The investigator determines one or more factors that are important to balance between the randomized groups (e.g. severity of disease, age, or sex). Eligible patients are stratified by these factors and randomization is then conducted in a pre-specified block size between the groups. For example, in an RCT stratified for age with block sizes of four, participants might be stratified as <50 years or ≥ 50 years of age and then each block of four patients under the age of 50 would be randomized, with two allocated to treatment A and two to treatment B.

Randomization will minimize the risk of confounding factors influencing the outcome of the trial, and will also help to reduce some biases. However, bias will still occur with this approach if participants, researchers, or those looking after the patient are aware of what intervention has been allocated [9]. This is particularly the case when the main outcome being assessed is subjective (e.g. improvement in dyspepsia symptoms). If the participant or researcher (or indeed both) knows that they have been allocated to the "new" treatment they may want the trial to succeed and feel that their symptoms have improved when there has, in fact, been little change.

It is ideal for all those involved in the trial to be masked as to the intervention the participant is receiving. This is achieved by using an identical placebo (or alternative treatment) that is the same size, shape, and taste as the active intervention. There are RCTs where this is not ethical (e.g. in a trial comparing fundoplication with proton pump inhibitor therapy in gastroesophageal reflux disease it would not be ethical to do a "sham" surgery procedure in those allocated to the drug) or would not be appropriate, for instance when the aim of the study is to assess the effect of knowledge of the intervention to which patients have been allocated on the outcome (e.g. when evaluating

whether having a helpline for inflammatory bowel disease patients improves patient satisfaction, it makes little sense to "blind" patients to the fact they have access to a helpline). Whilst it may not be feasible to blind participants or clinical staff to the treatment allocation it is usually possible to blind researchers assessing the outcome [9]. This will help reduce some biases, so the study should try to blind researchers even if they cannot successfully blind all involved in the trial.

Bias in the trial may also occur if participants drop out during the course of the study and the characteristics of those lost to follow-up is different between the groups. The only way to absolutely ensure that this does not happen is to have 100 % or close to 100 % follow-up and analyze in an intention-to-treat manner (i.e. analyze *all* participants in the trial according to the group they were originally allocated to, irrespective of what treatment they actually received). Complete follow-up may not be realistic, particularly with large population studies that follow up participants over many years. To guard against the possibility of differential drop-outs the investigator can impute missing data in the analysis or do a sensitivity analysis (e.g. assume all those that are lost to follow-up are treatment failures, or assume all those that are in the active group are treatment failures and all those in the placebo group are treatment successes) to see what impact this will have on conclusions.

The fundamental principles of human research and ethics govern the conduct of RCTs. These include respect for the individual, the security and wellbeing of the individual, to accrue benefit to society and the patient, and lastly justice (or treating subjects or patients fairly, sharing risks and benefits equally).

The accepted principles of medical ethics apply to the conduct of the actual study and are guided by good clinical practice guidelines [10]. These include a proper research design, randomization, and an appropriate and ethical use of placebo with careful and safe monitoring of the trial for its entire duration. This may include the appointment of an independent safety monitoring committee. The trial should be run by skilled and experienced clinical investigators who understand the harms and benefits of participation in the study. The trial should use carefully prepared informed consent, give fair and equal opportunity in the selection of subjects or patients, and offer fair and appropriate remuneration for time and costs incurred. However, this remuneration should avoid any possibility of it being perceived as an incentive. Appropriate monitoring and medical supervision should be in place for the duration of the study, with a physician available at short notice if necessary.

All volunteers or patients who agree to participate in a clinical trial should give their informed consent. This requires the investigator to explain in careful detail the purpose of the study and outline the premise on which the protocol is based. A detailed explanation of what will be involved and any possible risks discussed, including the expected rate of adverse events. They should also explain that the study will be conducted in accordance with the Declaration of Helsinki, which was originally published by the World Medical Association in 1964, and has provided the framework of ethical principles for the conduct of medical research. It has been updated by eight amendments, the most recent in 2008. It is helpful to explain the benefits that will result from a successful study and how this will likely fit into the overall management of the disease. It is good practice for informed consent to be obtained by an experienced investigator rather than a junior research assistant.

Stages of drug trial development

To evaluate a new drug requires several stages of evaluation in humans after the preclinical and toxicology studies predict human safety and the appropriate initial dose to test. Thus, traditionally, initial drug testing in humans has been divided into four trial phases, I to IV [11].

Phase I studies explore the effects of a new drug in humans for the first time, usually in a small group of healthy volunteers. This offers the opportunity to determine the safety, tolerability, and the pharmacodynamic effect(s) of the drug, while obtaining essential information on the pharmacokinetics. These studies are usually performed in a small clinical investigation unit where all aspects of safety can be properly monitored. Phase I studies also include single ascending and multiple ascending dose evaluations and studies of possible food interactions.

Phase II studies follow phase I studies and provide the opportunity to study larger groups of healthy volunteers and also introduce patients to the prospective

evaluation of efficacy. The larger numbers also increase the knowledge about the drug, especially the safety profile. Phase II studies are increasingly referred to as phase IIA or IIB studies. Phase IIA studies refer to those which are designed to determine the optimal dose of the drug while phase IIB studies are designed to determine how well the drug works at the chosen dose. In an attempt to reduce the development time and costs, some phase II trials attempt to combine both phase I and phase II studies to explore the safety, tolerability, dosing, and efficacy together.

Phase III trials are usually conducted on large numbers of patients and are prospective randomized controlled trials involving multiple centers. These studies are designed to assess the effectiveness of the new treatment in comparison with the current gold standard. Their size depends on the disease being treated, the calculated numbers required to determine a difference with the current treatment, and they often run over one or more years, making them time-consuming and very expensive. It has been usual for regulatory authorities to require two positive phase III trials, which have demonstrated a new drug's efficacy and safety, in order to approve the drug.

Phase IV studies are not RCTs but a description of the pharmacologic surveillance that is conducted to ensure that the drug is safe once it is released on the market. These usually rely on databases and are conducted to identify rare but serious adverse effects that may relate to the drug, but which are too rare to be detected in phase III RCTs.

Key issues relating to phase II and phase III trials

Objective

The objective of each study should be clear and uncomplicated. However, the objectives for phase II and III are somewhat different. Phase II studies usually evaluate pharmacokinetics and pharmacodynamics of a drug as well as measuring dose response. They often evaluate mechanistic primary endpoints rather than key clinical variables. For example, initial proton pump inhibitor studies to treat gastroesophageal reflux disease would focus first on efficacy of suppression of gastric acid measured through intragastric pH, and then healing of esophagitis, before finally evaluating resolution of heartburn symptoms as part of phase III studies.

These objectives often overlap so there are instances where phase II and III studies are combined. This may be appropriate when the sample sizes to meet the needs of phase II and III questions are similar but often the pressure to combine these phases relate to market forces on pharmaceutical companies. The financial consequence of a delay in release of a successful drug can be enormous. There are pressures to move to phase III trials for regulatory approval as soon as possible, and whilst pharmaceutical companies will be concerned with safety there may be the pressure to truncate phase II development and incorporate some of the aims with phase III before safety is fully established or the optimum dose completely elucidated.

Number of subjects

The sample size in phase II studies is generally much lower than in phase III studies. The number depends on the design and whether the study is an early phase II or later study. Moreover, placebo is often used as the comparator in phase II studies. For the simpler single stage design the numbers will also be less than a two stage, multiple stage or group-sequential study design. The usual principles of number calculations apply to phase II studies, as in other studies, with respect to the significance levels, power, confidence intervals, and so on. A randomized phase II study design may involve more than one dose of the test drug in addition to placebo or an active comparator such as the standard treatment. With such a design the patient numbers will increase substantially.

Inclusion/exclusion criteria

When undertaking phase II studies of a new drug treatment only a small number of patients will have been exposed to the drug during the initial phase I studies, and so it is common to do at least the early phase II studies in healthy volunteers. Each should be subjected to an extensive medical history and full physical examination, comprehensive blood work to exclude abnormalities of the hemopoietic, immunologic, hepatic and renal systems, and an EKG as minimal criteria for inclusion. It may be necessary to exclude past conditions, depending on the drug and systems being

evaluated. The aim of these strict inclusion criteria is to achieve *internal validity*; ensuring that participants are as similar and as well characterized as possible so that the trial is as well controlled and as rigorous as possible. The problem with narrow entry criteria is that the results then tend to apply to a very select group of individuals and may not relate to the general population.

In later phase II studies patients may be included, although the pharmacokinetic and pharmacodynamic endpoints are the same, or similar, to those conducted in healthy volunteers. The advantage of phase II studies in patients is that they can determine whether the pharmacokinetics or pharmacodynamics differ in the presence of disease, where the pathophysiology is different to that in healthy volunteers, such as the diastolic pressure in hypertension, or basal and stimulated gastric acid secretion in patients with duodenal ulcer. Phase III studies typical have less restrictive inclusion criteria, as the purpose is to evaluate whether the drug will be effective in a wider group of patients. Here *external validity* becomes more important; namely, that the trial evaluates whether the drug will be effective in a more general population.

Types of randomized controlled trial

There are a number of different designs for RCTs depending on the aim of the study; examples include parallel group, factorial design, or crossover and cluster randomized trials [12]. This list is not exhaustive, but these examples have been chosen as they are the most commonly used. The most classic design is the parallel group trial. Patients are randomized to two or more intervention groups and remain in those groups until the end of the trial, when outcomes are assessed. This design is usually used to establish whether one intervention is equivalent or superior to another (or to placebo).

A factorial design is similar to a parallel group design in that two or more interventions are evaluated separately but they are also evaluated together. For example, in a 2×2 factorial design evaluating drug A and drug B, one group will get all placebos, another will be given active drug A and placebo drug B, the third group will be allocated placebo drug A and active drug B, and the fourth group will be given both active drugs. The advantage of this design is that it is a more efficient approach to assessing a number of interventions at the same time and it can also evaluate whether there is interaction between treatments.

A crossover design is different from the parallel group trial in that the same patients are given both interventions, one after the other in a random order, usually with an intervening washout period. The advantage of this design is that the patients act as their own control. This reduces the number of subjects needed for the trial and, furthermore, as the patient obviously has the same characteristics from one treatment period to the next there is less variability in participants and so the statistical analysis has more power. This approach is useful for rare disease for these reasons, but has the disadvantage that there may be a carry-over effect of the previous treatment. The washout period, where the patient is receiving no drug is designed to minimize this, but sometimes it is not possible to know if a carry-over effect still exists. For example, if one is evaluating antidepressant therapy versus placebo in irritable bowel syndrome, some of the observed effects may relate to the treatment of anxiety or depression. If the patient is allocated to receive the antidepressant first, the impact on psychological parameters may be long-lasting and, when the participant is switched to placebo, the benefit may continue, giving the spurious impression that placebo is also effective. Indeed if the treatment one is assessing permanently cures the disease then a crossover design cannot be used.

A cluster randomized trial refers to designs where the unit of randomization is not the subject, but some other variable such as a primary care center or a clinician. This is an appropriate design where the intervention is not at the subject level but the outcome is related to the patient. For example, an investigator may want to assess whether primary care clinician education regarding the best approach to managing risk of nonsteroidal anti-inflammatory drug complications reduces peptic ulcer bleeds. The best design for this is a cluster randomized trial where primary care doctors are randomized to have an education package or no intervention. The trial will then follow up whether patients treated by clinicians having the education package have less peptic ulcer bleeds than those receiving no education. The important point to take home from this type of design is that the analysis is more complex as one cannot treat this as an individual patient trial.

Conclusions

The RCT has transformed the evaluation of medical interventions. Its conduct and design depends on the question being evaluated, and choosing the right methodology can be challenging. Nevertheless, this approach is more rigorous than any other type of study design, and can allow the clinician to be confident that a treatment is effective even when the impact is relatively modest.

References

1 Stolberg HO, Norman G, Trop I. Randomized controlled trials. *Am J Radiol* 2004;183:1539–44.
2 Kendall JM. Designing a research project: randomized controlled trials and their principles. *Emerg Med J* 2003;20:164–8.
3 Altman DG, Bland JM. Treatment allocation in randomized trials: why randomise? *BMJ* 1999;318: 1209.
4 Jadad A. (1998) *Randomized Controlled Trials*, BMJ Books.
5 Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing quality of controlled clinical trials. *BMJ* 2001;323:42–6.
6 Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002;359:515–19.
7 Forder PM, Gebski VJ, Keech AC. Allocation concealment and blinding: when ignorance is bliss. *Med J Aust* 2005;182:87–9.
8 Lachin JM, Matts JP, Wei LJ. Randomization in clinical trials: conclusions and recommendations. *Control Clin Trials* 1988;9:365–74.
9 Haahr MT, Hróbjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials* 2006;3:360–5.
10 http://ichgcp.net (accessed October 15, 2012).
11 Stanley K. Design of randomized controlled trials. *Circulation* 2007;115:1164–9.
12 Hopewell S, Dutton S, Yu LM, et al. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010; doi: 10.1136/bmj.c723.

PART FOUR

Epidemiology of Major GI Diseases

13 Epidemiology of GERD, Barrett's esophagus and esophageal cancer

David Armstrong
Farncombe Family Digestive Health Research Institute and Division of Gastroenterology, McMaster University, Hamilton, ON, Canada

Key points

- The prevalence of GERD is increasing worldwide but it remains greater in Western countries than in other regions.
- The incidence of esophageal adenocarcinoma continues to increase worldwide.
- The risk of esophageal adenocarcinoma in nondysplastic Barrett's esophagus is lower than had been reported previously.
- Epidemiologic studies of GERD and Barrett's esophagus are hampered by the absence of standardized criteria for the diagnosis of symptomatic GERD and the finding of erosive reflux esophagitis and Barrett's esophagus in asymptomatic individuals.
- Further epidemiologic and mechanistic studies are needed to elucidate the role of obesity, alongside lifestyle and environmental factors in the pathogenesis of GERD and its sequelae.

Introduction

Gastroesophageal reflux disease (GERD), the most common cause of esophageal symptoms and injury, affects about 10–20 % of the population [1] although there are widespread geographical differences in diagnosis, incidence, and sequelae [2]. GERD is associated with impaired health-related quality of life [3], decreased global health and productivity and burgeoning societal costs. GERD may progress to Barrett's esophagus (BE) and, subsequently, esophageal adenocarcinoma (EAC); the prevalence of EAC and esophagogastric junction adenocarcinoma (EGJAC) is increasing more rapidly than any other common malignancy. GERD is not a risk factor for esophageal squamous cell carcinoma (ESCC) which is more prevalent worldwide, but less prevalent in the Western world than EAC or EGJAC.

Gastroesophageal reflux disease

GERD occurs "when the reflux of gastric contents causes troublesome symptoms and/or complications" [2]. Epidemiologic studies in GERD are hampered by the absence of a diagnostic gold standard. Diagnosis is, generally, based on the presence of heartburn or regurgitation [2,4,5] but GERD is associated with a range of symptoms and neither symptom-based strategies nor objective tests deliver more than modest accuracy [6]. A systematic review of symptom-based studies reported a GERD prevalence ranging from 6.5 to 9.5 % when the diagnostic criterion was GERS at least

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

weekly up to nearly 30 % when various other criteria were used [7].

Prevalence of GERD

Gastroesophageal Reflux Symptoms (GERS) and Erosive Reflux Esophagitis (ERE)

Symptom-based epidemiologic studies cannot determine whether subjects have injury and endoscopy-based studies cannot determine that a normal endoscopy precludes prior ERE, especially if there has been prior acid suppression therapy. Patients with ERE and nonerosive reflux disease (NERD) differ with respect to esophageal acid exposure, histology [8], cytokine profiles [9], and clinical features [10,11]. However, it is difficult to diagnose NERD or functional heartburn (FH) definitively and determine their prevalence [12]; thus, although FH may constitute less than 10 % of the population with heartburn, it is difficult to ascertain whether ERE and NERD are different conditions.

The reported prevalence of GERD ranges from about 5 % in Asia to 10–20 % in the Western world [1]. The prevalence of GERS, occurring at least weekly, varies from 2.5 % to 6.6 % in eastern Asia, from 9.3 % to 20 % in western and southern Asia, from 10.3 % to 25.9 % in northern Europe, from 7.7 % to 9.8 % in southern Europe, and from 13.8 % to 28.8 % in North America [13].

Secular trends in GERD

In Norway, GERS prevalence increased by 47 % between 1995–1997 and 2006–2009, cumulative incidence rates exceeding the cumulative spontaneous loss rates [14] (Figure 13.1). In the United Kingdom GPRD, GERD prevalence increased with age [15] although this finding was not supported by later studies [16]. In Asia, GERD prevalence increased two- to fourfold over 10 years (Figure 13.2) [17]; in Taiwan, GERD incidence increased over 6 years [18] and, in Japan, too, GERS prevalence increased from 1990 to 2006 [19].

In a prospective German cohort study, 24.7 % of NERD patients developed ERE over 5 years whilst 60.8 % of ERE patients regressed to NERD [20]; however, as most patients received therapy, this, like other reported trends, may not reflect the natural history of GERD.

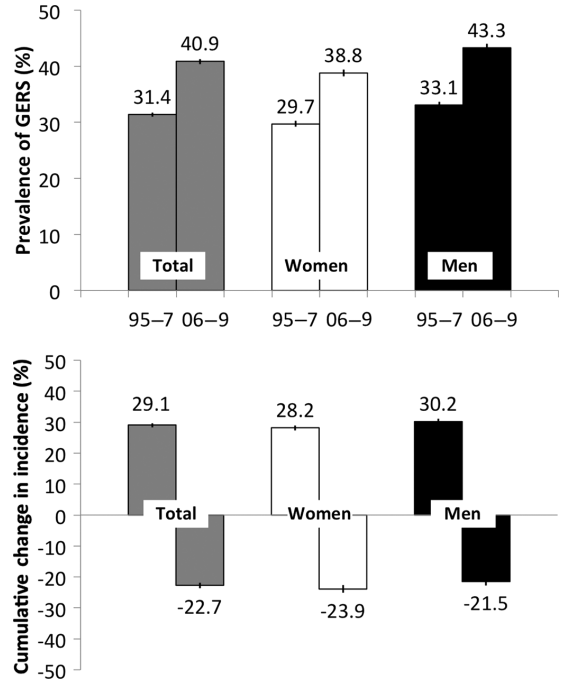


Figure 13.1 Increase in the prevalence of gastroesophageal reflux symptoms (GERS) (upper panel), related to the spontaneous incidence and loss of symptoms (lower panel) between 1995–1997 (95–7) and 2006–2009 (06–9) for all study subjects, women and men in Norway. Adapted from Ness-Jensen et al. 2012 [14].

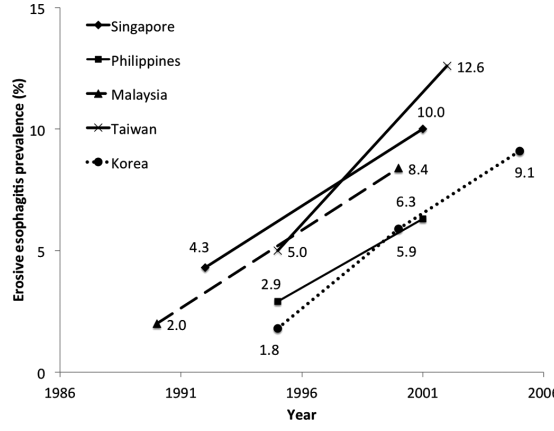


Figure 13.2 Increases in the prevalence of erosive reflux esophagitis (ERE) in various Asian countries between 1992 and 2005. Source: Goh 2011 [17]. Reproduced with permission of John Wiley & Sons.

Overlap between GERD and other diseases

Diagnostic overlap between GERD, dyspepsia, functional dyspepsia (FD), and irritable bowel syndrome (IBS) [21] may arise due to the coincidence of common unrelated conditions, the presence of common pathogenetic mechanisms, the generation of symptoms typical of multiple different conditions [22], or a lack of symptom specificity. Pathologic acid exposure, reported in 31.7 % of Chinese FD patients [23] suggests that these patients' true diagnosis may have been GERD, atypical FD or both GERD and FD, despite their symptom profile. Strict application of diagnostic criteria in Japanese individuals led to reduced diagnostic overlap (1 %) for GERD and FD and a diagnosis in 25 % whereas less strict application increased both the diagnostic rate (54.3 %) and the overlap (21.7 %) [24]. Currently, neither objective tests nor validated questionnaires can differentiate reliably between NERD, FH, and FD [25].

Risk factors for GERD

Reported risk factors for GERD include age, sex, ethnicity, obesity, smoking, alcohol, dietary factors, *Helicobacter pylori*, NSAIDs, hiatus hernia, and nationality (Figure 13.3), but the presence and strength of the associations have been inconsistent [26–31] and unhelpful in elucidating the pathogenesis of GERD.

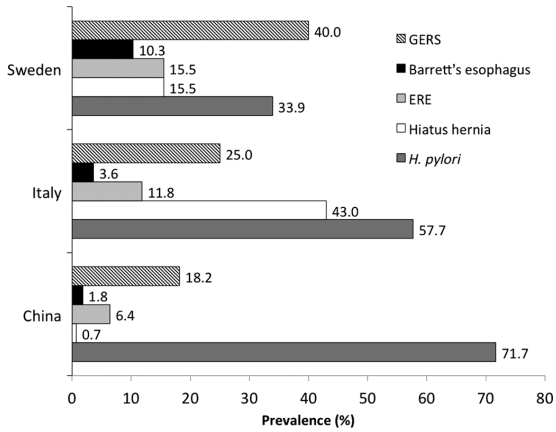


Figure 13.3 The prevalence of gastroesophageal reflux symptoms (GERS), Barrett's esophagus, erosive reflux esophagitis (ERE), hiatus hernia and *Helicobacter pylori* infection in three population-based endoscopy studies [28–31].

Seasonal variations in GERD incidence in Taiwan [18] were associated, in some individuals, with changes in humidity but other factors, such as intake of nitrates [32], ascorbic acid, and fruit and vegetables [33] may also vary seasonally and with respect to geographical location [34].

GERD, asthma, and post-traumatic stress disorder (PTSD) were reported in persons exposed to the World Trade Center (WTC) terrorist attacks of 11 September 2001 [35]. Exposed individuals reported cumulative incidences for new onset and persistent GERS of 20 % and 13 %, respectively, at 6 years [36] whilst rescue and recovery workers reported a cumulative incidence of 39.3 % for GERD, 42.3 % for sinusitis, and 27.6 % for asthma [37]. GERS were associated, in both groups, with greater exposure to WTC debris but the relative importance of environmental exposure and stress was unclear. In the United Kingdom, patients with depression had a higher incidence of GERD than controls and tricyclic antidepressants were associated with a greater risk of GERD [38].

Helicobacter pylori may protect against GERD by inducing gastritis and reducing the acidity of GER [39]. *H. pylori* prevalence is often associated inversely with GERD [40] but there is significant heterogeneity between studies [41]. In Japan, *H. pylori* eradication was associated with reduced GERS [19]. A meta-analysis identified no association between *H. pylori* eradication and new onset GERD [42] although ERE was associated with *H. pylori* eradication in patients with prior peptic ulcer disease [42] suggesting a complex, time-dependent relationship between *H. pylori* infection and GERD, ERE, BE, and EAC. Because upper GI symptoms are not diagnostic, there may be misattribution of symptoms and diagnoses in studies of *H. pylori* eradication or NSAID use [26].

GERD is more prevalent and more severe in older individuals and in males [14,27,43] but the reported magnitude of the risks differs between studies.

Many studies have reported an association between obesity and ERE in adults [44,45] and between obesity and GERS [46]. Furthermore, increased BMI over time is associated with a greater incidence of new onset GERS [47]. BMI, waist circumference, and waist-to-hip ratios may prove to be suboptimal indicators of obesity if visceral adiposity [48] or insulin resistance [49] are confirmed as risk factors for GERD.

Diet is implicated in the pathogenesis of GERD [50,51] but no single factor or food group is dominant

or readily identifiable [52]. Dietary fiber may reduce GERS [53,54] whilst starch intake is positively- and sugar intake is negatively associated with ERE [55]. GERS are more common in newly diagnosed celiac disease patients than controls but, despite a diet-related decrease in anti-TTG titres [56], the associated reduction in GERS may be attributable to reduced intake of dietary carbohydrates. Because of the diagnostic overlap between GERD, EoE, and other GI disorders, GERS may respond to dietary modification for EoE [57] or other food sensitivities [58] even if dietary factors do not affect GER.

Familial clustering occurs for GERS and increased acid exposure in relatives of BE patients [59] and the prevalence of ERE is higher in patients who have a family history of GERD [60]. Concordance for GERD is higher in monozygotic than in dizygotic twins [61,62], and although the underlying mechanisms are unknown, genetic studies have linked GERD to polymorphisms for interleukin-1 gene (IL-1B) and IL-1RN associated with *H. pylori*-related inflammation [63] as well as for 4-aminobutyrate aminotransferase (ABAT) related to LES function [64].

Barrett's esophagus

Barrett's esophagus is defined, in North America, as "a change in the distal esophageal epithelium of any length that can be recognized as columnar type mucosa at endoscopy and is confirmed to have intestinal metaplasia by biopsy of the tubular esophagus" [65] and, in the United Kingdom, as "an endoscopically apparent area above the oesophagogastric junction that is suggestive of Barrett's which is supported by the finding of columnar lined oesophagus on histology. The presence of areas of intestinal metaplasia (IM), although often present, is not a requirement for diagnosis" [66]. Endoscopic features of BE in the absence of histological confirmation have been termed "endoscopically suspected esophageal metaplasia (ESEM)" [2]. Differences in definition [2] are important considerations when comparing BE prevalence data between studies.

Prevalence of BE

Generally, BE presents in older adults [65] and is more prevalent in Western countries, ranging from 10.3 % (Sweden) and 3.6 % (Italy) to 1.8 % (China) in population-based endoscopy studies and from 0.0 % to 3.4 % in Asian health-check studies [28]. For individuals without symptom-defined GERD, the prevalence of ESEM was 9.4 % (Sweden), 2.8 % (Italy) and 1.8 % (China) [28] making it difficult to estimate the true population prevalence of BE.

In a computer simulation, the US population BE prevalence was estimated at 5.6 % [67]; this estimate needs confirmation as it is, for example, higher than the prevalence of ESEM (5.1 %) and BE (2.4 %) in Canadian primary care patients [68] and higher than the prevalence of histologically confirmed BE in Sweden (1.6 %) [69], Finland (1 %) [70] and Holland (0.8 %) [71]. Some variability may be due to national differences, as histologically confirmed BE is less prevalent in Swedish GERD patients (2.5 %) [28] than in German GERD patients (4.9 %) [72]. The reported prevalence of BE in Asia – ranging from less than 1 % in China [73], Japan and Korea [74], to 1–2 % in Malaysia, Singapore and Turkey, 3.7 % in Iran, 3.8 % in Taiwan [75], and 7.3 % in Egypt [76,77] – may have been affected by the definition of BE; for example, depending on diagnostic criteria, short segment BE prevalence in Asia ranges from 0.04 % to 37.4 % whilst long segment BE prevalence ranges from 0.02 % to 6.6 % [17].

In Germany, progression to endoscopic or confirmed BE occurred, over 5 years, in 9.7 % of GERD patients who had had NERD or ERE at baseline [20] suggesting an incidence of about 1.8 % annually.

Risk factors for BE

GERD is associated with an increased risk of BE, as are bile reflux, hiatus hernia, impaired LES function, and GERS frequency. ERE is associated with a greater risk than NERD of developing BE in Korea [74], the United Kingdom [78] and Sweden (Relative Risk ratio: 5.2; 1.2 to 22.9) [79]. GERS are associated with a fivefold increased risk of long segment BE but not of short segment BE [80].

Males have a two- to threefold greater risk of BE than females [43,78]; age [78] and obesity are also associated with BE [81]. Risk factors for BE are similar to those for ERE [43].

In the CORI endoscopic database, BE was more prevalent in Whites (5 %) than Hispanics (2.9 %), Asians and Pacific Islanders (1.8 %), and Blacks

(1.5 %) [82]. Ethnic differences in BE prevalence may be environmental in origin, rather than genetic; other factors including diet, fiber intake [55], body habitus, and geographic effects may also be relevant. Tobacco smoking is, generally, considered to increase the risk of BE although this was not supported by a US case-control study [83]; similarly, alcohol is not a consistent, independent risk factor [84,85].

The role of dietary factors remains unclear [86], perhaps because dietary assessments are imprecise. In a US study, high consumption of vegetables or fruit was associated with a lower BE risk [87], but in Northern Ireland, combined intake of vitamin A, vitamin C, carotenoids, and selenium was not [88], whilst in Australia, folate intake was associated with an increased risk of BE or EAC [89].

Overweight and obesity are associated with BE [45] although there is evidence of a threshold effect rather than a dose-response relationship for obesity and BE [81]. Central obesity and metabolic syndrome are more common in long segment than in short segment BE [90].

NSAIDs and statins are associated with a reduced risk of progression to EAC but there is no clear evidence that they affect the development of BE [91,92].

BE is more common than expected in first-degree relatives of patients with BE [93,94] and those with EAC or high-grade dysplasia (HGD) [95].

H. pylori is associated with a decreased risk of BE [96]; however, although *H. pylori* is associated with a lower risk of BE (OR 0.27), particularly for Cag A +ve strains (OR 0.08), its effect may be indirect as it is not an independent risk factor for BE compared with GERD controls [97].

Esophageal cancer

Esophageal cancers are divided, based on histological criteria, into squamous cell carcinoma and adenocarcinoma although they differ, also, with respect to etiology, epidemiology, and natural history.

Prevalence of esophageal cancer

Esophageal cancer is the eighth most common cancer worldwide, with 481,000 new cases (3.8 % of the total) estimated in 2008, and the sixth most common cause of death from cancer with 406,000 deaths (5.4 % of the total) [98].

Until 40 years ago, ESCC was the most common esophageal malignancy but the incidence of EAC has increased in the United States, from 0.4 to 3 per 100,000 person-years over the last 35 years [99], and across the Western world [100], as has the incidence of EGJAC [101,102].

The age standardized incidence rates (ASIR) of esophageal cancer vary internationally more than 15-fold in men (22.3 per 100,000 in southern Africa compared to 1.4 in western Africa), and almost 20-fold in women (11.7 in southern Africa compared to 0.6 in Micronesia/Polynesia) [98] (Figure 13.4).

EAC is now the most common esophageal malignancy in the West; in Asia, ESCC remains dominant although EAC incidence is increasing. In the US, esophageal cancer incidence has increased only in Whites, between 1977 and 2005; ESCC rates have decreased in virtually all racial and ethnic groups during the same period [99] whilst EAC has increased in both white and black males [103].

In Singapore, the ASIR (per 100,000 persons) for ESCC fell from 8.3 to 3.9 for men and from 3.4 to 0.81 for women, whilst those for EAC rose from 0.0 to 0.54 for men and from 0.03 to 0.13 for women [104]. In Sweden, the incidence of EAC increased by 2.6 % annually from 1970 to 1993, and by 11.5 % annually from 1993 to 2001; however, the incidence has been stable since 2001 (1.1 % annually) with a decrease over the last few years [105]. Data from the US SEER database suggested that the incidence had decreased from 8.2 % annually before 1996 to 1.3 % annually in subsequent years [106]. However, subsequent analyses concluded that EAC incidence in men had increased between 1994 and 2008 in Australia and between 1998 and 2008 in the United States whilst remaining stable in Sweden, and that EAC incidence in women had increased steadily from 1994 to 2008 in all three countries [107]. Regional cancer registry data from England and Wales showed a threefold increase in incidence for men and women between 1971 and 2001 [108].

Risk factors

The incidence of both ESCC and EAC [109] increases with age although ESCC generally presents about a decade later than EAC [110]; both cancers are

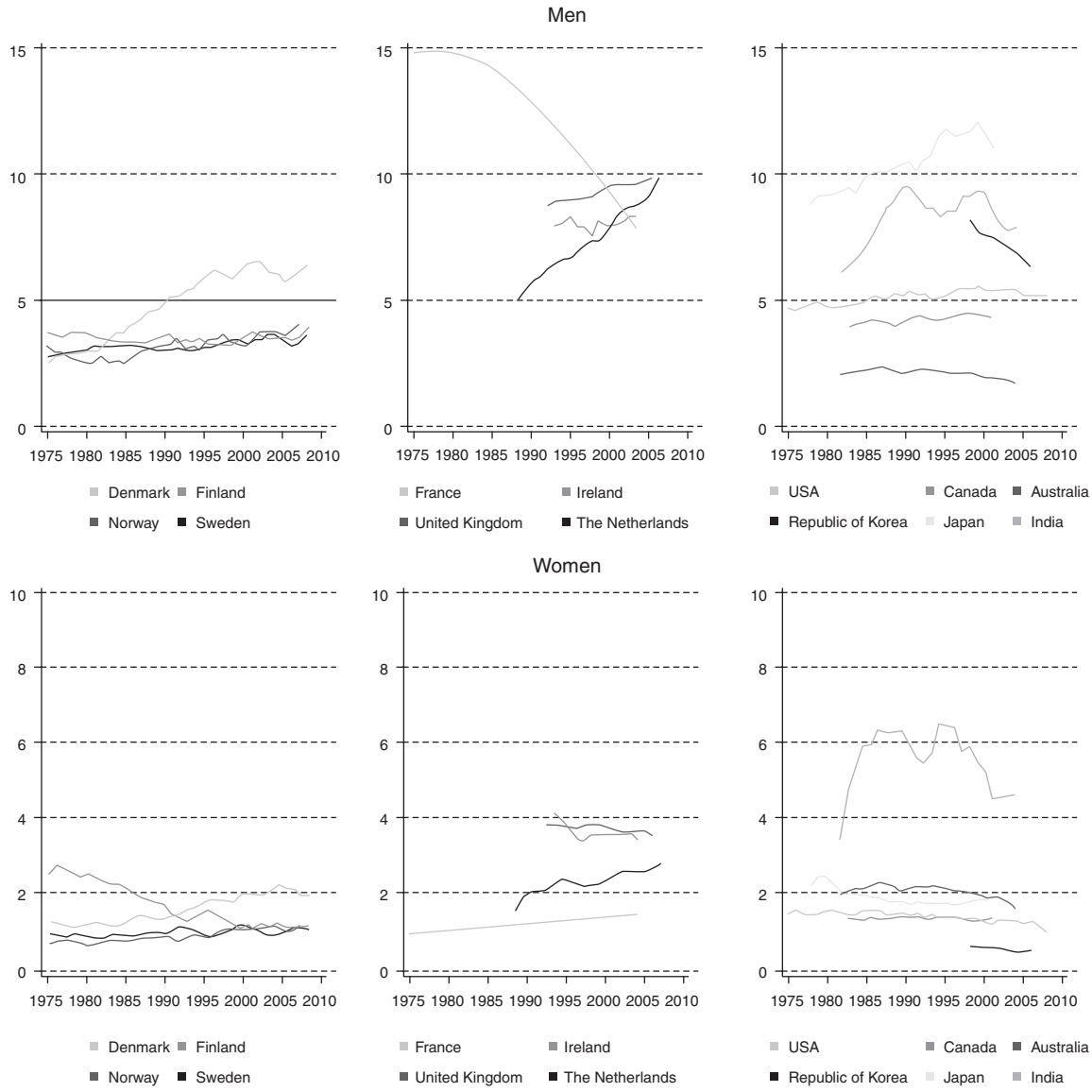


Figure 13.4 International trends in the incidence of esophageal cancer (age-standardized rate per 100,000 population) for men (upper 3 panels) and women (lower 3 panels). Source: Adapted from Ferlay J, Shin HR, Bray F, et al. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet], International Agency for Research on Cancer, Lyon, France; 2010.

more common in males [98,110] but the relative risk is lower for ESCC than for EAC [101] (Table 13.1). Breastfeeding may be protective against EAC [111].

Lifestyle factors, including obesity, smoking, and alcohol have been implicated as risk factors for the development of EAC although the role of alcohol is less clear for EAC than for ESCC [112–114]. Data from Northern Ireland confirmed the increased risk of EAC in current smokers but did not show an increased risk related to alcohol ingestion [109]. In an Australian study, heavy smokers had a higher risk of

Table 13.1 Potential risk factors for the development of esophageal cancers

Esophageal squamous cell carcinoma (ESCC)	Esophageal adenocarcinoma (EAC)
Geography/Race • Southeastern Africa • "Esophageal cancer belt" ○ Turkey ○ Southern Russia ○ Northern China • USA: black > white males	Geography/Race • Western Europe • Australia • North America • USA: white > black males
Gender • Male	Gender • Male
Age	Age
Tobacco	Tobacco
Alcohol	Obesity • Increased BMI • Central obesity
Helicobacter pylori • Presence	*Helicobacter pylori* • Absence
Nutritional factors • Thermal injury • *N*-nitroso compounds (nitrates) • Betel quid • Vitamin and mineral deficiencies • Decreased fruits and vegetables	Nutritional factors • *N*-nitroso compounds (nitrates) • Vitamin and mineral deficiencies • Decreased fruits and vegetables • Decreased fiber and antioxidants
Heredity • Tylosis ○ Howel–Evans syndrome • Family history of ESCC	Heredity • EGF polymorphisms • Family history of EAC or BE • MEN Type 1 or Zollinger–Ellison syndrome
Medical history • Caustic injury • Achalasia of the cardia • Current or prior ESCC • Plummer–Vinson syndrome • Zenker diverticulum	Medical history • Gastroesophageal reflux disease ○ Duration, symptom frequency • Barrett's esophagus ○ Length, dysplasia • Medications ○ No ASA, NSAIDs or statins ○ LES inhibitors • Myotomy or dilation of LES • Scleroderma

EAC than did nonsmokers in the presence of frequent GERS [115].

Based on the hypothesis that vitamin D reduces cancer risk, Australian subjects enrolled in a national, population-based, case-control study were evaluated for lifetime exposure to solar ultraviolet (UV) radiation; patients with EAC (OR 0.59; 95 % CI 0.35–0.99) or EGJAC (OR 0.55; 95 % CI 0.34–0.90) but not those with ESCC (OR 0.91; CI 0.51–1.64) were less likely than population controls to have high levels of lifetime exposure to solar UV radiation [116] (Figure 13.5).

Overweight and obesity are associated with an increased risk of EAC [45] although the magnitude of

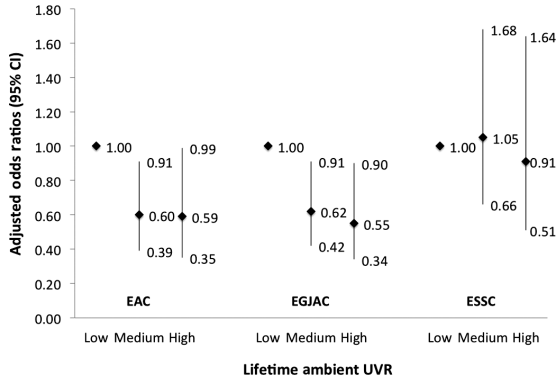


Figure 13.5 The association between cumulative lifetime ultraviolet radiation (UVR) exposure and the risk of esophageal adenocarcinoma (EAC), esophagogastric junction adenocarcinoma (EGJAC), and esophageal squamous cell carcinoma (ESCC) presented as odds ratios (95 % CI) relative to low exposure (reference), adjusted for age, sex, body mass index, education, state, reflux symptoms, smoking, alcohol, and *H. pylori*. Source: Tran et al. 2012 [116]. Reproduced with permission of Nature Publishing Group.

the risk in individual studies is quite variable [117], ranging from 2.27 [118] to 11.3 [119]. Although obesity is associated with a number of malignancies, the association with EAC is greater suggesting that there may be a specific pathogenic mechanism (e.g. GERD) by which adiposity increases risk. The association between obesity and EGJAC is weaker and, if anything, obesity is inversely related to the risk of ESCC [117].

Data from 1940 to 2007 in Connecticut suggest that the initial rise in EAC incidence predated the increase in obesity in the United States by over a decade [120]. In addition, a computer simulation model suggests that only a small proportion (6.5–7.6 %) of the rise in EAC incidence is attributable to secular trends in obesity [121]. However, weight and BMI may be suboptimal indices of obesity; abdominal obesity, defined by waist circumference, was associated with an increased risk of EAC, even for subjects with a normal BMI [122] suggesting that other obesity-related factors, possibly related to the metabolic syndrome, may promote EAC and other cancers [117].

The role of dietary factors in the pathogenesis of EAC is poorly understood. The risk of EAC is related, inversely, to dietary fiber intake and total carbohydrate intake, but foods with a high glycemic index are associated with an increased risk [55]. Higher toenail clipping concentrations of the trace elements zinc and cobalt were associated with BE but not EAC, whilst there was no risk reduction with increased levels of selenium [123]. In a population-based, case-control study, frequent GERS were associated with 6.4-fold (EAC), 4.6-fold (EGJAC), and 2.2-fold (ESCC) increased risks of cancer [115]. A meta-analysis indicated that the risk of EAC was increased in individuals who had weekly (OR 4.92) or daily (OR 7.40) GERS [124].

Consistent with a recent meta-analysis [125], a population-based, case-control Australian study [126] reported a reduced risk of EAC (OR 0.45) and EGJAC (OR 0.41) but not ESCC (OR 1.04) in association with *H. pylori*. However, unlike previous studies on *H. pylori* and gastric cancer [127,128], IL-1B or TNF-a polymorphisms did not modify the *H. pylori*-related risk of EAC or EGJAC [126].

ASA and NSAIDs were associated with reduced risks for EAC, EGJAC, and ESCC in an Australian study [129] and with reduced risks of EAC alone [130–132]. However, ASA and non-ASA NSAIDs [133], celecoxib [134] and ASA [135] have shown mixed results or no benefit in other studies although a meta-analysis did conclude that ASA or non-ASA NSAIDs were associated with a lower rate of esophageal cancers [133]. In two studies [131,132], statins were associated, independently, with decreased progression from BE to EAC. A subsequent, prospective cohort study, adjusted for NSAID use, also demonstrated a lower risk of EAC in BE patients taking statins [136]. Further studies are needed to elucidate the relevant mechanisms and determine, prospectively, whether ASA or statins affect progression from BE to EAC and there is an ongoing randomized control trial evaluating whether aspirin reduces the risk of EAC in BE patients [137].

Although polymorphisms for IL-1B and TNF-a were not associated with EAC, other single nucleotide polymorphisms in cancer-related genes involving apoptotic pathways (NOS3, BCL2, and CASP8) are associated, in a cumulative "dose-dependent" manner, with the early onset EAC [138]. Similarly, individuals who are homozygous for the A/A TNF-beta genotype have a greater risk of developing BE and EAC than the healthy population [139].

Esophageal squamous dysplasia (ESD) is a precursor lesion to ESCC [140] whilst intestinal metaplasia (IM), characteristic of BE, is a precursor lesion to EAC

[141]. The risk of EAC is 30 to 40 times greater in BE patients than in the general population [142]; however, pooled estimates of EAC incidence in BE, ranging from 0.41 % to 0.61 % annually, have been subject to publication bias and lack of adjustment for baseline dysplasia [141,143]. A recent meta-analysis, excluding patients with baseline dysplasia, reported lower pooled EAC incidence rates of 0.33 % annually for all BE and 0.19 % annually for short segment BE [143]. The risk of EAC increases 10- to 20-fold if HGD is present with a weighted annual incidence of 6.58 % [144] although progression from BE to EAC is not inevitable. Data from the Netherlands suggest that progression is slow and that the "incubation period" for EAC in BE is greater than 30 years [145].

Multiple choice questions

1 Epidemiologic studies of the prevalence of gastroesophageal reflux disease:
- **A** Indicate that the prevalence of GERD is increasing predominantly in Asia
- **B** Use standard criteria for diagnosis of GERD
- **C** Indicate that the prevalence of GERD is increasing worldwide
- **D** Use standard criteria for recording the frequency of GERD symptom occurrence
- **E** Indicate that the prevalence of GERD is increasing predominantly in North America and Europe

2 Which of the following has been identified, unequivocally, as a risk factor for Barrett's esophagus:
- **A** *Helicobacter pylori* infection
- **B** Symptomatic gastroesophageal reflux disease
- **C** Folic acid deficiency
- **D** Body mass index
- **E** Alcohol consumption

3 The pooled incidence of esophageal adenocarcinoma in Barrett's esophagus, in the absence of dysplasia or short-segment disease, is:
- **A** Greater than 1 % per annually
- **B** Between 0.6 % and 1.0 % annually
- **C** Between 0.4 % and 0.6 % annually
- **D** Between 0.2 % and 0.4 % annually
- **E** Less than 0.2 % annually

4 The incidence of esophageal adenocarcinoma:
- **A** Is greater in women than men
- **B** Is decreasing to a greater extent than that of esophageal squamous cell carcinoma in Asia
- **C** Is now stable in Sweden, having decreased steadily over the last 20 years
- **D** Is now greater than any other esophageal malignancy in the West
- **E** Is now greater than esophageal squamous cell cancer in Asia

References

1 Dent J, El-Serag HB, Wallander MA, et al. Epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* 2005;54:710–17.

2 Vakil N, van Zanten SV, Kahrilas P, et al. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol* 2006;101:1900–20.

3 Tack J, Becher A, Mulligan C, Johnson DA. Systematic review: the burden of disruptive gastro-oesophageal reflux disease on health-related quality of life. *Aliment Pharmacol Ther* 2012;35:1257–66.

4 Armstrong D, Marshall JK, Chiba N, et al. Canadian Consensus Conference on the management of gastroesophageal reflux disease in adults: Update 2004. *Can J Gastroenterol* 2005;19:15–35.

5 Kahrilas PJ, Shaheen NJ, Vaezi MF. American Gastroenterological Association Institute technical review on the management of gastroesophageal reflux disease. *Gastroenterology* 2008;135:1392–413.

6 Dent J, Vakil N, Jones R, et al. Accuracy of the diagnosis of GORD by questionnaire, physicians and a trial of proton pump inhibitor treatment: the Diamond Study. *Gut* 2010;59:714–21.

7 Kinoshita Y, Adachi K, Hongo M, Haruma K. Systematic review of the epidemiology of gastroesophageal reflux disease in Japan. *J Gastroenterol* 2011;46:1092–103.

8 Vela MF, Craft BM, Sharma N, et al. Refractory heartburn: comparison of intercellular space diameter in documented GERD vs. functional heartburn. *Am J Gastroenterol* 2011;106:844–50.

9 Altomare A, Ma J, Guarino MPL, et al. Platelet-activating factor and distinct chemokines are elevated in mucosal biopsies of erosive compared with non-erosive reflux disease patients and controls. *Neurogastroenterol Motil* 2012;24:943–51.

10 Lee ES, Kim N, Lee SH, et al. Comparison of risk factors and clinical responses to proton pump inhibitors in patients with erosive esophagitis and non-erosive reflux disease. *Aliment Pharmacol Ther* 2009;30:154–64.

11 Wu JCY, Lai LH, Chow DKL, et al. Concomitant irritable bowel syndrome is associated with failure of

step-down on-demand proton pump inhibitor treatment in patients with gastroesophageal reflux disease. *Neurogastroenterol Motil* 2011;23:155–60, e31.
12 Zerbib F, Bruley des Varannes S, Simon M, Galmiche JP. Functional heartburn: definition and management strategies. *Curr Gastroenterol Rep* 2012;14:181–8.
13 Sharma P, Wani S, Romero Y, et al. Racial and geographic issues in gastroesophageal reflux disease. *Am J Gastroenterol* 2008;103:2669–80.
14 Ness-Jensen E, Lindam A, Lagergren J, Hveem K. Changes in prevalence, incidence and spontaneous loss of gastro-oesophageal reflux symptoms: a prospective population-based cohort study, the HUNT study. *Gut* 2012;61:1390–7.
15 Ruigomez A, Wallander MA, Johansson S, et al. Natural history of gastroesophageal reflux disease diagnosed in UK general practice. *Aliment Pharmacol Ther* 2004;20:751–60.
16 Becher A, Dent J. Systematic review: ageing and gastro-oesophageal reflux disease symptoms, oesophageal function and reflux oesophagitis. *Aliment Pharmacol Ther* 2011;33:442–54.
17 Goh KL. Gastroesophageal reflux disease in Asia: a historical perspective and present challenges. *J Gastroenterol Hepatol* 2011;26(Suppl 1):2–10.
18 Chen KY, Lou HY, Lin HC, Lee SH. Seasonal variation in the incidence of gastroesophageal reflux disease. *Am J Med Sci* 2009;338:453–8.
19 Fujiwara Y, Arakawa T. Epidemiology and clinical characteristics of GERD in the Japanese population. *J Gastroenterol* 2009;44:518–34.
20 Malfertheiner P, Nocon M, Vieth M, et al. Evolution of gastro-oesophageal reflux disease over 5 years under routine medical care: the ProGERD study. *Aliment Pharmacol Ther* 2012;35:154–64.
21 Zagari RM, Law RG, Fuccio L, et al. Dyspeptic symptoms and endoscopic findings in the community: the Loiano-Monghidori Study. *Am J Gastroenterol* 2010;105:565–71.
22 Gasiorowska A, Poh CH, Fass R. Gastroesophageal reflux disease (GERD) and irritable bowel syndrome (IBS) – is it one disease or an overlap of two disorders? *Dig Dis Sci* 2009;54:1829–34.
23 Xiao YL, Peng S, Tao J, et al. Prevalence and symptom pattern of pathologic esophageal acid reflux in patients with functional dyspepsia based on the Rome III criteria. *Am J Gastoenterol* 2010;105:2626–31.
24 Ohara S, Kawano T, Kusano M, Kouzo T. Survey on the prevalence of GERD and FD based on the Montreal definition and the Rome III criteria among patients presenting with epigastric symptoms in Japan. *J Gastroenterol* 2011;46:603–11.
25 Savarino E, Pohl D, Zentilin P, et al. Functional heartburn has more in common with functional dyspepsia than with non-erosive reflux disease. *Gut* 2009;58:1185–91.
26 Pandeya N, Green AC, Whiteman DC for the Australian Cancer Study. Prevalence and determinants of frequent gastroesophageal reflux symptoms in the Australian community. *Dis Esoph* 2012;25:573–83.
27 Horiki B, Omata F, Ninomiya K, et al. Caucasian ethnicity as a risk factor for more severe mucosal damage in gastroesophageal reflux disease. *Esophagus* 2012;9:153–7.
28 Dent J, Becher A, Sung J, et al. Systematic review: patterns of reflux-induced symptoms and esophageal endoscopic findings in large surveys. *Clin Gastroenterol Hepatol* 2012;10:863–73.
29 Ronkainen J, Aro P, Storskrubb T, et al. High prevalence of gastroesophageal reflux symptoms and esophagitis with or without symptoms in the general adult Swedish population: a Kalixanda study report. *Scand J Gastroenterol* 2005;40:275–85.
30 Zagari RM, Fuccio L, Wallander MA, et al. Gastro-oesophageal reflux symptoms, oesophagitis and Barrett's oesophagus in the general population: the Loiano-Monghidoro study. *Gut* 2008;57:1354–9.
31 Zou D, He J, Ma X, et al. Epidemiology of symptom-defined gastroesophageal reflux disease and reflux esophagitis: the systematic investigation of gastrointestinal diseases in China (SILC). *Scand J Gastroenterol* 2011;46:133–41.
32 Nasseri-Moghaaddam S, Mofid A, Razjouyan H. Dietary nitrate may have a role in development of gastro-esophageal reflux disease. *Arch Iran Med* 2011;14:312–14.
33 Mostaghni A, Mehrabani D, Khademolhosseini F, et al. Prevalence and risk factors of gastroesophageal reflux disease in Qashqai migrating nomads, southern Iran. *World J Gastroenterol* 2009;15:961–5.
34 He J, Ma X, Zhao Y, et al. A population-based survey of the epidemiology of symptom-defined gastroesophageal reflux disease: the systematic investigation of gastrointestinal diseases in China. *BMC Gastroenterol* 2010;10:94.
35 Perlman Se, Friedman S, Galea S, et al. Short-term and medium-term effects of 9/11. *Lancet* 2011;378:925–34.
36 Li J, Brackbill RM, Stellman SD, et al. Gastroesophageal reflux symptoms and comorbid asthma and posttraumatic stress disorder following the 9/11 terrorist attacks on World Trade Center in New York City. *Am J Gastroenterol* 2011;106:1933–41.
37 Wisnivesky JP, Teitelbaum SL, Todd AC, et al. Persistence of multiple illnesses in World Trade Center rescue and recovery workers: a cohort study. *Lancet* 2011;378:888–97.
38 Martin-Merino E, Ruigomez A, Garcia Rodriguez LA, et al. Depression and treatment with antidepressants are

associated with the development of gastro-oesophageal reflux disease. *Aliment Pharmacol Ther* 2010;31:1132–40.

39 Blaser MJ. Disappearing microbiota: *Helicobacter pylori* protection against esophageal adenocarcinoma. *Cancer Prev Res (Phila PA)* 2008;1:308–11.

40 Raghunath A, Hungin APS, Wooff D, Childs S. Prevalence of *Helicobacter pylori* in gastro-oesophageal reflux disease: systematic review. *BMJ* 2003;326:736–9.

41 Leontiadis GI, Kadis S, Sharma VK, Howden CW. *Helicobacter pylori* in gastro-oesophageal reflux disease needs comparator [letter]. *BMJ* 2003;326:1460.

42 Yaghoobi M, Farrokhyar F, Yuan Y, Hunt RH. Is there an increased risk of GERD after *Helicobacter pylori* eradication?: a meta-analysis. *Am J Gastroenterol* 2010;105:1007–13.

43 Ford AC, Forman D, Reynolds PD, et al. Ethnicity, gender, and socioeconomic status as risk factors for esophagitis and Barrett's esophagus. *Am J Epidemiol* 2005;162:454–60.

44 El-Serag HB, Graham DY, Satia JA, Rabeneck L. Obesity is an independent risk factor for GERD symptoms and erosive esophagitis. *Am J Gastroenterol* 2005;100:1243–50.

45 Hampel H, Abraham NS, El-Serag HB. Meta-analysis: obesity and the risk for gastroesophageal reflux disease and its complications. *Ann Intern Med* 2005;143:199–211.

46 Murray L, Johnston B, Lane A, et al. Relationship between body mass and gastroesophageal reflux symptoms: the Bristol Helicobacter Project. *Int J Epidemiol* 2003;32:645–50.

47 Ford AC, Forman D, Bailey AG, et al. The natural history of gastro-oesophageal reflux symptoms in the community and its effects on survival: a longitudinal 10-year follow-up study. *Aliment Pharmacol Ther* 2013;37:323–31.

48 Nam SY, Choi IJ, Ryu KH, et al. Abdominal visceral adipose tissue volume is associated with increased risk of erosive esophagitis in men and women. *Gastroenterology* 2010;139:1902–11.

49 Tai CM, Lee YC, Tu HP, et al. The relationship between visceral adiposity and the risk of erosive esophagitis in severely obese Chinese patients. *Obesity* 2010;18:2165–9.

50 Vemulapalli R. Diet and lifestyle modifications in the management of gastroesophageal reflux disease. *Nutr Clin Pract* 2008;23:293–8.

51 Roman S, Pandolfino JE. Environmental – lifestyle related factors. *Best Pract Res Clin Gastroenterol* 2010;24:847–59.

52 Dibley LB, Norton C, Jones R. Don't eat tomatoes: patient's self-reported experiences of causes of symptoms in gastro-oesophageal reflux disease. *Fam Practice* 2010;27:410–17.

53 Locke GR 3rd, Talley NJ, Fett SL, et al. Risk factors associated with symptoms of gastroesophageal reflux. *Am J Med* 1999;106:642–9.

54 El-Serag HB, Satia JA, Rabeneck L. Dietary intake and the risk of gastro-oesophageal reflux disease: a cross sectional study in volunteers. *Gut* 2005;54:11–17.

55 Mulholland HG, Cantwell MM, Anderson LA, et al. Glycemic index, carbohydrate and fiber intakes and risk of reflux esophagitis, Barrett's esophagus and esophageal adenocarcinoma. *Cancer Causes Control* 2009;20:279–88.

56 Nachman F, Vázquez H, González A, et al. Gastroesophageal reflux symptom in patients with celiac disease and the effects of a gluten-free diet. *Clin Gastroenterol Hepatol* 2011;9:214–19.

57 Gonsalves N, Yang GY, Doerfler B, et al. Elimination diet effectively treats eosinophilic esophagitis in adults: food reintroduction identifies causative factors. *Gastroenterology* 2012;142:1451–9.

58 Pomiecinski F, Yang AC, Navarro-Rodrigues T, et al. Sensitization to foods in gastroesophageal reflux disease and its relation to eosinophils in the esophagus: is it of clinical importance? *Ann Allergy Asthma Immunol* 2010;105:359–63.

59 Trudgill NJ, Kapur KC, Riley SA. Familial clustering of reflux symptoms. *Am J Gastroenterol* 1999;94:1172–8.

60 Tsibouris P, Moussia M, Kalantzis C, et al. Endoscopic esophagitis is more severe in gastroesophageal reflux patients with a positive family history. *J Clin Gastroenterol* 2012;46:201–8.

61 Cameron AJ, Lagergren J, Henriksson C, et al. Gastroesophageal reflux disease in monozygotic and dizygotic twins. *Gastroenterology* 2002;122:55–9.

62 Mohammed I, Cherkas LF, Riley SA, et al. Genetic influences in gastro-oesophageal reflux disease: a twin study. *Gut* 2003;52:1085–9.

63 Chourasia D, Achyut BR, Tripathi S, et al. Genotypic and functional roles of IL-1B and IL-1RN on the risk of gastroesophageal reflux disease: the presence of IL-1B-511∗T/IL-1RN∗1 (T1) haplotype may protect against the disease. *Am J Gastroenterol* 2009;104:2704–13.

64 Jirholt J, Asling B, Hammond P, et al. 4-aminobutyrate aminotransferase (ABAT): genetic and pharmacological evidence for an involvement in gastro-esophageal reflux disease. *PLoS ONE* 2011;6:e19095.

65 Spechler SJ, Sharma P, Souza RF, et al. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* 2011;140:1084–91.

66 Playford RJ. New British Society of Gastroenterology (BSG) guidelines for the diagnosis and management of Barrett's oesophagus. *Gut* 2006;55:442–3.
67 Hayeck TJ, Kong CY, Spechler SJ, et al. The prevalence of Barrett's esophagus in the US: estimates from a simulation model confirmed by SEER data. *Dis Esophagus* 2010;23:451–7.
68 Veldhuyzen van Zanten SJO, Thomson ABR, Barkun AN, et al. The prevalence of Barrett's oesophagus in a cohort of 1040 Canadian primary care patients with uninvestigated dyspepsia undergoing prompt endoscopy. *Aliment Pharmacol Ther* 2006;23:595–9.
69 Ronkainen J, Aro P, Storskrubb T, et al. Prevalence of Barrett's esophagus in the general population: an endoscopic study. *Gastroenterology* 2005;129:1825–31.
70 Voutilainen M, Sipponen P, Mecklin JP, et al. Gastroesophageal reflux disease: prevalence, clinical, endoscopic and histopathological findings in 1,128 consecutive patients referred for endoscopy due to dyspeptic and reflux symptoms. *Digestion* 2000;61:6–13.
71 Flameling RD, Numans ME, ter Linde J, et al. Different characteristics of patients with gastro-oesophageal reflux disease on their path through healthcare: a population follow-up study. *Eur J Gastroenterol Hepatol* 2010;22:578–82.
72 Malfertheiner P, Lind T, Willich S, et al. Prognostic influence of Barrett's oesophagus and *Helicobacter pylori* infection on healing of erosive gastro-oesophageal (GORD) and symptom resolution in non-erosive GORD: report from the ProGORD Study. *Gut* 2005;54:746–51.
73 Xiong LS, Wang JP, Wang JH, et al. Prevalence and risk factors of Barrett's esophagus in patients undergoing endoscopy for upper gastrointestinal symptoms. *J Dig Dis* 2010;11:83–7.
74 Lee IS, Choi SCC, Shim KN, et al. Prevalence of Barrett's esophagus remains low in the Korean population: nationwide cross-sectional prospective multicenter study. *Dig Dis Sci* 2010;55:1932–9.
75 Kuo CJ, Lin CH, Liu NJ, et al. Frequency and risk factors for Barrett's esophagus in Taiwanese patients: a prospective study in a tertiary referral centre. *Dig Dis Sci* 2010;55:1337–43.
76 Fouad YM, Makhlouf MM, Tawfik HM, et al. Barrett's esophagus: Prevalence and risk factors in patients with chronic GERD in Upper Egypt. *World J Gastroenterol* 2009;15:3511–5.
77 Rajendra S. Barrett's oesophagus in Asians – are ethnic differences due to genes or the environment? *J Intern Med* 2011;270:421–7.
78 Menon S, Jayasena H, Nightingale P, Trudgill NJ. Influence of age and sex on endoscopic findings of gastroesophageal reflux disease: an endoscopy database study. *Eur J Gastroenterol Hepatol* 2011;23:389–95.
79 Ronkainen J, Talley NJ, Storskrubb T, et al. Erosive esophagitis is a risk factor for Barrett's esophagus: a community-based endoscopic follow-up study. *Am J Gastroenterol* 2011;106:1946–52.
80 Taylor JB, Rubenstein JH. Meta-analyses of the effect of symptoms of gastroesophageal reflux on the risk of Barrett's esophagus. *Am J Gastroenterol* 2010;105:1730–7.
81 Edelstein ZR, Bronner MP, Rosen SN, Vaughan TL. Risk factors for Barrett's esophagus among patients with gastroesophageal reflux disease: a community clinic-based case-control study. *Am J Gastroenterol* 2009;104:834–42.
82 Wang A, Mattek NC, Holub JL, et al. Prevalence of complicated gastroesophageal reflux disease and Barrett's esophagus among racial groups in a multi-center consortium. *Dig Dis Sci* 2009;54:964–71.
83 Kubo A, Levin TR, Block G, et al. Cigarette smoking and the risk of Barrett's esophagus. *Cancer Causes Control* 2009;20:303–11.
84 Kubo A, Levin TR, Block G, et al. Alcohol types and sociodemographic characteristics as risk factors for Barrett's esophagus. *Gastroenterology* 2009;136:806–15.
85 Winberg H, Lindblad M, Lagergren J, Dahlstrand H. Risk factors and chemoprevention in Barrett's esophagus: an update. *Scand J Gastroenterol* 2012;47:397–406.
86 Kubo A, Block G, Quesenberry CP Jr, et al. Effects of dietary fiber, fats and meat intakes on the risk of Barrett's esophagus. *Nutrition Cancer* 2009;61:607–16.
87 Thompson OM, Beresford SA, Kirk EA, Vaughan TL. Vegetable and fruit intakes and risk of Barrett's esophagus in men and women. *Am J Clin Nutr* 2009;89:890–6.
88 Murphy SJ, Anderson LA, Ferguson HR, et al. Dietary antioxidant and mineral intake in humans is associated with reduced risk of esophageal adenocarcinoma but not reflux esophagitis or Barrett's esophagus. *J Nutr* 2010;140:1757–63.
89 Ibiebele TI, Hughes MC, Pandeya N, et al. High intake of folate from food sources is associated with reduced risk of esophageal cancer in an Australian population. *J Nutr* 2011;141:274–83.
90 Ryan AM, Healy LA, Power DG, et al. Barrett esophagus: prevalence of central adiposity, metabolic syndrome, and a proinflammatory state. *Ann Surg* 2008;247:909–15.
91 Anderson LA, Johnston BT, Watson RG, et al. Non-steroidal anti-inflammatory drugs and the esophageal inflammation-metaplasia-adenocarcinoma sequence. *Cancer Res* 2006;66:4975–82.

92 Thrift AP, Pandeya N, Smith KJ, et al. The use of nonsteroidal anti-inflammatory drugs and the risk of Barrett's oesophagus. *Aliment Pharmacol Ther* 2011; 34:1235–44.
93 Chak A, Lee T, Kinnard MF, et al. Familial aggregation of Barrett's oesophagus, oesophageal adenocarcinoma, and oesophagogastric junctional adenocarcinoma in Caucasian adults. *Gut* 2002;51:323–8.
94 Romero Y, Cameron AJ, Schaid DJ, et al. Barrett's esophagus: prevalence in symptomatic relatives. *Am J Gastroenterol* 2002;97:1127–32.
95 Juhasz A, Mittal SK, Lee TH, et al. Prevalence of Barrett esophagus in first-degree relatives of patients with esophageal adenocarcinoma. *J Clin Gastroenterol* 2011;45:867–71.
96 Huang JQ, Zheng GF, Sumanac K, et al. Meta-analysis of the relationship between cagA seropositivity and gastric cancer. *Gastroenterology* 2003;125: 1636–44.
97 Corley DA, Kubo A, Levin TR, et al. *Helicobacter pylori* infection and the risk of Barrett's oesophagus: a community-based study. *Gut* 2008;57:727–33.
98 GLOBOCAN 2008 (IARC). Oesophageal cancer incidence and mortality worldwide in 2008: summary. http://globocan.iarc.fr/factsheets/cancers/oesophagus .asp (accessed January 19, 2013).
99 Cook MB, Chow WH, Devesa SS. Oesophageal cancer incidence in the United States by race, sex, and histologic type, 1977–2005. *Br J Cancer* 2009;101:855–9.
100 Vizcaino AP, Moreno V, Lambert R, Parkin DM. Time trends incidence of both major histologic types of esophageal carcinomas in selected countries, 1973–1995. *Int J Cancer* 2002;99:860–8.
101 Blot WJ, Devesa SS, Kneller RW, et al. Rising incidence of adenocarcinoma of the esophagus and gastric cardia. *JAMA* 1991;265:1287–9.
102 van Blankenstein M, Looman CW, Siersema PD, et al. Trends in the incidence of adenocarcinoma of the oesophagus and cardia in the Netherlands 1989–2003. *Br J Cancer* 2007;96:1767–71.
103 Holmes RS, Vaughan TL. Epidemiology and pathogenesis of esophageal cancer. *Semin Radiat Oncol* 2006;17:2–9.
104 Fernandes ML, Seow A, Chan YH, Ho KY. Opposing trends in incidence of oesophageal squamous cell carcinoma and adenocarcinoma in a multi-ethnic Asian country. *Am J Gastroenterol* 2006;101:1430–6.
105 Lagergren J, Mattsson F. No further increase in the incidence of esophageal adenocarcinoma in Sweden. *Int J Cancer* 2011;129:513–16.
106 Pohl H, Sirovich B, Welch HG. Esophageal adenocarcinoma incidence: are we reaching the peak? *Cancer Epidemiol Biomarkers Prev* 2010;19:1468–70.
107 Thrift AP, Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. *Ann Oncol* 2012;23:3155–62.
108 Lepage C, Rachet B, Jooste V, et al. Continuing rapid increase in esophageal adenocarcinoma in England and Wales. *Am J Gastroenterol* 2008;103:2694–9.
109 Coleman HG, Bhat S, Johnston BT, et al. Tobacco smoking increases the risk of high-grade dysplasia and cancer among patients with Barrett's esophagus. *Gastroenterology* 2012;142:233–40.
110 Eslick GD. Epidemiology of esophageal cancer. *Gastroenterol Clin N Am* 2009;38:17–25.
111 Rutegard M, Lagergren P, Nordenstedt H, Lagergen J. Oesophageal adenocarcinoma: the new epidemic in men? *Maturitas* 2011;69:244–8.
112 Freedman ND, Abnet CC, Leitzmann MF, et al. A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. *Am J Epidemiol* 2007;165:1424–33.
113 Cook MB, Kamangar F, Whiteman DC, et al. Cigarette smoking and adenocarcinomas of the esophagus and esophagogastric junction: a pooled analysis from the international BEACON consortium. *J Natl Cancer Inst* 2010;102:1344–53.
114 Freedman ND, Murray LJ, Kamangar F, et al. Alcohol intake and risk of oesophageal adenocarcinoma: a pooled analysis from the BEACON Consortium. *Gut* 2011;60:1029–37.
115 Pandeya N, Webb PM, Sadeghi S, et al. for the Australian Cancer Study. Gastro-oesophageal reflux symptoms and the risks of oesophageal cancer: are the effects modified by smoking, NSAIDs or acid suppressants? *Gut* 2010;59:31–8.
116 Tran B, Lucas R, Kimlin M, et al. for the Australian Cancer Study. Association between ambient ultraviolet radiation and risk of esophageal cancer. *Am J Gastroenterol* 2012;107:1803–13.
117 Ryan AM, Duong M, Healy L, et al. Obesity, metabolic syndrome and esophageal adenocarcinoma: epidemiology, etiology and new targets. *Cancer Epidemiol* 2011;35:309–19.
118 Abnet CC, Freedman ND, Hollenbeck AR, et al. A prospective study of BMI and risk of oesophageal and gastric adenocarcinoma. *Eur J Cancer* 2008;44:465–71.
119 Ryan AM, Rowley SP, Fitzgerald AP, et al. Adenocarcinoma of the oesophagus and gastric cardia: male preponderance in association with obesity. *Eur J Cancer* 2006;42:1151–8.
120 Abrams JA, Sharaiha RZ, Gonsalves L, et al. Dating the rise of esophageal adenocarcinoma: analysis of Connecticut Tumor Registry Data, 1940–2007. *Cancer Epidemiol Biomarkers Prev* 2011;20:183–6.

121 Kong CY, Nattinger KJ, Hayeck TJ, et al. The impact of obesity on the rise in esophageal adenocarcinoma incidence: estimates from a disease simulation model. *Cancer Epidemiol Biomarkers Prev* 2011;20:2450–6.
122 O'Doherty MG, Freedman ND, Hollenbeck AR, et al. A prospective cohort study of obesity and risk of oesophageal and gastric adenocarcinoma in the NIH-AARP Diet and Health Study. *Gut* 2012;61:1261–8.
123 O'Rorke MA, Cantwell MM, Abnet CC, et al. Toenail trace element status and risk of Barrett's oesophagus and oesophageal adenocarcinoma: results from the FINBAR study. *Int J Cancer* 2012;131:1882–91.
124 Rubenstein JH, Taylor JB. Meta-analysis: the association of oesophageal adenocarcinoma with symptoms of gastro-oesophageal reflux. *Aliment Pharmacol Ther* 2010; 32:1222–7.
125 Rokkas T, Pistiolas D, Sechopoulos P, et al. Relationship between *Helicobacter pylori* infection and esophageal neoplasia: a meta-analysis. *Clin Gastroenterol Hepatol* 2007;5:1413–17.
126 Whiteman DC, Parmar P, Fahey P, et al. Association of *Helicobacter pylori* infection with reduced risk for esophageal cancer is independent of environmental and genetic modifiers. *Gastroenterology* 2010;139:73–83.
127 El-Omar EM, Carrington M, Chow WH, et al. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* 2000;404:398–402.
128 Gorouhi F, Islami F, Bahrami H, et al. Tumour-necrosis factor-a polymorphisms and gastric cancer risk: a meta-analysis. *Br J Cancer* 2008;98:1443–51.
129 Sadeghi S, Bain CJ, Pandeya N, et al. Aspirin, nonsteroidal anti-inflammatory drugs, and the risks of cancers of the esophagus. *Cancer Epidemiol Biomarkers Prev* 2008;17:1169–78.
130 Vaughan TL, Dong LM, Blount PL, et al. Non-steroidal anti-inflammatory drugs and risk of neoplastic progression in Barrett's oesophagus: a prospective study. *Lancet Oncol* 2005;6:945–52.
131 Kastelein F, Spaander MC, Biermann K, et al. Nonsteroidal anti-inflammatory drugs and statins have chemopreventative effects in patients with Barrett's esophagus. *Gastroenterology* 2011;141:2000–8.
132 Nguyen DM, Richardson P, El-Serag HB. Medications (NSAID, statins, PPI) and the risk of esophageal adenocarcinoma in patients with Barrett's esophagus. *Gastroenterology* 2010;138:2260–6.
133 Abnet CC, Freedman ND, Kamangar F, et al. Nonsteroidal anti-inflammatory drugs and risk of gastric and oesophageal adenocarcinomas: results from a cohort study and a meta-analysis. *Br J Cancer* 2009;100:551–7.
134 Heath EI, Canto MI, Piantadosi S, et al. Secondary chemoprevention of Barrett's esophagus with celecoxib: results of a randomized trial. *J Natl Cancer Inst* 2007;99:545–57.
135 Gatenby PA, Ramus JR, Caygill CP, et al. Aspirin is not chemoprotective for Barrett's adenocarcinoma of the oesophagus in multicentre cohort. *Eur J Cancer Prev* 2009;18:381–4.
136 Kantor ED, Onstad L, Blount PL, et al. Use of statin medications and risk of esophageal adenocarcinoma in persons with Barrett's esophagus. *Cancer Epidemiol Biomarkers Prev* 2012;21:456–61.
137 Jankowski J, Moayyedi P. Cost effectiveness of aspirin chemoprevention for Barrett's esophagus. *J Natl Cancer Inst* 2004;96:885–7.
138 Wu IC, Zhao Y, Zhai R, et al. Association between polymorphisms in cancer-related genes and early onset of esophageal adenocarcinoma. *Neoplasia* 2011;13: 386–92.
139 Menke V, van Zoest KPM, Moons LMG, et al. NcoI TNF-b gene polymorphism and TNF expression are associated with an increased risk of developing Barrett's esophagus and esophageal adenocarcinoma. *Scand J Gastroenterol* 2012;47:378–86.
140 Wang GQ, Abnet CC, Shen Q, et al. Histological precursors of oesophageal squamous cell carcinoma: results from a 13-year prospective follow-up study in a high risk population. *Gut* 2005;54:187–92.
141 Shaheen NJ, Crosby MA, Bozymski EM, et al. Is there publication bias in the reporting of cancer risk in Barrett's esophagus? *Gastroenterology* 2000;119:333–8.
142 Wani S, Falk G, Hall M, et al. Patients with nondysplastic Barrett's esophagus have low risks for developing dysplasia or esophageal adenocarcinoma. *Clin Gastroenterol Hepatol* 2011;9:220–7.
143 Desai TK, Krishnan K, Samala N, et al. The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut* 2012;61:970–6.
144 Rastogi A, Puli S, El-Serag HB, et al. Incidence of esophageal adenocarcinoma in patients with Barrett's esophagus and high-grade dysplasia: a meta-analysis. *Gastrointest Endosc* 2008;67:394–8.
145 den Hoed CM, van Blankenstein M, Dees J, Kuipers EJ. The minimal incubation period from the onset of Barrett's oesophagus to symptomatic adenocarcinoma. *Br J Cancer* 2011;105:200–5.

Answers to multiple choice questions

1. C
2. B
3. D
4. D

14 Epidemiology of *Helicobacter pylori* infection, peptic ulcer disease and gastric cancer

Grigorios I. Leontiadis[1] *& Olof Nyrén*[2]

[1]Department of Medicine, Division of Gastroenterology, McMaster University, Hamilton, ON, Canada

[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Key points

- Approximately half of all humans harbor *Helicobacter pylori.*
- Persistent *H. pylori* infection is a cause of peptic ulcer, gastric adenocarcinoma, and MALT lymphoma.
- *H. pylori* prevalence in each birth cohort reflects the risk of acquisition that prevailed during the cohort members' childhoods.
- Peptic ulcer may affect, at some point in life, 4–12 % of the adult population, and the population attributable risk (PAR) for *H. pylori* has been estimated to be 48 %.
- Due to its poor prognosis, stomach cancer (adenocarcinoma) ranks number two among all causes of cancer death (10 % of all cancer deaths), with more than two-thirds of the cases occurring in developing countries.
- There has been a steep downward trend for distal stomach cancer in white men and women, but this decline does not seem to include cardia cancer.

Helicobacter pylori infection

Clinical microbiology and expression

Helicobacter pylori infection is an established cause of both peptic ulcer disease and gastric cancer. The *Helicobacter* genus consists of over 20 recognized species, including *H. pylori*. The latter is a curved bacterium, 2.5–4.0 μm long, that produces urease, which is thought to make short-term survival possible in the highly acidic intragastric environment. In contrast to many other bacterial pathogens, *H. pylori* is genetically heterogeneous, a result of several mechanisms for DNA rearrangement, including introduction and deletion of foreign sequences. The genetic heterogeneity is thought to reflect the microorganism's extraordinary ability for adaptation, both to the inhospitable acidic environment and to various attacks from the host's immune system [1]. Some regions of the 1.7-Mbp bacterial genome are more variable than others. A striking example is the *cag* pathogenicity island (*cag* PAI), a 37–40-kb genetic element that contains the *cagA* gene [2]. It is present in approximately 50–70 % of *H. pylori* strains and was linked early on to a higher inflammatory response and to a particularly elevated risk of manifest diseases such as peptic ulcer or cancer in the host [3]. The entire island may be restored or lost through transformation [4]. The genes on the *cag* PAI encode 27–31 proteins, among them CagA and the components of a type IV secretion system (TFSS). The latter injects CagA into the host's epithelial cells [5], where it is phosphorylated and interacts with a range of host signaling molecules. This, in turn, leads to morphologic changes and proliferation of the epithelial cells. The intimate interaction of the microorganism with the host cells results

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

in the induction of potent proinflammatory cytokines such as interleukin 8 (IL-8) through the activation of the intracellular innate immune receptor Nod1 and nuclear factor kappa B (NF-κB) [6]. The *cag* PAI may be incomplete, and thus not fully functional. Comparisons between cancer or precancer-related *H. pylori* strains and noncancer strains have indicated that the cancer-related ones tend to have more complete *cag* PAIs [7,8].

The primary histologic lesion caused by *H. pylori* is gastritis. As opposed to the acute gastritis that follows initial colonization and that tends to be associated with transient nonspecific symptoms, the ensuing chronic gastritis is essentially symptomless in most individuals.

Persistent *H. pylori* infection may lead to peptic ulcer, gastric adenocarcinoma, and mucosa-associated lymphoid tissue (MALT) lymphoma. Although several known and yet unidentified cofactors may be required for these respective outcomes, the causal relationships with the infection are widely accepted [9–11]. In the case of peptic ulcer and MALT lymphoma, the causality is supported by eradication studies demonstrating disease control after *H. pylori* eradication [12,13]. It is estimated that *H. pylori*-positive patients have a 10–20 % lifetime risk of developing peptic ulcer disease and a 1–2 % risk of developing gastric cancer [14]. Hence, the overwhelming majority of infected individuals will never develop any clinically manifest *H. pylori*-related disease. Numerous studies have explored possible associations between the infection and a variety of extragastric conditions, but with the possible exception of iron deficiency anemia, no conclusive evidence has emerged [10,15,16].

Distribution of *H. pylori* infection in the general population

Phylogeographic studies indicate that humans have been colonized by *H. pylori* for more than 58,000 years, before their migration from east Africa [17]. Nowadays, approximately half of all humans harbor *H. pylori* [18], but the prevalence shows large geographic variations. Whilst generally less than 40 % of people in industrialized countries are *H. pylori* positive, the prevalence of the infection in various developing countries is more than 80 % [19]. The range is even greater among child populations, with prevalence rates varying from below 10 % to over 80 % in high-income and low-income countries, respectively [20]. This means that children in many impoverished countries rapidly – typically before adolescence – reach the prevalence prevailing in the adult population. In several such populations, a prevalence of 50 % is reached by the age of 5 years [21–23]. However, pediatric prevalence studies need to be interpreted with caution. Compared to adults, young children demonstrate significant variability in their immune response to this infection. As a result, serologic assays for *H. pylori* antibodies are less accurate, with particularly low sensitivity in young children [24]. Furthermore, longitudinal studies have unveiled complex dynamics; in a US-Mexican cohort of infants, who were followed with ^{13}C-urea breath tests during the first two years of life, the initial acquisition of detectable *H. pylori* infection occurred at a rate of 20 % per year, but most of these infections did not persist [25].

Whereas in developing countries the prevalence ceiling is reached before or during adolescence, *H. pylori* prevalence continues to rise with age in the adult population of industrialized countries. At the same time, there are strong indications that the overall prevalence in the latter countries is rapidly declining over calendar time [26]. Studies on stored sera suggest that this fall in prevalence is mainly explained by a birth cohort-wise decline in early acquisition of the infection [27,28]. Accordingly, the *H. pylori* prevalence in each birth cohort (generation) reflects the risk of acquisition that prevailed during the cohort members' childhood. Because this risk seems to have fallen dramatically in developed countries during the twentieth century, the subsequent prevalence in any given calendar year is expected to be inversely related to year of birth and, consequently, positively related to age. Nevertheless, the *H. pylori* prevalence among children in developed countries may not continue to fall at the same rate as previously. A recent study from the Netherlands suggested that the *H. pylori* prevalence in birth cohorts of children remained stable (around 9 %) from 1993 to 2005, despite a previously documented decreasing trend from 1978 to 1993 [29].

The seroconversion rate, marking the incidence of new *H. pylori* infections in these adult populations, has been estimated to be 1–2 per 200 persons and year [18,30], thus contributing little to the age effect. There are also seroreversions, that is, serologic indications of *H. pylori* disappearance. This rate was approximately 3 per 200 persons and year in both Sweden

and Japan [30,31]. Hence, spontaneous disappearance of *H. pylori* will tend to balance the addition of new infections in adult populations. Interestingly, the prevalence of CagA(+) *H. pylori* strains among young adults and children in Western countries seems to be decreasing much faster than that of CagA(−) strains [29,32].

Transmission of *H. pylori*

The mode of transmission of the infection has remained elusive, as have the mechanisms involved. Decades of intense research have failed to identify any important reservoir for the microorganism other than the human stomach. This implies that direct human-to-human transmission is the principal – perhaps the only – way by which the *H. pylori* species secures its continued existence. However, although challenged in some studies [33,34], the infectivity in adulthood seems to be limited [18,35]. Most infected individuals, no doubt, have contracted their infection during childhood [36], but a Swedish study revealed strain concordance upon molecular typing in approximately one-fifth of married couples [37]. *H. pylori* has been detected in saliva, dental plaque, vomitus, gastric refluxate, and feces, but there is no conclusive evidence for predominant transmission via any of these vehicles [14]. Thus, it appears that the transmission can occur via both the oral-oral and fecal-oral route.

The family stands out as the most important framework for transmission, at least in developed countries [36]. Family size (both while growing up and as an adult), presence of infected family members, familial connections to high-prevalence regions, and residential crowding are all factors that are associated with an increased risk of being infected [22,38–41]. Clustering of *H. pylori* infection in sibships is consistent with transmission between siblings [22,37,40]. Presence of infected siblings was an independent strong risk factor for infection among 11- to 13-year-old children in Sweden, even after control for parental infection status [38]. Furthermore, in a molecular typing study from Sweden, siblings were frequently infected with the same strains [37]. However, in these families, it was common that the mother also carried the same strain. Thus, it is still possible that the mother might have been the common source. An *H. pylori*-infected mother is a much stronger risk factor for the child than an infected father [39,38], suggesting that close contacts are more important than possible genetic predisposition. Interestingly, close contacts with infected children outside the family, such as with peers at day-care centers or at school, were not associated with an increased risk of infection in studied index children in Sweden [39], while day-care attendance was a risk factor in urban Sardinia [42] (Figure 14.1).

Although exposure opportunity in the form of close contacts with an infected family member may be more important than genetic factors, this does not mean that the host's genetic predisposition is unimportant. The concordance within adult twin pairs with regard to *H. pylori* seropositivity was considerably greater in monozygotic (81 %) than in dizygotic (63 %) twins

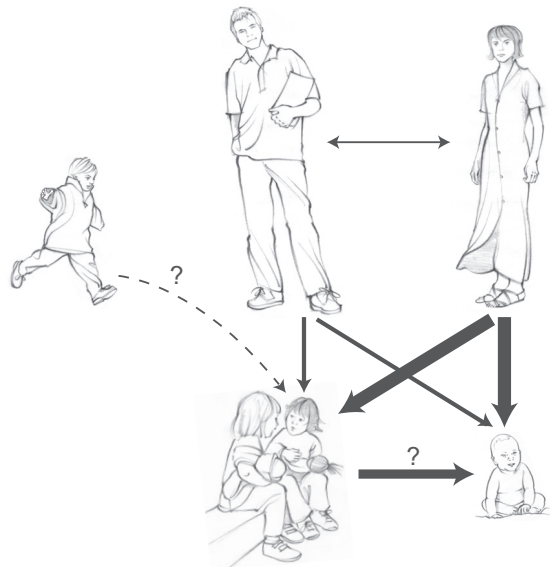


Figure 14.1 Data from affluent Western populations point to the family as the most important framework for *H. pylori* transmission. An *H. pylori*-infected mother is a much stronger risk factor for the child than an infected father. As uninfected adults rarely contract the infection, transmission between spouses is rare. Clustering of *H. pylori* infection in sibships is often observed and siblings are frequently infected with the same strains, but the mother may be the common source. However, presence of infected siblings seems to be an independent strong risk factor for infection even after control for parental infection status. Close contacts with infected children outside the family, such as with peers at day-care centers or at school, do not seem to be associated with an increased risk of infection in well-developed countries.

[43], suggesting that genetic mechanisms in the host may be involved. The exact nature of these mechanisms remains to be clarified. Although not universally confirmed, studies in Japan and Sweden have demonstrated that presence of the *0102 allele of the human leukocyte antigen (HLA) class II *DR-DQA1* locus is inversely and significantly associated with *H. pylori* seropositivity [44,45]. An Italian study of polymorphisms in the interleukin (IL) gene cluster (*IL1B*, *IL1RN*), interleukin-10 gene (*IL10*), tumor necrosis factor alpha gene (*TNF-A*), and interferon gamma gene (*IFNG*) found the *TNF-A* –308AG genotype to be associated with an increased overall prevalence of *H. pylori* infection, and the *IFNG* + 874AA genotype to be linked specifically to *cagA*-positive infections, while the other studied polymorphisms were unrelated to *H. pylori* status [46]. As these studies were all of a cross-sectional nature and conducted among adults, the genetic predisposition could equally well pertain to persistence of the infection as to initial acquisition. Because blood group antigens mediate bacterial adhesion to the gastric mucosa, and *H. pylori* strains may have adapted their binding affinity in accordance with the blood group antigen expression of different human populations [47], the blood group phenotype of the host is potentially of interest. However, the results of a handful of studies on Lewis genotypes and phenotypes, as well as ABO phenotypes, are inconsistent.

Risk factors for *H. pylori* infection in the adult population

The literature on risk factors for *H. pylori* seropositivity in adult life is large but generally cross-sectional and therefore unable to distinguish between effects on *H. pylori* acquisition and persistence. The possibility of reverse causation must also be borne in mind, for instance when anthropometric measures and dietary habits are considered as risk factors. Among the studies that can be characterized as population-based, there is overwhelming consensus about the importance of *age* (or, indirectly, birth year) and *socioeconomic status* (including various indices of domestic crowding and/or underprivileged home during childhood). In the United States (USA), young African Americans have more than a threefold increased *H. pylori* prevalence compared with their Caucasian peers [48]. It also appears that *men*, when compared with women, generally have a slightly higher risk of being infected, at least in Western populations [49,50]. Although not confirmed by all investigators, *family size* during childhood may be important, but data on the significance of birth order are conflicting. *Smoking* has generally been found to be unrelated to *H. pylori* status, but a few exceptions exist. Some investigators have found that a moderate *alcohol intake* may be associated with a decreased *H. pylori* seroprevalence [51–53], while others found no association, or even an increased seroprevalence [54]. Low intake of *fruit and vegetables* tended to be a risk factor in a number of studies [55–57], but the strength of the relationships varied widely, from small nonsignificant associations to up to 19-fold risk gradients. As ascorbic acid inhibits the growth of *H. pylori* in animal models [58], the link between serum ascorbic acid levels and *H. pylori* seroprevalence has attracted much attention, but most clinical studies were unable to establish a clear relationship.

Peptic ulcer

Clinical outline

Peptic ulcers in the stomach or duodenum are defined as benign mucosal lesions that penetrate deeply into the gut wall, beyond the muscularis mucosae, and form craters surrounded by acute and chronic inflammatory cell infiltrates. Criteria for size of the lesion vary, but ≥5 mm is a common cutpoint. *Duodenal ulcers* are located in the upper portions of the duodenum and are generally associated with antrum-predominant gastritis, which contributes to a high and somewhat dysregulated acid output from the stomach. *Gastric ulcers* are located in the stomach, frequently along the lesser curvature and, in particular, in the transition zone from corpus to antrum mucosa. As opposed to duodenal ulcer disease, gastric ulcer tends to be preceded by pangastritis (affecting the entire stomach), often atrophic in character, resulting in low acid production.

Peptic ulcers tend to have a chronic remitting course; the ulcers come and go, often with imperfect correlation between symptoms and presence of an open crater. Among 224 community-based Australian patients with duodenal ulcer followed for up to 7 years, dyspepsia was present during 20 % of the time if untreated, and during 15 % if they were on

antiulcer treatment [59]. Asymptomatic ulcer occurrences are quite common, and complications may arise without any forewarning.

H. pylori eradication is the preferred treatment when definite cure and elimination of ulcer recurrence is the goal. Such treatment is associated with three- to fivefold higher success rates compared with placebo for both duodenal and gastric ulcer recurrence, and it is superior to pharmacologic acid suppression in duodenal ulcer healing [12].

Bleeding and perforation are the main complications. Gastric outlet obstruction is an increasingly rare complication, mainly restricted to duodenal ulcers. While the overwhelming majority of ulcer patients do not die of their disease, it has been estimated that the cure of active peptic ulcer increases life expectancy by 2.3 years in persons aged 40–44 years and 121 days in persons aged 70–74 years [60]. Among cases with newly diagnosed uncomplicated peptic ulcer in Funen County, Denmark, during 1993–2002, the standardized mortality ratio (SMR), which can be seen as the cases' relative risk of dying in comparison with the matching general population, was 2.5 (95 % confidence interval (CI) 2.3–2.7) during year 2–10 after initial diagnosis [61]. The corresponding SMR among new cases with complicated ulcer (bleeding or perforated) was 2.6, suggesting that if a patient only survives the acute phase of the complication, the survival is similar to that among patients with uncomplicated disease.

Occurrence of peptic ulcer in the general population

There are a number of problems involved in the assessment of incidence and prevalence of peptic ulcer. In particular, many ulcers are asymptomatic. What is observed in health care may only be the tip of an iceberg. Moreover, dramatic changes in the management of peptic ulcer in the past decades have imposed calendar period-dependent selection forces that complicate comparisons of hospitalizations or outpatient visits over time. Mortality from peptic ulcer is low and confounded by age distribution among affected individuals, comorbidity, and changes in management practices. Because only a minority of individuals with dyspepsia suggestive of peptic ulcer do in fact have the disease, and invasive tests in the form of radiology or gastroscopy are needed for a reliable diagnosis, self-reports form a shaky basis for calculations of incidence and prevalence. The superior way of investigating these matters is by means of population-based endoscopic surveys. Such surveys, on the other hand, may be severely biased unless a high participation rate is attained. The only such study that reasonably fulfills high-level quality requirements was conducted in northern Sweden [62]. The prevalence of peptic ulcer was 4.1 %, with an equal contribution of gastric and duodenal ulcers. Interestingly, epigastric pain/discomfort was not a significant predictor of peptic ulcer disease. It should be noted that the accumulated nonparticipation rate corresponded to 46 %. The final participants were, on average, older and were more likely to have symptoms, compared with the initial sample. Therefore, the prevalence may have been somewhat overestimated, but the proportion of all ulcers that were asymptomatic was presumably underestimated.

Secular trends in peptic ulcer occurrence

To summarize the secular trends as reflected by statistics of complications and mortality, the rates of peptic ulcer increased among successive birth cohorts in the nineteenth century to reach a peak among people born in around 1870–1920 (somewhat varying between populations), with an earlier peak among men than among women, and with the peak for gastric ulcer preceding that for duodenal ulcer (Figure 14.2) [63]. The subsequent calendar time-wise occurrence of ulcer deaths and complications is largely consistent with the birth cohort pattern, with falling rates among younger age groups, irrespective of gender and ulcer type, and a general – albeit not universal – tendency for downward trends also among elderly men, while the rates among women do not yet seem to diminish (Figure 14.3). This has also shifted the much-cited 2:1 male-to-female ratio towards unity. However, increasing overall death rates, particularly attributed to complicated ulcer among women, in several populations in which the most risky birth cohorts are expected to be disappearing, suggest that another trend is superimposed on the pure birth-cohort pattern. This trend could tentatively reflect external exposures that were introduced or increased during the last decades of the twentieth century; implicated factors include aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs), estimated to account

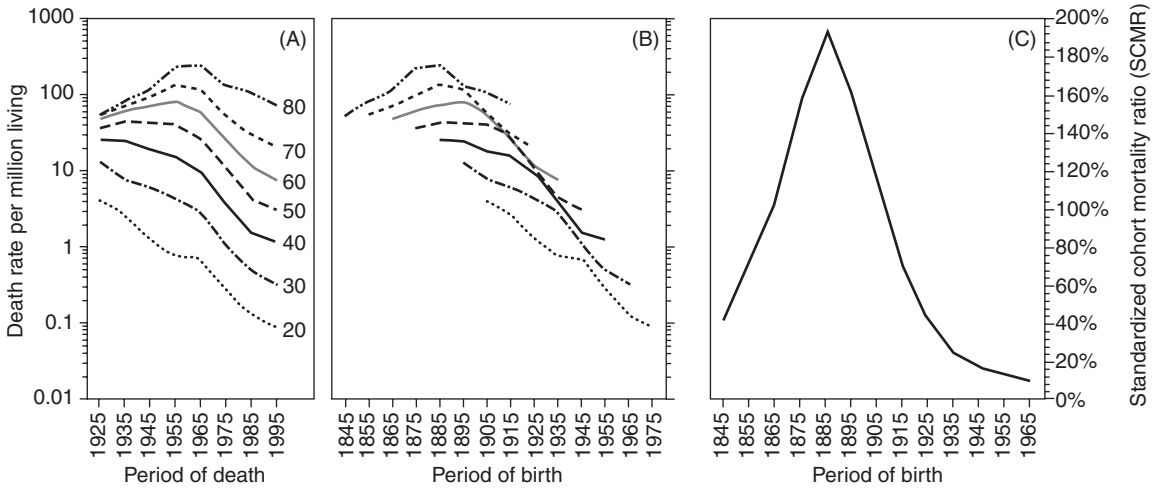


Figure 14.2 Age-specific death rates from duodenal ulcer in the USA between 1921 and 1998 plotted as (a) period-age contours, (b) as cohort-age contours, and (c) as standardized cohort mortality ratios. Every point of the standardized cohort mortality ratio curve represents an average (standardized) death rate among individuals of different ages born during the same time period. Thus (c) shows how successive birth cohorts during the early and mid-nineteenth century showed increasing duodenal ulcer mortality up to the birth cohort born around 1885. Subsequent birth cohorts experienced successively falling mortality. The different lines (full, dashed, dotted, etc.) refer to the same ages in both (a) and (b). Source: Cucino 2002 [63]. Reproduced with permission of Blackwell Publishing.

for one-third of the overall risk of bleeding ulcer and its complications [64–67], antiplatelet agents including clopidogrel and low-dose aspirin [65–68], selective serotonin reuptake inhibitors (SSRIs) [69,70], and oral anticoagulants [65,67]. Although smoking may also be a risk factor for ulcer disease [71], it has not been consistently shown to be an independent risk factor for ulcer-related complications [67].

Incidence and prevalence of peptic ulcer from the healthcare perspective

Studies of peptic ulcer incidence, that is, the frequency of new disease occurrences among individuals without a previous history, have been based on self-reports of physician-diagnosed ulcers, searches in healthcare archives, or a combination of both. The information is typically obtained through follow-up of defined cohorts/populations. A recent systematic review and meta-analysis of 31 studies that were published in the last three decades provided estimates regarding the incidence of peptic ulcer or peptic ulcer complications in the general population [72]. The pooled incidence per 1000 person-years was 0.90 (95 % CI 0.78–1.04) for peptic ulcer (with higher rates among men than among women), 0.57 (95 % CI 0.49–0.65) for peptic ulcer bleeding, and 0.10 (95 % CI 0.08–0.13) for peptic ulcer perforation. There was a statistically significant trend for a decrease in incidence rates by progressing calendar year. The variable sources of information contributed at least as much to the variation in study results as the genuine geographic variability. It is notable that studies that identified cases exclusively by medical records data or inpatient hospitalizations, or required validation of case ascertainment tended to report smaller incidence estimates for peptic ulcer.

Prevalence rates of peptic ulcer tend to be higher when based on self-reports of physician-diagnosed ulcers than when the information is obtained through searches in medical archives. This may reflect a higher sensitivity of self-reports, but it is also conceivable that their specificity is poorer. A recent systematic review [73] indentified only two studies that had assessed lifetime period prevalence of peptic ulcer in the general population. Both studies were conducted in the USA and were based on self-reports of physician's diagnoses from the National Health Interview Survey [74,75]. Lifetime prevalence of peptic ulcer was

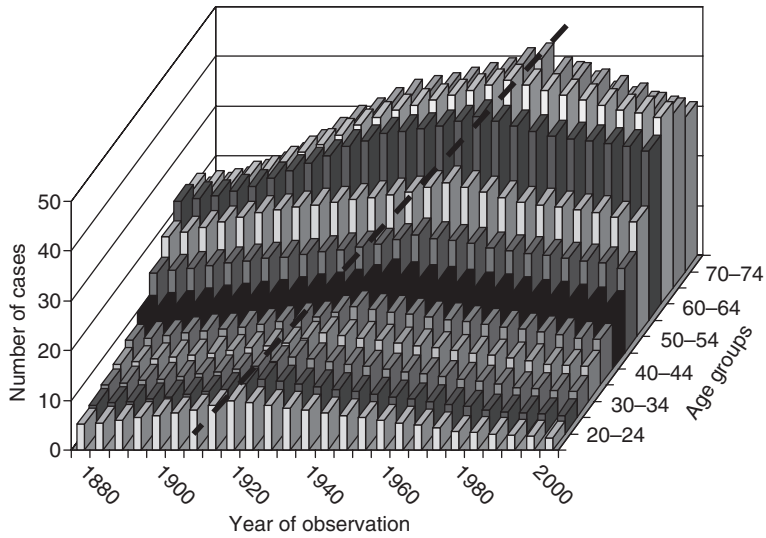


Figure 14.3 Hypothetical distribution of peptic ulcer cases across calendar years and age groups in a Western male population. There has been a peak among men born around 1900. These birth cohorts have accounted for the peak in all subsequent years, but for each consecutive year, the peak shifted towards older age groups as the high-risk birth cohorts grew older. The birth cohort phenomenon is visible as an oblique ridge in the 3D chart, marked with a dashed line. The occurrence increased with age in each consecutive year of observation, but declined in the oldest age groups because the oldest were dying off. However, because of the birth cohort effect, with declining rates in younger age groups and a shift of the high-risk birth cohorts into older ages, combined with a general increase in life expectancy, elderly people are presently much more dominating in the peptic ulcer population than they were in the distant past. Because the birth cohorts with the highest risk came approximately 20 years later in women than in men, the "peptic ulcer epidemic" does not yet seem to have culminated among elderly women, and there are so far no certain indications of a downward trend.

estimated to be 10.4 % (95 % CI 10–10.6 %) from the 1989 survey, and 8.4 % (CI was not reported) from the 1997–2003 surveys.

Healthcare utilization

Rates of hospitalizations and outpatient visits may only partly mirror the true epidemiology of the disease but are nonetheless of interest because they are indicators of the burden falling on health care. In 1995, 4 million physician visits for peptic ulcer were recorded (each patient could have visited a physician more than once), corresponding to a rate of 1500 per 100,000 of the US population per year [76]. This rate represents a marked decline since 1958, mainly due to a reduction of the visits for duodenal ulcers, while visits for gastric ulcer remained largely unchanged. Between the mid 1990s and 2005/2006 hospitalization rates for all ulcer types fell noticeably in the USA [77,78]. This downward trend was evident also in other Western countries during the same period, including Spain [79] and Sweden [80]. During this period, 30-day case-fatality from complicated peptic ulcers remained stable or changed slightly (not always in the same direction among studies). A recent systematic review found that the weighted mean of 30-day case-fatality from complicated peptic ulcers was 8.6 % (95 % CI 5.8–11.4) [67].

The average direct medical costs of peptic ulcer disease based on estimates from several Western countries were US$163–866 per patient; the costs of complicated peptic ulcers were US$1883–25,444 per patient [81]. The overall direct medical costs attributed to peptic ulcers was US$3.1 billion in the USA in 2002 [82], and US$29–94 million in Sweden in 2001 [83]. The indirect cost of work loss due to peptic ulcer disease was equivalent to US$1.37 billion in the USA in 1993 [84].

Peptic ulcer in Asian populations

Most of the literature on peptic ulcer epidemiology emanates from Western countries, but the disease occurs at an approximately equal rate also in the East. It appears that the rapid rise that was seen in the West around the turn of the century (when the birth cohorts with the highest risks came into their "ulcer ages") occurred simultaneously in the East. However, the decline in the East appears to have started considerably later than in the West [85]. The male-to-female ratio is higher and the duodenal-to-gastric ulcer ratio exhibits a greater variation in the East, compared to the West.

Risk factors for peptic ulcer: environmental exposures

A large body of literature concerns risk factors for peptic ulcer. Increasing age, male gender, and low socioeconomic status/income/educational attainment or underprivileged race/ethnic group are consistently linked to a higher risk, suggesting that factors linked to these circumstances, like *H. pylori* infection and smoking, may be etiologically important. Indeed, *H. pylori* infection, smoking, and aspirin/NSAID use are the overshadowing risk factors for both gastric and duodenal ulcer. In an excellent systematic review and meta-analysis of the literature up to 1995, Kurata and Nogawa [71] reported that the overall risk ratio for total peptic ulcer among *H. pylori*-infected individuals relative to uninfected was 3.3 (95 % CI 2.6–4.4). The risk ratio for serious upper GI events (bleeding, perforation or other GI events related to peptic ulcer disease resulting in hospitalization or death) among NSAID users relative to nonusers was 3.7 (95 % CI 3.5–3.9) with little variation between sexes and across age groups. The smoking-related overall risk ratio for peptic ulcer was 2.2 (95 % CI 2.0–2.3), again remarkably similar among men and women and among younger and older people. Using exposure prevalence rates from US populations, Kurata and Nogawa estimated population attributable risk percent (PAR) for *H. pylori* at 48 %. PAR expresses the percent of the studied outcome disease that can be attributed to the exposure under study, or in other words, the percent of all cases that might be prevented by eliminating this risk factor. The corresponding statistics for NSAID use and smoking, respectively, were 24 % and 23 %. Taken together, these three risk factors were thus deemed to be responsible for 89–95 % of the total peptic ulcer-related risk in the US general population [88]. The studies published after 1995 do not materially change the risk ratio estimates [67,86], but due to decreasing rates of *H. pylori* infection and smoking the PARs for these exposures are likely to be falling.

Indeed, recent studies have shown that up to 50 % of peptic ulcers in North America were *H. pylori*-negative, while in areas with higher prevalence of *H. pylori* infection, such as southern Europe and Asia, less than 5 % of the ulcers were *H. pylori*-negative [87,88]. The characteristics of patients with *H. pylori*-negative ulcers have not been adequately defined. The majority of these patients use aspirin or NSAIDs, surreptitiously on some occasions. Other recognized causes of ulceration in the upper GI tract (other ulcerogenic medications, malignancy/lymphoma, Crohn's disease, unusual infectious agents, gastrinoma) probably account for only a small proportion of *H. pylori*-negative NSAID/aspirin-negative ulcers, most of which currently remain "idiopathic". *H. pylori* prevalence is reportedly particularly low in complicated ulcer disease [89–91]. However, the prevalence of the infection may have been underestimated in patients with bleeding ulcers, due to reduced sensitivity of diagnostic tests during or soon after the bleeding episode [92].

Diet

Associations of duodenal and gastric ulcer with liver cirrhosis and pancreatic diseases suggest that alcohol may be a common underlying risk factor [93]. A similar link with high blood pressure and stroke indirectly implicates salt intake [94]; although difficult to quantify on an individual level, studies with direct assessment of salt intake support the importance of salt in the etiology of gastric ulcer [95,96]. No association between self-reported alcohol intake and risk of duodenal ulcer was found in a Swedish population-based case-control study [97], nor could alcohol intake be confirmed as a risk factor for any peptic ulcer type in a cohort of American men of Japanese ancestry in Hawaii [95]. An association between duodenal ulceration and a low fiber intake and a high refined carbohydrate diet has been reported, but the association with fiber intake was attenuated after control for confounding in a British study [98]. In a Swedish

cross-sectional study with careful dietary assessment [99], the presence of verified peptic ulcer was associated with a low intake of fruit and vegetables and consequently a low fiber and vitamin C intake, but no adjustments were made for potentially confounding factors. The latter study, further, found a positive association with regular intake of milk, possibly an expression of reversed causation. Intake of fermented milk, on the other hand, was associated with a reduced prevalence of peptic ulcer [99]. The consumption of fermented milk is relatively high in Sweden, and it has been speculated that lactobacilli in these products might have suppressed *H. pylori* growth. Analyses of the possible association of peptic ulcer with intake of fat and essential fatty acids have yielded conflicting results.

Psychological factors

Is psychological stress an important cause of peptic ulcer as has been widely believed? The evidence remains meager [100,101]. A population-based Swedish case-control study was unable to confirm any links with psychiatric morbidity, marital status, personal worries, type-A behavior, or experience of a hectic or psychologically demanding job in either sex [97]. Results from a Danish occupational cohort study indicated that low employment status and non-daytime work were associated with an increased risk of gastric ulcer [102], but confounding from socioeconomic status, particularly during childhood with possible consequences for *H. pylori* status, is difficult to rule out.

Genetic predisposition

Possible genetic components in the etiology of peptic ulcer disease have been addressed in several ways. A Finnish twin study found no more than modest familial aggregation but unveiled a significantly higher concordance among monozygotic than among dizygotic twin pairs [100]. Thirty-nine percent of the liability to peptic ulcer disease was explained by genetic factors and 61 % by individual environmental factors. Very little of the liability was explained by shared environmental factors. Thus, the familial aggregation was attributable almost solely to genetic factors, while environmental effects not shared by family members were dominating predictors of disease. Investigators of associations between genetically determined phenotypic expressions and presence of peptic ulcer disease noted in the 1950s a modest excess ulcer prevalence among subjects with the ABO blood group O, and among subjects with ABH nonsecretor status. A more recent Danish study [103] showed that carriers of ABO blood group A have a risk elevation that is comparable to that among individuals with blood group O. These investigators, and a Finnish group alike [104], found that people with the Lewis (a^+b^-) phenotype also have an increased risk of the same magnitude. The role of functional polymorphisms in genes that code for various cytokines involved in the inflammatory response to *H. pylori* infection has attracted considerable attention in recent years, but published studies have yielded mixed and partly contradictory results. Therefore, it appears that the relationship between polymorphisms of the *IL1* gene cluster and risk of peptic ulcer is incompletely understood at present. With the need for confirmation in rigorous epidemiologic studies in mind, it is worth mentioning that two studies have shown a positive association of carriage of the variant *A* allele of the *IL8* –251 locus with prevalence of gastric [105] and duodenal ulcer [106]. The gene product, IL-8, a major host mediator inducing neutrophil chemotaxis and activation, plays an important role in the pathogenesis of *H. pylori* infection.

Gastric cancer

Clinical outline

This section will only discuss adenocarcinoma, which is the dominating gastric neoplasm. Other types, such as lymphomas, carcinoids, and leiomyosarcomas account for less than 5 % of gastric neoplasms.

There are several classifications of gastric adenocarcinoma, but the one most used in epidemiologic research is that proposed by Laurén [107]. It distinguishes between two main histologic types: (i) the intestinal type, with glandular epithelium composed of absorptive cells and goblet cells; and (ii) the diffuse type, with poorly differentiated small cells in a dissociated noncohesive growth pattern. In addition, mixed and unclassifiable tumors occur. Adenocarcinomas occurring in the gastroesophageal junction or immediately below are referred to as gastric cardia cancers. There is no unanimous agreement about

which cancers to include in the latter category, and the definitions of cardia cancer vary between authors. The cardia cancers seem to behave differently compared with noncardia gastric cancer, both in terms of secular trends and risk factor pattern. This may be at least partly explained by heterogeneity among the cardia cancers; the cardia cancer category likely consists of a mix of genuine cardia cancer emanating from cardia epithelium (in view of the typical length of the segment occupied by such epithelium, the proportion of genuine cardia cancer is likely to be small), proximal noncardia gastric cancers that invade the gastroesophageal junction from below, and low esophageal adenocarcinomas that invade the same area from above. Unfortunately, there are no good morphologic or biochemical markers to help us distinguish between these tentative subgroups.

Stomach cancer has long belonged to the most deadly cancers. Five-year relative survival (i.e. survival adjusted for expected normal life expectancy) varied between 10 % and 20 % among patients diagnosed during the 1980s in the USA and Europe. This means that the survival at 5 years was no more than 10–20 % of the survival among the age-, sex- and calendar period-matched general population. Despite the lack of major therapeutic breakthroughs, there has been a noticeable improvement in the past 30 years [108]. In the USA, the 5-year relative survival has increased from 16 % in 1975–77 to 27 % in 1999–2006 [109]. This increase was statistically significant.

A disappearing disease?

As opposed to peptic ulcer, the incidence of stomach cancer is relatively easy to study thanks to the existence of well-functioning cancer registration in several countries or regions. In the USA it is easy to get the impression that stomach cancer is disappearing entirely. After having been the most common cancer until the 1940s, stomach cancer now ranks number 12 among men and number 14 among women as far as incidence is concerned [109]. In 2011, an estimated 13,120 and 8400 new cases of stomach cancer were diagnosed in men and women in the USA, respectively. In terms of deaths, stomach cancer ranks number 11 and 9 among US men and women, respectively, with 6260 and 4080 deaths [109]. Falling rates have been noted in most populations (Figure 14.4). The decline in the age-specific incidence of stomach cancer seems to have begun in the early 1930s in the Western Hemisphere and thereafter spread eastward. The secular trend seems to fit well with a log-linear model, that is, the incidence decreases by a fixed percentage each year [110]. As for peptic ulcer, the decline is best explained

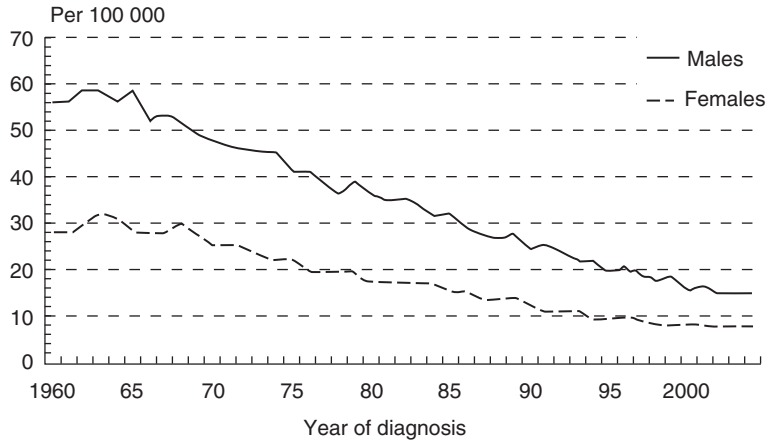


Figure 14.4 Swedish data on gender-specific incidence of stomach cancer 1960–2004. There is an unabated decline among both men and women. The secular trend seems to fit well with a log-linear model, that is, the incidence decreases by a fixed percentage each year. Source: Reproduced with permission from *Cancer Incidence in Sweden 2005*. Centre for Epidemiology. National Board of Health and Welfare. Official statistics of Sweden. Health and diseases 2005:9. Published at http://www.socialstyrelsen.se/Publicerat/2005/9042/2005-42-9.htm [January, 2006].

by a marked fall in incidence in successive birth cohort [110,111]. Notwithstanding this remarkable spontaneous global decline, stomach cancer, with an estimated 989,000 new cases in 2008, is still the fourth most frequent cancer worldwide, surpassed only by cancer of the lung, breast, and colorectum [112]. Due to its poor prognosis stomach cancer ranks second, after lung cancer among all causes of cancer death. In fact, with approximately 740,000 deaths annually, it accounts for 10 % of all cancer deaths. Over 70 % of the cases occur in developing countries. The worldwide estimates of age-adjusted incidence (22.0 per 100,000 person-years in men and 10.3 per 100,000 person-years in women in 2002) are about 15 % lower than the values estimated in 1985.

Geographic distribution

With reservations for possible differences in the availability of medical services, diagnostic methods and registration practices, the national incidence rates of stomach cancer vary approximately 10-fold, with the lowest reliable rates observed among North Americans (age-standardized incidence of 5.8 per 100,000 person-years in men and 2.8 per 100,000 person-years in women, 2008) and the highest in Japan, where screening is ongoing. The age-standardized incidence in East Asia was 42.4 per 100,000 person-years in men and 18.3 per 100,000 person-years in women [112]. With few exceptions, the incidence among women is approximately half that among men, regardless of geographic area, culture, and religion. While the risk of stomach cancer seems to co-vary with socioeconomic conditions, there is no clear correlation between national level of economic development and national incidence rates. However, suspected underreporting may have deflated figures from poorly developed countries. Although the highest rates are observed in East Asia, low rates (<10 per 100,000 person-years) are reported from South and Southeast Asia and from Africa. High incidence rates also are found in tropical Central and South America and in Central/Eastern Europe [112].

Demographic distribution

In the USA the incidence is twice as high in African Americans as in white people, and three to six times higher among Japanese Americans than among US-born white people [113]. Immigrant Koreans have an incidence that is eightfold higher than that among white people [114], while the incidence among Filipino men, regardless of birthplace, is only 60 % that of US-born white males [113]. Another example of marked differences within a limited geographic area comes from Singapore, where the incidence rates among men of Malay and Chinese descent vary more than threefold [115]. When people move between populations with different risks of stomach cancer, their risk patterns are usually retained or only slightly modified, regardless of their country of origin and country of destination. In the succeeding generation, the rates adjust to that prevailing in the new environment, but this adaptation appears to be somewhat slower for stomach cancer than for colorectal and some other cancers. Though the patterns of risk in relation to migration are complex and defy simple dietary or other interpretation, it appears that early-life exposures are important for the future risk of gastric cancer.

Opposing secular trend for cardia cancer?

While the decline in incidence of gastric carcinoma overall has abated in the USA [116], a closer look at the data reveals two coinciding trends: the steep downward trend seems to persist for distal stomach cancer in white men and women, but this decline is balanced by an increase in the incidence of cardia cancer [117]. Increasing incidence rates of cardia cancer have been noted in a number of cancer registers in Europe and the USA in the past 20–30 years. However, considerable misclassification of the site within the stomach has been demonstrated [118]; following careful classification of all tumors, no increasing trend could be confirmed for cardia cancer [119]. Some other studies have also failed to verify any upward trend, and even in the USA, the trend seems to have leveled off in the 1990s [116]. Regardless of whether the incidence curve for cardia cancer is flat or turning up, it clearly differs from the descending one for distal gastric cancer.

Risk factors for stomach cancer

Helicobacter pylori

In the past 20 years, numerous observational studies of various designs have demonstrated a positive

association between presence of anti-*H. pylori* antibodies and risk of stomach cancer. A review of published meta-analyses [120] showed that serologic evidence of *H. pylori* infection is associated with pooled odds ratios of stomach cancer ranging between 1.92 and 2.56, with little heterogeneity. In other words, carriers of antibodies to *H. pylori* allegedly run a risk for stomach cancer that is 2–3 times higher than among people without such antibodies. However, because some infections disappear spontaneously due to changes in the gastric microenvironment during the precancerous stages, it looks as if the strength of the association with stomach cancer risk may be underestimated [121,122]. Moreover, it appears that the association is confined to noncardia gastric cancer, whereas the infection might even be inversely related to the risk of cardia cancer [123,124]. In studies that restricted the outcome to noncardia stomach cancer and which took measures to overcome the misclassification of exposure, the relative risk linked to the infection was 20-fold or greater [122,125,126]. According to such studies, the PAR may be 70 % or higher even in Western populations [122], while an American case-control study with conventional serotesting reported a PAR of 10.4 % [127]. The risk seems to be particularly elevated among carriers of CagA-positive strains (and among carriers of CagA-positive strains in those with strains having the "A-B-D-type" CagA typically seen in Asian high-risk populations [128]), although CagA-negative strains are not without risk [129]. The *vacA* gene of *H. pylori*, encoding a vacuolating cytotoxin, comprises two variable regions: the s (signaling) and the m (mid) regions. *H. pylori vacA* type s1 and m1 strains appear to be more carcinogenic than strains with other *vacA* types [130]. Although the ultimate proof of causality is still missing, a growing number of randomized trials have either shown trends towards reduced gastric cancer incidence or indications of slowing progression of precancerous lesions after *H. pylori* eradication [131–137], thus gradually adding to our confidence in a causal inference.

Smoking

A relationship between smoking and risk of stomach cancer is well established [138,139]. The excess risk among current smokers is 1.5–2.5-fold and increases with higher doses and/or longer duration of cigarette smoking [140–142]. It appears that the risk returns to baseline relatively soon after quitting smoking [143], but in a pooled analysis of two Japanese cohorts, a significant risk elevation remained for up to 14 years after cessation [141]. While some studies suggest that smoking is more strongly related to cardia cancer risk [140,142,144,145], others indicate that the link with distal stomach cancer is not appreciably weaker, and in Japan it might even be stronger [141]. The PAR for smoking varies with the exposure prevalence, and thus between men and women, but within sexes the variation between American and European data is surprisingly small; thus, the PAR among men varied between 21.5 % and 28.6 %, and among women between 11 % and 14 % in three recent studies emanating from the USA and Europe [127,140,146].

Alcohol

The most authoritative review of the literature published up until the mid-1990s [147] noted that the bulk of evidence weighed against the possibility of a substantial effect of alcohol consumption on the risk of stomach cancer. The literature on the relationships between the different types of alcoholic beverages and stomach cancer risk has been reviewed, but no consistent pattern was found [148]. However, a recent large multicenter prospective cohort study in Europe found that heavy (but not light or moderate) consumption of alcohol (mainly from beer) was associated with an increased risk of intestinal type of noncardia gastric cancer in men [149]. A meta-analysis in 2012 [150] arrived at a modestly increased summary relative risk estimate (1.14, 95 % CI 1.08–1.21; and 1.45, 95 % CI 1.31–1.62, for intake of 50 g and 100 g alcohol per day, respectively, relative to no intake).

Diet

Until recently, the most consistent nutritional epidemiology finding in relation to stomach cancer has been inverse associations with fruit and vegetable intake. In 1997, an international expert panel at the World Cancer Research Fund-American Institute for Cancer Research concluded that there was convincing evidence that high intake of vegetables, particularly raw vegetables and allium vegetables, reduces the risk of stomach cancer [147]. A similar conclusion was also drawn with regard to high fruit intake. A more recent meta-analysis, however, noted that the protective effect seemed to be weaker in cohort investigations than in case-control studies [151], suggesting

that recall bias might have pushed the relative risk estimates away from the null value in the latter. The estimated overall relative risks that were based on all study types were 0.81 (95 % CI 0.75–0.87) and 0.74 (95 % CI 0.69–0.81) per 100 g intake per day of vegetables and fruit, respectively. Although heterogeneity was observed in essentially all analyzed substrata, the estimates for both fruit and vegetables were always less than unity [151]. A recent large European multinational cohort study (European Prospective Investigation into Cancer and Nutrition, EPIC) with careful dietary assessments and a fairly wide range of exposure [149], failed to verify any overall association of stomach cancer risk with total or category-specific vegetable or fruit intake. The most recent addition to the literature, a large Dutch cohort study, also failed to show a protective effect of total vegetable or fruit intake on stomach cancer risk (an inverse association was found only between citrus fruit and gastric cardia adenocarcinoma) [152]. Even though the estimation of portion size and frequency of consumption of a wide range of vegetables is rather difficult and non-differential misclassification may bias the relative risk estimates toward the null value, it is reasonable to assume that the more recent studies, particularly the cohort studies with increasingly sophisticated dietary assessments, are less affected by such bias compared with earlier studies. Therefore, it must be suspected that previous research may have overestimated the protection conferred by these plant foods.

Moreover, whereas there is almost total consensus among case-control studies that vitamin C intake is strongly protective, only one [153] out of four prospective studies [153–156] reported a significant inverse association between estimated vitamin C intake and stomach cancer. The summary estimate of relative risk in a meta-analysis, however, was still statistically significant (relative risk among subjects with the highest intake, relative to those with the lowest, was 0.77 (95 % CI 0.61–0.97) [157]. A similar meta-analysis of the three prospective studies concerned with pre-disease blood levels of vitamin C [155,158,159] yielded a summary estimate that was also statistically significant (RR 0.64; 95 % CI 0.41–0.98) [157].

Vitamin E (tocopherol), another important antioxidant in plant foods, has been investigated with regard to its relationship with stomach cancer risk in at least 18 case-control studies, close to half of which reported a statistically significant inverse association while the others were unable statistically to confirm any relationship at all. Among four prospective studies that related estimated dietary vitamin E intake with stomach cancer risk, only one – conducted among Finnish smokers [154] – showed a significantly reduced risk of noncardia stomach cancer among individuals with the highest intake, but this study also noted an *increased* risk for cancer of the gastric cardia. Six prospective studies that proceeded from pre-disease blood levels of tocopherols yielded mixed results; a recent large European study reported a strong and statistically significant inverse relationship with stomach cancer risk, albeit seemingly limited to the diffuse histologic type [160], while a Chinese study showed a *positive* association with noncardia cancer risk but no relationship with cardia cancer [161]. Thus, the effect of vitamin E on risk of stomach cancer remains uncertain.

At least 15 case-control studies have addressed the relationship between intake of total vitamin A (retinol and provitamin A carotenoids) and risk of stomach cancer, and the overwhelming majority of them have shown a trend towards an inverse association (in five such studies this trend was statistically significant) [157]. The results of prospective studies, particularly those that examined associations specifically with retinol or β-carotene, have been somewhat less persuasive [157].

Unfortunately, not even randomized intervention trials have been able to provide an unambiguous answer regarding the protective effect of the antioxidative vitamins in plant foods. Two such studies argue in favor of a protective effect; a Chinese study, performed in subjects who were likely to be vitamin deficient, showed a reduced incidence of gastric cancer mortality after administration of a combination of β-carotene, vitamin E and selenium [162]. In the other study, carried out in South America [131], treatment with either β-carotene or ascorbic acid significantly increased the rates of regression of atrophic gastritis and intestinal metaplasia. However, two other randomized intervention studies, conducted among Finnish male smokers [163,164] and American male physicians [165], respectively, showed no effect on prevention of stomach cancer incidence during or after supplementation with either β-carotene or α-tocopherol, the most active form of vitamin E. Moreover, two additional randomized Chinese intervention studies did not observe any significant

reductions in stomach cancer incidence or mortality after daily supplementation with 14 vitamins and 12 minerals for 6 years [166,167] or a combination of vitamin C, vitamin E, and selenium every second day for 7.3 years [136].

Fiber and carbohydrates

Several investigators have found a decreased risk of stomach cancer among people with a high consumption of fiber. A particularly strong inverse association has been demonstrated between cereal fiber intake and risk of cardia cancer [168], possibly attributable to the nitrite scavenging properties of wheat fiber. However, the only prospective study addressing the relationship between intake of whole-grain foods and stomach cancer mortality was negative [169]. High-starch/carbohydrate diets, on the other hand, were reportedly linked to an increased risk of stomach cancer in some studies, but others showed no association. It is conceivable that the association noted in the positive studies may be explained by residual confounding by socioeconomic status.

Salt

Most textbooks list salt intake as an established risk factor for stomach cancer. Ecological studies provide support for a relatively strong correlation between urinary salt excretion and stomach cancer mortality [170–172]. Further, there are abundant case-control and cohort data on intake of salt or salty foods and risk of stomach cancer. Although the results are somewhat divergent, the bulk of evidence weighs towards a positive association, albeit not particularly strong. However, confounding is a major concern; in some of the studied populations, consumption of salted foods may have correlated inversely with socioeconomic status, access to refrigeration and consumption of fruits and vegetables, and positively with the prevalence of *H. pylori* infection. Moreover, salted foods tend to contain significant amounts of *N*-nitroso compounds (NOCs), which may be the true culprits. The relative risk estimates in cohort and case-control studies are mostly in the range where undetected confounding might well explain the association. It should also be noted that there is no laboratory evidence that salt *per se* is a carcinogen for any site of the body [173].

N-*nitroso compounds*

N-nitroso compounds (NOCs) have been found to be carcinogenic in multiple organs in at least 40 animal species. Humans are exposed to NOCs from diet (processed meats, smoked preserved foods, pickled and salty preserved foods, and foods dried at high temperatures such as the constituents of beer, whisky, and dried milk), tobacco smoke and other environmental sources, but a large proportion (typically more than 50 %) comes from endogenous synthesis. The results of epidemiologic investigations addressing the possible association between estimated nitrite exposure (the precursor substance) and stomach cancer risk have been mixed. Similarly, studies of estimated NOC intake in relation to stomach cancer risk have yielded discrepant results, although the majority of case-control studies suggested a positive association. A recent analysis of data from the EPIC study [174] estimated exposure to endogenously formed NOCs and found a statistically significant association with the risk of noncardia stomach cancer (relative risk associated with a 40 μg day^{-1} increase in endogenous NOC exposure was 1.42, 95 % CI 1.14–1.78) but not with the risk of cardia cancer. Thus, the epidemiologic literature has been unable to unequivocally confirm a link between nitrite or NOC exposure and risk of gastric cancer, but the data are clearly suggestive of such a link.

Meat intake

Whilst meat consumption has been associated with increased risks of cancer of the colorectum, breast, and possibly prostate, the epidemiologic evidence for a relationship with stomach cancer risk has so far been considered insufficient. However, recent cohort studies have reported substantial risk elevations among subjects in the highest intake categories, relative to those in the lowest. A meta-analysis that encompassed six prospective cohort studies and nine case-control studies [175] arrived at a summary estimate of relative risk for stomach cancer per 30 g day^{-1} increase in processed meat consumption of 1.15 (95 % CI 1.04–1.27) among cohort studies and 1.38 (95 % CI 1.19–1.60) among case-control studies. Thus, it appears that high intake of processed meat should be added to the list of known – but moderately strong – risk factors for stomach cancer. In the EPIC cohort, the association with processed meat was confined to noncardia stomach cancer, with a relative risk of 2.45 for every 50 g day^{-1} increase in processed meat intake [176]. The latter study also noted positive associations with nonprocessed meat.

Aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs)

Several epidemiologic studies have noted small reductions in risk of stomach cancer among users of aspirin and/or NSAIDs. A meta-analysis of observational studies (13 case-control studies and eight cohort studies) yielded a summary adjusted risk ratio of 0.81(95 % CI 0.73–0.89) [177]. The risk of gastric cancer was significantly reduced in all subgroups analyses (according to study design, type of drug, site of cancer, or sample source). A placebo-controlled randomized trial among 40,000 US women failed to provide evidence of a protective effect of low-dose aspirin on the risk of stomach cancer over a duration of 10 years [178]. However, the study did not have enough power to prove lack of an effect. The confidence interval for the effect (RR 1.00; 95 % CI 0.42–2.40) was not tight enough to rule out clinically important risk modifications due to aspirin use.

Genetic risk factors

Familial aggregation of stomach cancer has been reported in the epidemiologic literature. Typically, a 50–130 % excess risk was observed among subjects with a positive family history. In an analysis of 44,788 Scandinavian twin pairs, the risk of stomach cancer among dizygotic twins with a partner who developed the same cancer was 6.6 times higher than among dizygotic twins whose partner did not have stomach cancer [179]. The corresponding excess was 10-fold in monozygotic pairs. It was estimated that inherited genes contribute 28 % to the risk of stomach cancer, shared environmental effects contribute 10 %, and nonshared environmental factors make up the remaining 62 % of the risk. Therefore, studies on twins predict the involvement of major environmental factors plus minor genetic components.

An aggregation of two or more stomach cancers in the same family is noted in about 10 % of all stomach cancer cases. Among them, a number of syndromes can be identified; the most notable is the hereditary diffuse gastric cancer (HDGC – requiring two or more documented cases of diffuse stomach cancer in first/second-degree relatives, with at least one diagnosed before the age of 50; *or* three or more cases of documented diffuse stomach cancer in first/second-degree relatives, independently of age) [180]. The term "familial diffuse gastric cancer" (FDGC) is used for families with aggregation of stomach cancer and an index case with diffuse stomach cancer, but not otherwise fulfilling the criteria for HDGC, for instance due to unknown histologic type of the related cases. In a recent review of the accumulated literature [181], HDGC and FDGC accounted for 27 % and 24 %, respectively, of 439 screened families with familial aggregation of stomach cancer. Germline truncating mutations in the gene for the cell–cell adhesion protein E-cadherin (*CDH1*) were found in 36 % of families with HDGC and in 13 % of families with FDGC. In about two-thirds of HDGC families, a large proportion of FDGC families, and in the majority of families with aggregation not fulfilling criteria for HDGC or FDGC, cancer susceptibility is caused by presently unknown genetic defects.

The literature on genetic polymorphisms and stomach cancer risk is limited by a common lack of appropriate control of potential sources of bias; few studies are population-based, and the sample sizes are often insufficient even for the statistical verification of moderate main effects, let alone gene–environment interactions. Besides, information on exposure to relevant cofactors such as *H. pylori* infection, diet, and smoking is often lacking. The role of functional polymorphisms in genes that code for various cytokines involved in the inflammatory response to *H. pylori* infection has attracted considerable attention in recent years. One of the key cytokines is interleukin-1 beta (IL-1β), which is an important driving force in the inflammatory responses and also a potent inhibitor of gastric acid secretion. The *IL1B* gene encoding IL-1β is highly polymorphic. Two of the polymorphisms are in the promoter region at positions –511 and –31, representing C→T and T→C transitions, respectively. The variant alleles of these loci are associated with more severe inflammation. Another cytokine that has an important influence on IL-1β levels is the endogenous interleukin-1 receptor antagonist (IL-1ra), whose gene (*IL1RN*) is also known to be polymorphic. The *IL1RN* gene has a penta-allelic 86-bp tandem repeat polymorphism (variable number of tandem repeat, VNTR) in intron 2, of which the less common allele 2 (*IL1RN*2*) – associated with enhanced IL-1β production *in vitro* – is linked to several chronic inflammatory conditions. In a landmark case-control study from Poland, El-Omar and co-workers [182] demonstrated that carriers of the C allele of *IL1B* –31 (in positive linkage disequilibrium with *IL1B* –511T) and homozygous carriers of the *2 allele of *IL1RN* had

1.6- and 2.9-fold increased risks, respectively, of stomach cancer, compared with noncarriers of these variant alleles. Carriers of the *IL1B* –31T/*IL1RN**2 haplotype had an odds ratio of 4.4. These findings have been more or less replicated in several populations, among them Portuguese, American, Mexican, Italian, and Chinese. A study from Portugal with genotyping of archived gastric biopsies suggested that the combination of proinflammatory genotypes in the host with infection with high-risk *H. pylori* strains (see previous section about *H. pylori*) might involve major increases in risk, with relative risks as high as 87 [130]. However, others, including investigators from Japan, China, Taiwan, Korea, Holland, Italy, Finland, and Sweden, have failed to confirm any association of the *IL1B* –31, *IL1B* –511 and/or *IL1RN* polymorphisms with stomach cancer. It appears that the relationship, if any, between polymorphisms in the *IL1* gene cluster and stomach cancer risk may be more complex than first thought.

There is a large and rapidly expanding literature on links between genetic variation in a number of potentially important carcinogenic pathways (mucin production, cytokines other than those in the *IL1* cluster, human leukocyte antigen (HLA) classes I and II, metabolic phase I and II enzymes, DNA repair systems, cyclooxygenase system, oncogenes, and tumor suppressor genes) and risk of stomach cancer [183]. Unfortunately, the overall results become increasingly disappointing as the literature accumulates; notwithstanding the often apparent biologic plausibility, the results are remarkably divergent. Typically, promising reports of fairly large effects are followed by null or opposite findings. Whether this diversity is mainly due to an apparent variation in epidemiologic rigor, laboratory measurement errors, or to effect modification by race, ethnicity, or other external exposures cannot be determined at present. The positive findings that remain unopposed tend to be the ones that have been tested in no more than one single study. And this could, in turn, be a result of publication bias because negative studies are difficult to get published. There is an urgent need for more population-based studies with meticulous attention to epidemiologic fallacies. Carefully conducted meta-analyses of epidemiologically sound studies may also be helpful.

One notable exception is the literature on polymorphisms in the gene coding for the enzyme 5,10-methylenetetrahydrofolate reductase (MTHFR). The enzyme irreversibly converts 5,10-methylene tetrahydrofolate to 5-methyltetrahydrofolate, the predominant form of folate in the circulation. Folate is a water-soluble B vitamin that plays an important role in the maintenance of DNA integrity. Increasing evidence suggests that a low folate intake and/or an impaired folate metabolism may be implicated in the development of gastrointestinal cancers. Two common functional polymorphisms of the *MTHFR* gene, 677C/T and 1298A/C, have been identified, associated with up to 70 % and 40 % reductions, respectively, of MTHFR activity among individuals who are homozygous for the variant alleles. A recent meta-analysis of 11 case-control and two cohort studies that examined the association between dietary folate intake and risk of stomach cancer arrived at statistically significant 30 % risk reductions for stomach cancer of both noncardia and cardia location [184]. However, this inverse relationship was confined to studies conducted in the USA and Europe, while studies done in other populations were essentially negative. The summary estimate of relative risk for stomach among individuals with the variant TT genotype of *MTHFR* –677, relative to those with the CC genotype, was 1.68 (95 % CI 1.29–2.19) and the corresponding estimate for gastric cardia cancer was 1.90 (95 % CI 1.38–2.60) [175]. Available studies of the 1298A/C polymorphism did not provide any indications of a statistical relationship with stomach cancer risk.

Multiple choice questions

1 The two most important risk factors for *H. pylori* infection are:

A Acid suppression medication and alcohol consumption
B Smoking and diet
C Age and socioeconomic status
D Race and gender
E Genetic factors and immune deficiency status

2 Peptic ulcer disease

A is inversely associated with regular milk ingestion
B is more common among carriers of ABO blood group B
C has a prevalence of less than 0.5 % among the general population in Western countries

D reached a peak prevalence in the early nineteenth century

E leads to significantly less hospitalizations nowadays compared to the early 1990s despite the increasing use of NSAIDs and aspirin

3 Gastric adenocarcinoma,

A worldwide, is the fourth most frequent cancer and ranks second among all causes of cancer death

B worldwide, is the second most frequent cancer, and ranks fourth among all causes of cancer death

C in the United States, is the fourth most frequent cancer, and ranks second among all causes of cancer death

D in the United States, is the second most frequent cancer, and ranks fourth among all causes of cancer death

E in most populations, is approximately twice as common among women compared to men

References

1 Baltrus DA, et al. *Helicobacter pylori* genome plasticity. *Genome Dyn* 2009;6:75.

2 Cover TL, et al. Characterization of and human serologic response to proteins in *Helicobacter pylori* broth culture supernatants with vacuolizing cytotoxin activity. *Infect Immun* 1990;58:603.

3 Blaser MJ, et al. CagA and the outcome of *Helicobacter pylori* infection. *Am J Clin Pathol* 1996;106:565.

4 Kersulyte D, et al. Emergence of recombinant strains of *Helicobacter pylori* during human infection. *Mol Microbiol* 1999;31:31.

5 Odenbreit S, et al. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* 2000;287:1497.

6 Backert S, et al. Molecular mechanisms of gastric epithelial cell adhesion and injection of CagA by *Helicobacter pylori*. *Cell Commun Signal* 2011;9:28.

7 Occhialini A, et al. Composition and gene expression of the cag pathogenicity island in *Helicobacter pylori* strains isolated from gastric carcinoma and gastritis patients in Costa Rica. *Infect Immun* 2001;69:1902.

8 Israel DA, et al. *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. *J Clin Invest* 2001;107:611.

9 Malfertheiner P, et al. Current concepts in the management of *Helicobacter pylori* infection: the Maastricht III Consensus Report. *Gut* 2007;56:772.

10 Chey WD, et al. American College of Gastroenterology guideline on the management of *Helicobacter pylori* infection. *Am J Gastroenterol* 2007;102:1808.

11 IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Schistosomes, liver flukes and *Helicobacter pylori*. *IARC Monogr Eval Carcinog Risks Hum* 1994;61:1.

12 Ford AC, et al. Eradication therapy for peptic ulcer disease in *Helicobacter pylori* positive patients. *Cochrane Database Syst Rev* 2006;CD003840.

13 Zullo A, et al. Eradication therapy for *Helicobacter pylori* in patients with gastric MALT lymphoma: a pooled data analysis. *Am J Gastroenterol* 2009;104:1932.

14 Kusters JG, et al. Pathogenesis of *Helicobacter pylori* infection. *Clin Microbiol Rev* 2006;19:449.

15 Leontiadis GI, et al. Non-gastrointestinal tract associations of *Helicobacter pylori* infection. *Arch Intern Med* 1999;159:925.

16 DuBois S, et al. Iron-deficiency anemia and *Helicobacter pylori* infection: a review of the evidence. *Am J Gastroenterol* 2005;100:453.

17 Linz B, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 2007;445:915.

18 Parsonnet J. The incidence of *Helicobacter pylori* infection. *Aliment Pharmacol Ther* 1995;9(Suppl 2):45.

19 Goh KL, et al. Epidemiology of *Helicobacter pylori* infection and public health implications. *Helicobacter* 2011;16(Suppl 1):1.

20 Torres J, et al. A comprehensive review of the natural history of *Helicobacter pylori* infection in children. *Arch Med Res* 2000;31:431.

21 Clemens J, et al. Sociodemographic, hygienic and nutritional correlates of *Helicobacter pylori* infection of young Bangladeshi children. *Pediatr Infect Dis J* 1996;15:1113.

22 Goodman KJ, et al. Transmission of *Helicobacter pylori* among siblings. *Lancet* 2000;355:358.

23 Hestvik E, et al. *Helicobacter pylori* in apparently healthy children aged 0–12 years in urban Kampala, Uganda: a community-based cross sectional survey. *BMC Gastroenterol* 2010;10:62.

24 Khanna B, et al. Use caution with serologic testing for *Helicobacter pylori* infection in children. *J Infect Dis* 1998;178:460.

25 Goodman KJ, et al. Dynamics of *Helicobacter pylori* infection in a US-Mexico cohort during the first two years of life. *Int J Epidemiol* 2005;34:1348.

26 Genta RM. Review article: after gastritis – an imaginary journey into a *Helicobacter*-free world. *Aliment Pharmacol Ther* 2002;16(Suppl 4):89.

27 Banatvala N, et al. The cohort effect and *Helicobacter pylori*. *J Infect Dis* 1993;168:219.
28 Cullen DJ, et al. When is *Helicobacter pylori* infection acquired? *Gut* 1993;34:1681.
29 den Hoed CM, et al. *Helicobacter pylori* and the birth cohort effect: evidence for stabilized colonization rates in childhood. *Helicobacter* 2011;16:405.
30 Kumagai T, et al. Acquisition versus loss of *Helicobacter pylori* infection in Japan: results from an 8-year birth cohort study. *J Infect Dis* 1998;178:717.
31 Akre K, et al. Risk for gastric cancer after antibiotic prophylaxis in patients undergoing hip replacement. *Cancer Res* 2000;60:6376.
32 Perez-Perez GI, et al. Evidence that cagA(+) *Helicobacter pylori* strains are disappearing more rapidly than cagA(−) strains. *Gut* 2002;50:295.
33 Melo ET, et al. Seroprevalence of *Helicobacter pylori* antibodies in medical students and residents in Recife, Brazil. *J Clin Gastroenterol* 2003;36:134.
34 Triantafillidis JK, et al. *Helicobacter pylori* infection in hospital workers over a 5-year period: correlation with demographic and clinical parameters. *J Gastroenterol* 2002;37:1005.
35 Weck MN, et al. Apparent incidence of *Helicobacter pylori* in adulthood: to what extent do new infections reflect misclassification? *Helicobacter* 2011;16:266.
36 Malaty HM, et al. Age at acquisition of *Helicobacter pylori* infection: a follow-up study from infancy to adulthood. *Lancet* 2002;359:931.
37 Kivi M, et al. Concordance of *Helicobacter pylori* strains within families. *J Clin Microbiol* 2003;41:5604.
38 Kivi M, et al. *Helicobacter pylori* status in family members as risk factors for infection in children. *Epidemiol Infect* 2005;133:645.
39 Tindberg Y, et al. *Helicobacter pylori* infection in Swedish school children: lack of evidence of child-to-child transmission outside the family. *Gastroenterology* 2001;121:310.
40 Fialho AM, et al. Younger siblings play a major role in *Helicobacter pylori* transmission among children from a low-income community in the Northeast of Brazil. *Helicobacter* 2010;15:491.
41 Weyermann M, et al. Acquisition of *Helicobacter pylori* infection in early childhood: independent contributions of infected mothers, fathers, and siblings. *Am J Gastroenterol* 2009;104:182.
42 Dore MP, et al. Risk factors associated with *Helicobacter pylori* infection among children in a defined geographic area. *Clin Infect Dis* 2002;35:240.
43 Malaty HM, et al. *Helicobacter pylori* infection: genetic and environmental influences. A study of twins. *Ann Intern Med* 1994;120:982.
44 Azuma T, et al. Genetic differences between duodenal ulcer patients who were positive or negative for *Helicobacter pylori*. *J Clin Gastroenterol* 1995;21(Suppl 1):S151.
45 Magnusson PKE, et al. Gastric cancer and human leukocyte antigen: distinct DQ and DR alleles are associated with development of gastric cancer and infection by *Helicobacter pylori*. *Cancer Res* 2001;61:2684.
46 Zambon CF, et al. Pro- and anti-inflammatory cytokines gene polymorphisms and *Helicobacter pylori* infection: interactions influence outcome. *Cytokine* 2005;29:141.
47 Aspholm-Hurtig M, et al. Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* 2004;305:519.
48 Malaty HM, et al. Natural history of *Helicobacter pylori* infection in childhood: 12-year follow-up cohort study in a biracial community. *Clin Infect Dis* 1999;28:279.
49 Moayyedi P, et al. Relation of adult lifestyle and socioeconomic factors to the prevalence of *Helicobacter pylori* infection. *Int J Epidemiol* 2002;31:624.
50 Nakajima S, et al. Changes in the prevalence of *Helicobacter pylori* infection and gastrointestinal diseases in the past 17 years. *J Gastroenterol Hepatol* 2010;25(Suppl 1):S99.
51 Murray LJ, et al. Inverse relationship between alcohol consumption and active *Helicobacter pylori* infection: the Bristol Helicobacter project. *Am J Gastroenterol* 2002;97:2750.
52 Gao L, et al. Alcohol consumption, serum gamma-glutamyltransferase, and *Helicobacter pylori* infection in a population-based study among 9733 older adults. *Ann Epidemiol* 2010;20:122.
53 Zhang L, et al. Relationship between alcohol consumption and active *Helicobacter pylori* infection. *Alcohol Alcohol* 2010;45:89.
54 Bazzoli F, et al. The Loiano-Monghidoro population-based study of *Helicobacter pylori* infection: prevalence by 13C-urea breath test and associated factors. *Aliment Pharmacol Ther* 2001;15:1001.
55 Fontham ET, et al. Determinants of *Helicobacter pylori* infection and chronic gastritis. *Am J Gastroenterol* 1995;90:1094.
56 Goodman KJ, et al. Nutritional factors and *Helicobacter pylori* infection in Colombian children. *J Pediatr Gastroenterol Nutr* 1997;25:507.
57 Russo A, et al. Determinants of *Helicobacter pylori* seroprevalence among Italian blood donors. *Eur J Gastroenterol Hepatol* 1999;11:867.
58 Sjunnesson H, et al. High intake of selenium, beta-carotene, and vitamins A, C, and E reduces growth

of *Helicobacter pylori* in the guinea pig. *Comp Med* 2001;51:418.
59 McIntosh JH, et al. Patterns of dyspepsia during the course of duodenal ulcer. *J Clin Gastroenterol* 1991;13:506.
60 Inadomi JM, et al. The impact of peptic ulcer disease and infection with *Helicobacter pylori* on life expectancy. *Am J Gastroenterol* 1998;93:1286.
61 Lassen A, et al. Complicated and uncomplicated peptic ulcers in a Danish county 1993–2002: a population-based cohort study. *Am J Gastroenterol* 2006;101: 945.
62 Aro P, et al. Peptic ulcer disease in a general adult population: the Kalixanda study: a random population-based study. *Am J Epidemiol* 2006;163:1025.
63 Cucino C, et al. The long-term time trends of peptic ulcer and ulcerative colitis are interrelated. *Am J Gastroenterol* 2002;97:2657.
64 Langman M. Population impact of strategies designed to reduce peptic ulcer risks associated with NSAID use. *Int J Clin Pract Suppl* 2003;38.
65 Lanas A, et al. Effect of antisecretory drugs and nitrates on the risk of ulcer bleeding associated with nonsteroidal anti-inflammatory drugs, antiplatelet agents, and anticoagulants. *Am J Gastroenterol* 2007;102: 507.
66 Bhatt DL, et al. ACCF/ACG/AHA 2008 expert consensus document on reducing the gastrointestinal risks of antiplatelet therapy and NSAID use. *Am J Gastroenterol* 2008;103:2890.
67 Lau JY, et al. Systematic review of the epidemiology of complicated peptic ulcer disease: incidence, recurrence, risk factors and mortality. *Digestion* 2011;84:102.
68 Garcia Rodriguez LA, et al. Risk of upper gastrointestinal bleeding with low-dose acetylsalicylic acid alone and in combination with clopidogrel and other medications. *Circulation* 2011;123:1108.
69 Loke YK, et al. Meta-analysis: gastrointestinal bleeding due to interaction between selective serotonin uptake inhibitors and non-steroidal anti-inflammatory drugs. *Aliment Pharmacol Ther* 2008;27:31.
70 Targownik LE, et al. Selective serotonin reuptake inhibitors are associated with a modest increase in the risk of upper gastrointestinal bleeding. *Am J Gastroenterol* 2009;104:1475.
71 Kurata JH, et al. Meta-analysis of risk factors for peptic ulcer. Nonsteroidal anti-inflammatory drugs, *Helicobacter pylori*, and smoking. *J Clin Gastroenterol* 1997;24:2.
72 Lin KJ, et al. Systematic review of peptic ulcer disease incidence rates: do studies without validation provide reliable estimates? *Pharmacoepidemiol Drug Saf* 2011;20:718.
73 Sobieraj DM, et al. US prevalence of upper gastrointestinal symptoms: a systematic literature review. *Am J Manag Care* 2011;17:e449.
74 Sonnenberg A, et al. The prevalence of self-reported peptic ulcer in the United States. *Am J Public Health* 1996;86:200.
75 Garrow D, et al. Risk factors for gastrointestinal ulcer disease in the US population. *Dig Dis Sci* 2010; 55:66.
76 Munnangi S, et al. Time trends of physician visits and treatment patterns of peptic ulcer disease in the United States. *Arch Intern Med* 1997;157:1489.
77 Feinstein LB, et al. Trends in hospitalizations for peptic ulcer disease, United States, 1998–2005. *Emerg Infect Dis* 2010;16:1410.
78 Wang YR, et al. Trends and outcomes of hospitalizations for peptic ulcer disease in the United States, 1993 to 2006. *Ann Surg* 2010;251:51.
79 Lanas A, et al. Time trends and impact of upper and lower gastrointestinal bleeding and perforation in clinical practice. *Am J Gastroenterol* 2009;104:1633.
80 Ahsberg K, et al. Hospitalisation of and mortality from bleeding peptic ulcer in Sweden: a nationwide time-trend analysis. *Aliment Pharmacol Ther* 2011;33: 578.
81 Barkun A, et al. Systematic review of the symptom burden, quality of life impairment and costs associated with peptic ulcer disease. *Am J Med* 2010;123:358.
82 Sandler RS, et al. The burden of selected digestive diseases in the United States. *Gastroenterology* 2002;122:1500.
83 Jonsson B, et al. Economic burden of NSAID-induced gastropathy in Sweden. *Scand J Gastroenterol* 2001;36:775.
84 Sonnenberg A, et al. Health impact of peptic ulcer in the United States. *Am J Gastroenterol* 1997;92:614.
85 Lam SK. Differences in peptic ulcer between East and West. *Baillieres Best Pract Res Clin Gastroenterol* 2000;14:41.
86 Huang JQ, et al. Role of *Helicobacter pylori* infection and non-steroidal anti-inflammatory drugs in peptic-ulcer disease: a meta-analysis. *Lancet* 2002;359: 14.
87 Howden CW, et al. Current indications for acid suppressants in *Helicobacter pylori*-negative ulcer disease. *Best Pract Res Clin Gastroenterol* 2001;15:401.
88 McColl KE. How I manage *H. pylori*-negative, NSAID/aspirin-negative peptic ulcers. *Am J Gastroenterol* 2009;104:190.
89 Gisbert JP, et al. *Helicobacter pylori* infection and perforated peptic ulcer prevalence of the infection and role of antimicrobial treatment. *Helicobacter* 2003; 8:159.

90 Gisbert JP, et al. *Helicobacter pylori* and bleeding peptic ulcer: what is the prevalence of the infection in patients with this complication? *Scand J Gastroenterol* 2003;38:2.
91 Vaira D, et al. What is the role of *Helicobacter pylori* in complicated ulcer disease? *Gastroenterology* 1997;113:S78.
92 Sanchez-Delgado J, et al. Has *H. pylori* prevalence in bleeding peptic ulcer been underestimated? A meta-regression. *Am J Gastroenterol* 2011;106:398.
93 Sonnenberg A, et al. Associations of peptic ulcer and gastric cancer with other diseases in US veterans. *Am J Public Health* 1995;85:1252.
94 Kurata JH, et al. A prospective study of risk for peptic ulcer disease in Seventh-Day Adventists. *Gastroenterology* 1992;102:902.
95 Kato I, et al. A prospective study of gastric and duodenal ulcer and its relation to smoking, alcohol, and diet. *Am J Epidemiol* 1992;135:521.
96 Watanabe Y, et al. Epidemiological study of peptic ulcer disease among Japanese and Koreans in Japan. *J Clin Gastroenterol* 1992;15:68.
97 Adami HO, et al. Is duodenal ulcer really a psychosomatic disease? A population-based case-control study. *Scand J Gastroenterol* 1987;22:889.
98 Katschinski BD, et al. Duodenal ulcer and refined carbohydrate intake: a case-control study assessing dietary fibre and refined sugar intake. *Gut* 1990;31:993.
99 Elmstahl S, et al. Fermented milk products are associated to ulcer disease. Results from a cross-sectional population study. *Eur J Clin Nutr* 1998;52:668.
100 Raiha I, et al. Lifestyle, stress, and genes in peptic ulcer disease: a nationwide twin cohort study. *Arch Intern Med* 1998;158:698.
101 Medalie JH, et al. The importance of biopsychosocial factors in the development of duodenal ulcer in a cohort of middle-aged men. *Am J Epidemiol* 1992;136:1280.
102 Tuchsen F, et al. Employment status, non-daytime work and gastric ulcer in men. *Int J Epidemiol* 1994;23:365.
103 Hein HO, et al. Genetic markers for peptic ulcer. A study of 3387 men aged 54 to 74 years: the Copenhagen Male Study. *Scand J Gastroenterol* 1997;32:16.
104 Sipponen P, et al. Chronic antral gastritis, Lewis(a+) phenotype, and male sex as factors in predicting coexisting duodenal ulcer. *Scand J Gastroenterol* 1989;24:581.
105 Ohyauchi M, et al. The polymorphism interleukin 8 -251 A/T influences the susceptibility of *Helicobacter pylori*-related gastric diseases in the Japanese population. *Gut* 2005;54:330.
106 Gyulai Z, et al. Genetic polymorphism of interleukin-8 (IL-8) is associated with *Helicobacter pylori*-induced duodenal ulcer. *Eur Cytokine Netw* 2004;15:353.
107 Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* 1965;64:31.
108 Hansson LE, et al. Survival in stomach cancer is improving: results of a nationwide population-based Swedish study. *Ann Surg* 1999;230:162.
109 American Cancer Society. *Cancer Facts & Figures 2011*, American Cancer Society, Atlanta. http://www.cancer.org/Research/CancerFactsFigures/CancerFactsFigures/cancer-facts-figures-2011 (accessed December 30, 2011).
110 Hansson LE, et al. The decline in the incidence of stomach cancer in Sweden 1960–1984: a birth cohort phenomenon. *Int J Cancer* 1991;47:499.
111 Aragones N, et al. Trends in oesophago-gastric cancer incidence in Spain: analysis by subsite and histology. *Ann Oncol* 2010;21(Suppl 3):iii69.
112 Jemal A, et al. Global cancer statistics. *CA Cancer J Clin* 2011;61:69.
113 Kamineni A, et al. The incidence of gastric carcinoma in Asian migrants to the United States and their descendants. *Cancer Causes Control* 1999;10:77.
114 Cho NH, et al. Ethnic variation in the incidence of stomach cancer in Illinois, 1986–1988. *Am J Epidemiol* 1996;144:661.
115 Peng LH, et al. Singapore Cancer Registry Interim Report : Trends in cancer incidence in Singapore 2002–2006. National Registry of Diseases Office (NRDO), 2007. www.hpb.gov.sg/uploadedFiles/HPB.../Cancer Trends2002-2006.pdf (accessed December 30, 2011).
116 Devesa SS, et al. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. *Cancer* 1998;83:2049.
117 Camargo MC, et al. Divergent trends for gastric cancer incidence by anatomical subsite in US adults. *Gut* 2011;60:1644.
118 Ekstrom AM, et al. Evaluating gastric cancer misclassification: a potential explanation for the rise in cardia cancer incidence. *J Natl Cancer Inst* 1999;91:786.
119 Ekstrom AM, et al. Decreasing incidence of both major histologic subtypes of gastric adenocarcinoma – a population-based study in Sweden. *Br J Cancer* 2000;83:391.
120 Eslick GD. *Helicobacter pylori* infection causes gastric cancer? A review of the epidemiological, meta-analytic, and experimental evidence. *World J Gastroenterol* 2006;12:2991.
121 Maeda S, et al. Assessment of gastric carcinoma risk associated with *Helicobacter pylori* may vary depending

on the antigen used: CagA specific enzyme-linked immunoadsorbent assay (ELISA) versus commercially available *H. pylori* ELISAs. *Cancer* 2000;88:1530.

122 Ekstrom AM, et al. *Helicobacter pylori* in gastric cancer established by CagA immunoblot as a marker of past infection. *Gastroenterology* 2001;121:784.

123 Hansen S, et al. *Helicobacter pylori* infection and risk of cardia cancer and non-cardia gastric cancer. A nested case-control study. *Scand J Gastroenterol* 1999;34: 353.

124 Kamangar F, et al. Opposing risks of gastric cardia and noncardia gastric adenocarcinomas associated with *Helicobacter pylori* seropositivity. *J Natl Cancer Inst* 2006;98:1445.

125 Uemura N, et al. *Helicobacter pylori* infection and the development of gastric cancer. *New Engl J Med* 2001;345:784.

126 Brenner H, et al. Is *Helicobacter pylori* infection a necessary condition for noncardia gastric cancer? *Am J Epidemiol* 2004;159:252.

127 Engel LS, et al. Population attributable risks of esophageal and gastric cancers. *J Natl Cancer Inst* 2003;95:1404.

128 Hatakeyama M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nat Rev Cancer* 2004;4:688.

129 Held M, et al. Is the association between *Helicobacter pylori* and gastric cancer confined to CagA-positive strains? *Helicobacter* 2004;9:271.

130 Figueiredo C, et al. *Helicobacter pylori* and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. *J Natl Cancer Inst* 2002;94:1680.

131 Correa P, et al. Chemoprevention of gastric dysplasia: randomized trial of antioxidant supplements and anti-*Helicobacter pylori* therapy. *J Natl Cancer Inst* 2000;92:1881.

132 Sung JJ, et al. Atrophy and intestinal metaplasia one year after cure of *H. pylori* infection: a prospective, randomized study. *Gastroenterology* 2000;119:7.

133 Wong BC, et al. *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. *JAMA* 2004;291:187.

134 Leung WK, et al. Factors predicting progression of gastric intestinal metaplasia: results of a randomised trial on *Helicobacter pylori* eradication. *Gut* 2004;53: 1244.

135 Mera R, et al. Long term follow up of patients treated for *Helicobacter pylori* infection. *Gut* 2005;54:1536.

136 You WC, et al. Randomized double-blind factorial trial of three treatments to reduce the prevalence of precancerous gastric lesions. *J Natl Cancer Inst* 2006;98: 974.

137 Fukase K, et al. Effect of eradication of *Helicobacter pylori* on incidence of metachronous gastric carcinoma after endoscopic resection of early gastric cancer: an open-label, randomised controlled trial. *Lancet* 2008;372:392.

138 Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum* 2004;83:1.

139 Tredaniel J, et al. Tobacco smoking and gastric cancer: review and meta-analysis. *Int J Cancer* 1997;72: 565.

140 Gonzalez CA, et al. Smoking and the risk of gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Int J Cancer* 2003;107: 629.

141 Koizumi Y, et al. Cigarette smoking and the risk of gastric cancer: a pooled analysis of two prospective studies in Japan. *Int J Cancer* 2004;112:1049.

142 Nomura AM, et al. The association of cigarette smoking with gastric cancer: the multiethnic cohort study. *Cancer Causes Control* 2012;23:51.

143 Chow WH, et al. Risk of stomach cancer in relation to consumption of cigarettes, alcohol, tea and coffee in Warsaw, Poland. *Int J Cancer* 1999;81:871.

144 Lagergren J, et al. The role of tobacco, snuff and alcohol use in the aetiology of cancer of the oesophagus and gastric cardia. *Int J Cancer* 2000;85:340.

145 Mao Y, et al. Active and passive smoking and the risk of stomach cancer, by subsite, in Canada. *Eur J Cancer Prev* 2002;11:27.

146 Chao A, et al. Cigarette smoking, use of other tobacco products and stomach cancer mortality in US adults: The Cancer Prevention Study II. *Int J Cancer* 2002;101:380.

147 World Cancer Research Fund/American Institute of Cancer Research (1997) *Food, Nutrition and the Prevention of Cancer: A Global Perspective*, BANTA Book Group, Menasha, USA.

148 Larsson SC, et al. Alcoholic beverage consumption and gastric cancer risk: a prospective population-based study in women. *Int J Cancer* 2007;120:373.

149 Duell EJ, et al. Alcohol consumption and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Am J Clin Nutr* 2011;94:1266.

150 Tramacere I, et al. A meta-analysis on alcohol drinking and gastric cancer risk. *Ann Oncol* 2012;23:28.

151 Riboli E, et al. Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. *Am J Clin Nutr* 2003;78:559S.

152 Steevens J, et al. Vegetables and fruits consumption and risk of esophageal and gastric cancer subtypes in the Netherlands Cohort Study. *Int J Cancer* 2011;129:2681.

153 Botterweck AA, et al. Vitamins, carotenoids, dietary fiber, and the risk of gastric carcinoma: results from a prospective study after 6.3 years of follow-up. *Cancer* 2000;88:737.
154 Nouraie M, et al. Fruits, vegetables, and antioxidants and risk of gastric cancer among male smokers. *Cancer Epidemiol Biomarkers Prev* 2005;14:2087.
155 Jenab M, et al. Plasma and dietary vitamin C levels and risk of gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *Carcinogenesis* 2006;27:2250.
156 Zheng W, et al. Retinol, antioxidant vitamins, and cancers of the upper digestive tract in a prospective cohort study of postmenopausal women. *Am J Epidemiol* 1995;142:955.
157 Larsson SC. (2006) Diet and gastrointestinal cancer. One carbon metabolism and other aspects. Thesis, Karolinska Institutet, Stockholm.
158 Stahelin HB, et al. Plasma antioxidant vitamins and subsequent cancer mortality in the 12-year follow-up of the prospective Basel Study. *Am J Epidemiol* 1991; 133:766.
159 Yuan JM, et al. Prediagnostic levels of serum micronutrients in relation to risk of gastric cancer in Shanghai, China. *Cancer Epidemiol Biomarkers Prev* 2004;13:1772.
160 Jenab M, et al. Plasma and dietary carotenoid, retinol and tocopherol levels and the risk of gastric adenocarcinomas in the European prospective investigation into cancer and nutrition. *Br J Cancer* 2006;95:406.
161 Taylor PR, et al. Prospective study of serum vitamin E levels and esophageal and gastric cancers. *J Natl Cancer Inst* 2003;95:1414.
162 Blot WJ, et al. Nutrition intervention trials in Linxian, China: supplementation with specific vitamin/mineral combinations, cancer incidence, and disease-specific mortality in the general population. *J Natl Cancer Inst* 1993;85:1483.
163 Malila N, et al. Effects of alpha-tocopherol and beta-carotene supplementation on gastric cancer incidence in male smokers (ATBC Study, Finland). *Cancer Causes Control* 2002;13:617.
164 Virtamo J, et al. Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation: a postintervention follow-up. *JAMA* 2003;290:476.
165 Hennekens CH, et al. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *New Engl J Med* 1996;334:1145.
166 Li JY, et al. Nutrition intervention trials in Linxian, China: multiple vitamin/mineral supplementation, cancer incidence, and disease-specific mortality among adults with esophageal dysplasia. *J Natl Cancer Inst* 1993;85:1492.
167 Dawsey SM, et al. Effects of vitamin/mineral supplementation on the prevalence of histological dysplasia and early cancer of the esophagus and stomach: results from the Dysplasia Trial in Linxian, China. *Cancer Epidemiol Biomarkers Prev* 1994;3:167.
168 Terry P, et al. Inverse association between intake of cereal fiber and risk of gastric cardia cancer. *Gastroenterology* 2001;120:387.
169 McCullough ML, et al. A prospective study of diet and stomach cancer mortality in United States men and women. *Cancer Epidemiol Biomarkers Prev* 2001;10:1201.
170 Kono S, et al. Nutrition and stomach cancer. *Cancer Causes Control* 1996;7:41.
171 Joossens JV, et al. Dietary salt, nitrate and stomach cancer mortality in 24 countries. European Cancer Prevention (ECP) and the INTERSALT Cooperative Research Group. *Int J Epidemiol* 1996;25:494.
172 Tsugane S. Salt, salted food intake, and risk of gastric cancer: epidemiologic evidence. *Cancer Sci* 2005; 96:1.
173 Cohen AJ, et al. Evaluation of the aetiological role of dietary salt exposure in gastric and other cancers in humans. *Food Chem Toxicol* 1997;35:271.
174 Jakszyn P, et al. Endogenous versus exogenous exposure to *N*-nitroso compounds and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST) study. *Carcinogenesis* 2006;27:1497.
175 Larsson SC, et al. Processed meat consumption and stomach cancer risk: a meta-analysis. *J Natl Cancer Inst* 2006;98:1078.
176 Gonzalez CA, et al. Meat intake and risk of stomach and esophageal adenocarcinoma within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J Natl Cancer Inst* 2006;98:345.
177 Tian W, et al. Meta-analysis on the relationship between nonsteroidal anti-inflammatory drug use and gastric cancer. *Eur J Cancer Prev* 2010;19:288.
178 Cook NR, et al. Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. *JAMA* 2005;294:47.
179 Lichtenstein P, et al. Environmental and heritable factors in the causation of cancer – analyses of cohorts of twins from Sweden, Denmark, and Finland. *New Engl J Med* 2000;343:78.
180 Caldas C, et al. Familial gastric cancer: overview and guidelines for management. *J Med Genet* 1999;36: 873.
181 Oliveira C, et al. Genetics, pathology, and clinics of familial gastric cancer. *Int J Surg Pathol* 2006;14:21.

182 El-Omar EM, et al. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* 2000;404:398.
183 Gonzalez CA, et al. Genetic susceptibility and gastric cancer risk. *Int J Cancer* 2002;100:249.
184 Larsson SC, et al. Folate intake, MTHFR polymorphisms, and risk of esophageal, gastric, and pancreatic cancer: a meta-analysis. *Gastroenterology* 2006;131:1271.

Answers to multiple choice questions

1. C
2. E
3. A

15 Epidemiology of dyspepsia

Alexander C. Ford[1] *& Nicholas J. Talley*[2]
[1]Leeds Teaching Hospitals Trust, Leeds, West Yorkshire, UK
[2]University of Newcastle, Callaghan, NSW, Australia

Key points

- Rome III criteria, divides functional dyspepsia into two separate syndromes: postprandial distress; and epigastric pain.
- Between 5–40 % of the population suffer with dyspepsia.
- Dyspepsia is associated with a reduction in quality of life and those with dyspepsia take on average almost 1 extra sick day per year compared with those without epigastric pain or discomfort.
- Clinicians are unable to reliably distinguish organic from functional causes of dyspepsia.
- The majority of patients with dyspepsia have a normal endoscopy so not all those with upper gastrointestinal symptoms need invasive investigations, unless they are over 50–55 years of age or have alarm features.

Disease definitions

Dyspepsia is a hybrid word, derived from Latin and Greek, meaning bad (*dys*) digestion (*pepsis*), and is a complex of symptoms referable to the upper gastrointestinal (GI) tract. Over the last 20 years definitions of the condition have been refined substantially, perhaps because not all clinicians and researchers agree which symptoms should be included in the syndrome of dyspepsia. The situation is further complicated by the fact that the classification of dyspepsia depends upon whether upper GI endoscopy has been performed and, if so, whether there were structural abnormalities. Individuals who have not undergone investigation are said to have uninvestigated dyspepsia, dyspeptic patients who undergo upper GI investigation and have findings that may be responsible for the symptoms, such as peptic ulcer, are classed as having organic dyspepsia, while those without a detectable cause are labeled as having functional dyspepsia.

In the late 1980s a report from a working party defined dyspepsia broadly according to the presence of any symptom thought to be referable to the upper GI tract [1]. However, in the early 1990s the Rome Foundation was instrumental in changing approaches to the classification of functional GI disorders, and at this time the definition was restricted to a feeling of pain or discomfort centered in the upper abdomen [2], with symptoms that were suggestive of heartburn excluded, as these were felt to be pathognomonic for gastroesophageal reflux disease (GERD), which was classified separately. The exclusion of these symptoms from the definition of dyspepsia was intended primarily for research purposes, in order to facilitate the recruitment of individuals with homogeneous symptoms into clinical trials of therapies for these conditions in secondary or tertiary care, as well as to enable the conduct of studies whose aims were to elucidate

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

potential underlying pathophysiologies. As a result, this approach has been criticized by some because this more restrictive definition of dyspepsia may not be generalizable to primary care, where individuals often present with numerous overlapping symptoms [3], which are not necessarily predictive of underlying pathology [4]. In fact, there is recent evidence to suggest that even when individuals with Rome-defined dyspepsia do undergo upper GI endoscopy, the commonest organic finding encountered is erosive esophagitis [5].

Despite these concerns, the Rome criteria have been revised on two subsequent occasions [6,7], and are considered to be the gold standard for defining the presence of dyspepsia. The latest of these revisions, the Rome III criteria [6], divides functional dyspepsia into two separate syndromes: postprandial distress; and epigastric pain. The former requires the presence of either bothersome postprandial fullness after normal-sized meals, or early satiation that prevents eating a regular meal. The latter consists of pain or burning localized to the epigastric region of at least moderate severity. In both cases symptoms need to have been present for at least 3 months, with onset at least 6 months prior to diagnosis, and there should be no evidence of structural abnormalities at upper GI endoscopy. These subgroups were created based on the results of symptom clusters reported in factor analysis studies [8–11]. However, there is evidence to suggest substantial overlap between these two syndromes [12,13], as well as between dyspepsia and other functional GI disorders, such as irritable bowel syndrome (IBS) [14].

Incidence and prevalence

Dyspepsia is common in the general population, yet the incidence of dyspepsia has not been widely reported to date. A Swedish study that followed up more than 1000 individuals from the community at 3 months, 1 year, and 7 years demonstrated incidences of new-onset dyspepsia of 0.8 %, 1 %, and 3 %, respectively [15,16]. In a UK-based study that followed almost 4000 individuals over 10 years, the incidence of new-onset dyspepsia in those asymptomatic at baseline was higher, at almost 3 % per year [17]. Finally, in a follow-up of residents of Olmsted County, Minnesota the onset rate for dyspepsia during the 12-year time frame of the study was close to 5 % [18]. As the onset of dyspeptic symptoms in previously asymptomatic individuals in all these studies was closely matched by the rates of symptom resolution in those who were symptomatic at initial study entry, the prevalence of dyspepsia remains remarkably stable over extended periods of follow-up.

There have been many cross-sectional surveys that have reported the prevalence of dyspepsia in various groups of individuals worldwide. A systematic review and meta-analysis assembled all studies published up to 2008 in order to estimate the worldwide prevalence of dyspepsia, identifying 157 eligible articles [19]. Prevalence varied from less than 5 % to greater than 40 % [17,20–22], depending on the criteria used to define the presence of dyspepsia, as well as the characteristics of the population under study. The majority of identified studies were conducted in North America and Europe, with a dearth of data concerning the prevalence of dyspepsia in Central America, Africa, South Asia, and the Middle East [23–27]. When data from all studies were pooled, the worldwide prevalence was 18 % using the Rome II criteria to define dyspepsia, compared with 31 % when a broad definition was used [19].

The prevalence according to geographical location of the studies identified in this meta-analysis, when either a broad definition of dyspepsia or the Rome II criteria were used, is summarized in Figure 15.1 and Figure 15.2. The striking variation in prevalence throughout the world, even when the same diagnostic criteria are used to define dyspepsia, highlights the importance of other factors such as genetic, ethnic, and cultural differences on the reporting of upper GI symptoms.

One major limitation of the surveys to date is a lack of detailed information on the relationship between meals and upper GI symptoms; indeed, surveys may not capture this relationship accurately and diaries may be required. Experimental data suggest that a key characteristic of functional dyspepsia is meal-induced symptoms, when measured after a test meal, even when this was not recognized by patients completing a baseline questionnaire [28].

Risk factors

Proposed risk factors for dyspepsia include *Helicobacter pylori* (*H. pylori*) infection, aspirin and

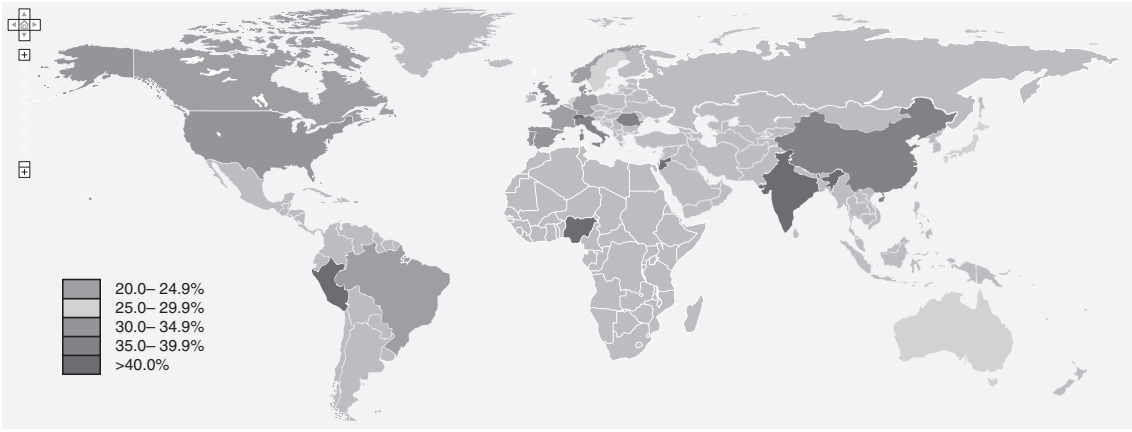


Figure 15.1 Prevalence of dyspepsia worldwide using a broad definition.

nonsteroidal anti-inflammatory drug (NSAID) use, tobacco and alcohol consumption, female gender, and lower socioeconomic status. Numerous cross-sectional surveys have examined the influence of these, and other, sociodemographic variables [29–39]. A large population-based study in the United Kingdom suggested that 5 % of dyspepsia in the general population is attributable to *H. pylori* [31], but others have not replicated these data [33,34]. In this study [31], as well as another UK-based study [32], analgesic drug use was significantly higher in those reporting dyspepsia, whilst Talley et al. demonstrated an almost twofold increase in odds for aspirin use among those with dyspepsia [37]. Tobacco use, but not alcohol, was also associated with dyspepsia in this study [37], but another study conducted among German blood donors did not demonstrate any association between either tobacco or alcohol use and dyspepsia [40]. Finally, a survey of 15,000 Australian adults suggested that dyspepsia was commoner among those of lower socioeconomic status [41], but a UK-based study showed no clear association [22].

These conflicting results from individual studies emphasize that the interpretation of potential risk factors for dyspepsia in the population can be difficult. A systematic review and meta-analysis conducted

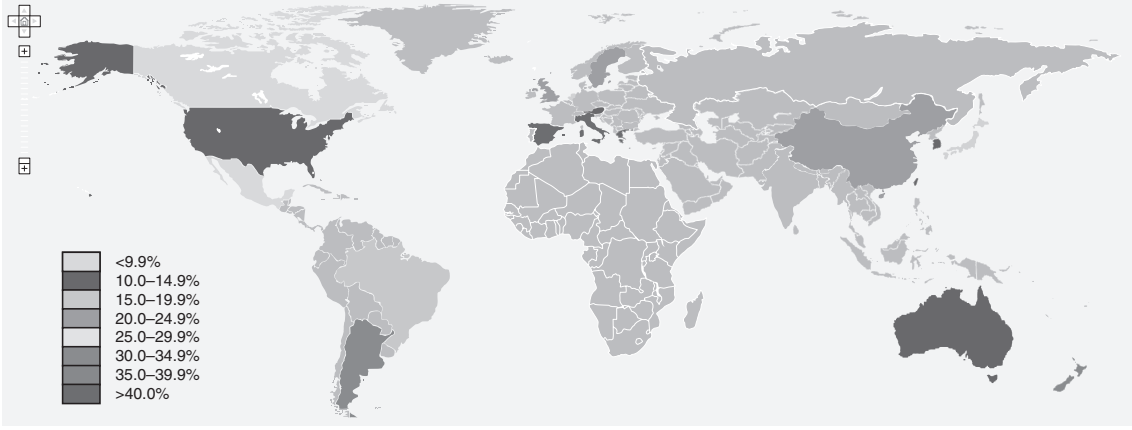


Figure 15.2 Prevalence of dyspepsia worldwide using the Rome II criteria.

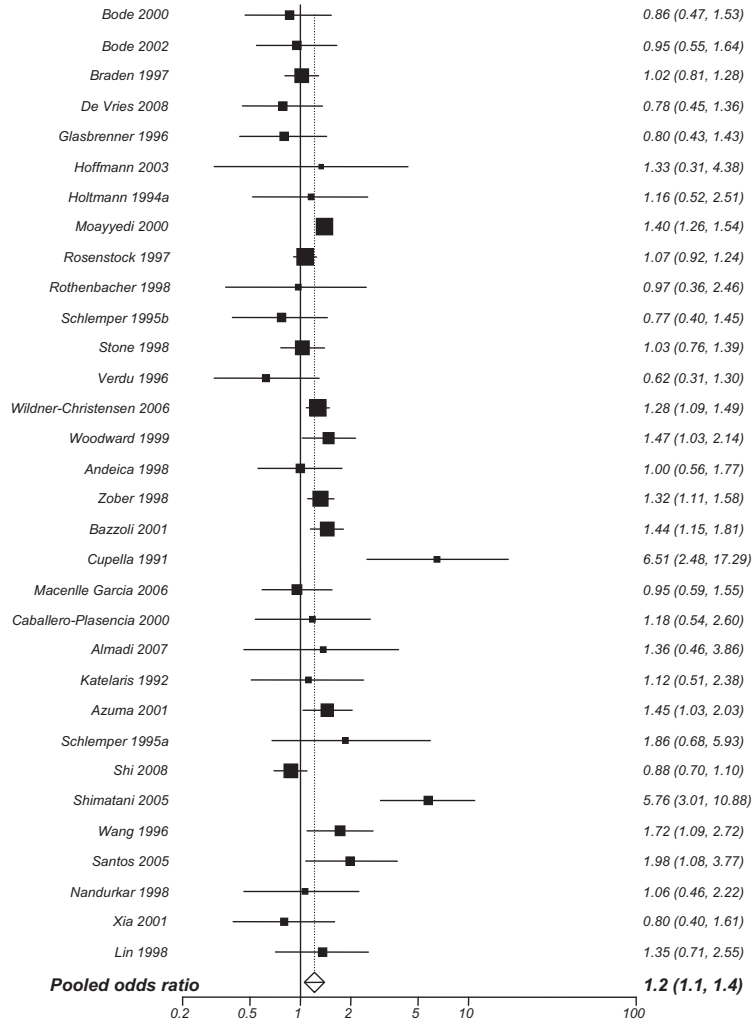


Figure 15.3 Forest plot of studies reporting on the effect of *Helicobacter pylori* status on prevalence of dyspepsia.

in 2009 attempted to synthesize all available data for some of these risk factors [42]. The authors identified 70 studies reporting on the effect of gender on dyspepsia, 32 that examined influence of *H. pylori* status, and 18 that studied NSAID or aspirin use. The pooled prevalence of dyspepsia in females was 26 %, compared with 24 % in males, giving an odds ratio for dyspepsia in females compared with males of 1.2 (95 % confidence interval (CI) 1.1–1.4). In *H. pylori*-positive individuals the prevalence of dyspepsia was 34 %, compared with 30 % in *H. pylori*-negative individuals, giving an odds ratio for dyspepsia in *H. pylori*-positive individuals of 1.2 (95 % CI 1.1–1.4) (Figure 15.3). Finally, the prevalence of dyspepsia in NSAID or aspirin users was 39 %, compared with 30 % in nonusers, with an odds ratio for dyspepsia of 1.5 (95 % CI 1.3–1.7) (Figure 15.4). Whilst this study confirms that female gender, presence of *H. pylori* infection, and NSAID or aspirin use are all statistically significantly associated with dyspepsia, causation cannot be implied from observational studies such as these. In addition, these data highlight how modest the absolute increase in the prevalence of dyspepsia is when these risk factors are present, suggesting there are other, perhaps more important, underlying determinants of symptom status.

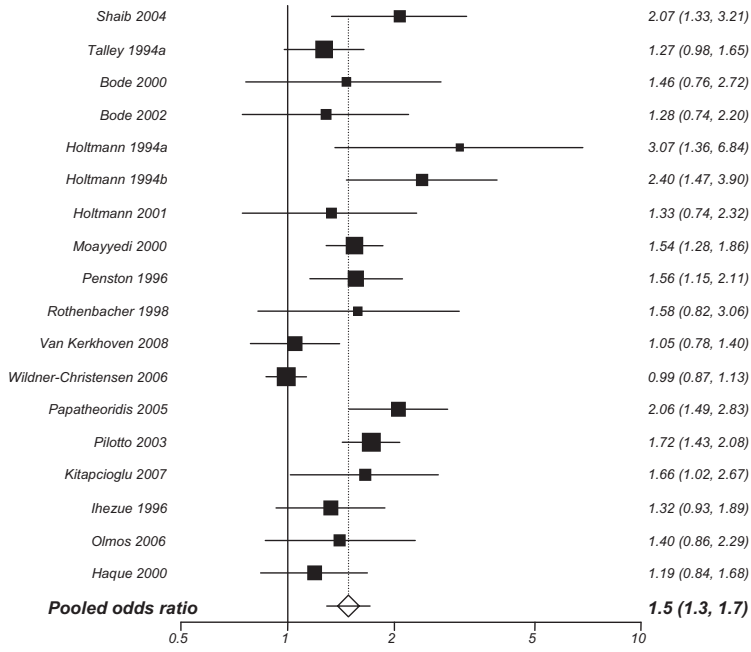


Figure 15.4 Forest plot of studies reporting on the effect of nonsteroidal anti-inflammatory drug or aspirin use on prevalence of dyspepsia.

There is good evidence to support the hypothesis that acute bacterial or viral infection can trigger chronic lower GI symptoms compatible with IBS in some individuals [43–45], a phenomenon termed post-infectious IBS. In a series of dyspeptic patients undergoing investigation, 17 % reported an acute onset of symptoms accompanied by other symptoms suggestive of an infective process, such as fever, myalgia, diarrhea, or vomiting [46]. These individuals were labeled as having presumed post-infectious dyspepsia. Interest in this phenomenon led other investigators to study the association between acute enteric illness and dyspepsia [47–50]. In a follow-up of a cohort of over 300 individuals admitted to a single hospital with infective diarrhea, 12 % developed symptoms compatible with a functional GI disorder at 5 years, and 12 % of these individuals met diagnostic criteria for dyspepsia [49].A case-control study reported that the odds ratio for dyspepsia in cases with bacterial gastroenteritis, confirmed by a positive stool culture, compared with nonexposed individuals was 2.9 at 6 months post-infection [48], although this did not reach statistical significance, perhaps because of a small sample size. In a 1-year follow-up of a Spanish community, some of whom were exposed to an outbreak of *Salmonella*, the odds of dyspepsia in infected individuals at 12 months was 6 [47].

The largest study to examine this association, to date, was conducted in over 1000 individuals from the town of Walkerton, Ontario in Canada. Here, the municipal water supply was contaminated by livestock fecal residue, leading to an outbreak of bacterial dysentery in May 2000. When individuals were followed up 8 years post-infection, the prevalence of dyspepsia among exposed individuals was 49 %, compared with 30 % in nonexposed individuals, giving an odds ratio for dyspepsia in those exposed to dysentery of 2.3 (95 % CI 1.7–3.0) [50]. The biologic explanation for the increased prevalence of dyspepsia in individuals following acute gastroenteritis is unclear, but these studies suggest that acute enteric infections have the ability to trigger symptoms that affect the upper, as well as the lower, GI tract with long-lasting consequences. One hypothesis is the site of inflammation dictates the clinical outcome, because of increased permeability and recruitment of an inflammatory and cytokine response; if the distal small intestine and/or colon is inflamed, IBS can result, if only the proximal small intestine is involved, functional dyspepsia can occur, and if the entire small

intestine becomes inflamed then an overlap syndrome may follow [51]. However, this hypothesis needs to be tested rigorously.

Differential diagnosis

There are a myriad of diseases that lead to upper GI symptoms that may be compatible with dyspepsia. Some of these, such as biliary and pancreatic disorders, do not relate to the esophago-gastro-duodenal region at all. Some studies have suggested that there is an increased prevalence of celiac disease in individuals that report symptoms compatible with dyspepsia [52,53]. A recent systematic review and meta-analysis of case-control studies that examined this issue reported that the prevalence of positive celiac serology in individuals with symptoms suggestive of dyspepsia was 8 %, compared with 4 % in controls without [54]. The prevalence of biopsy-proven celiac disease was 3 % in those with dyspepsia, compared with 1 % in controls. Despite these differences in absolute prevalence, neither were statistically significant, with odds ratios of 1.9 (95 % CI 0.9–4.0) for positive celiac serology and 2.9 (95 % CI 0.6–13.4) for biopsy-proven celiac disease in individuals meeting criteria for dyspepsia.

In terms of diseases of the esophagus, stomach, and duodenum that may cause dyspepsia, there have been several cross-sectional surveys that have reported the prevalence of organic findings at upper GI endoscopy in individuals with dyspepsia, as well as in controls without dyspepsia [23,36,55–61]. A systematic review and meta-analysis of these studies was performed in 2010 [62]. Seven studies reported the prevalence of erosive esophagitis in 2067 individuals. This was the commonest finding, with a pooled prevalence of 13 %. Six studies reported a pooled prevalence of Barrett's esophagus in 1982 subjects of only 1 %. There was only one esophageal cancer reported among 1982 individuals with dyspepsia undergoing upper GI endoscopy, and four gastric cancers, giving a pooled prevalence for upper GI malignancy of 0.25 %. Finally, nine studies provided data for peptic ulcer in 2597 individuals with dyspepsia, with a pooled prevalence of 8 %. Six of these studies reported gastric and duodenal ulcer separately, with a pooled prevalence of 3 % for both.

When the prevalence of clinically significant endoscopic findings was compared between those with dyspepsia and those without, the prevalence of peptic ulcer was significantly higher in individuals with dyspepsia, with an odds ratio of 2.1 (95 % CI 1.5–2.8), although this was only the case for duodenal ulcer (odds ratio 3.1; 95 % CI 1.8–5.3) There were trends towards a higher prevalence of any clinically significant endoscopic finding, erosive esophagitis, and Barrett's esophagus in those with dyspepsia, but these did not reach statistical significance. When the effect of definition of dyspepsia used in the studies on prevalence of endoscopic findings was studied, the pooled prevalence of erosive esophagitis in individuals with dyspepsia was 20 % when a broad definition was used, compared with only 6 % when the Rome criteria were used, and the pooled prevalence of peptic ulcer was 6 % in studies using a broad definition of dyspepsia, compared with 11 % in studies that used the Rome criteria.

These data highlight that the majority of individuals with dyspepsia in the community have no structural abnormality to explain their symptoms at upper GI endoscopy, and that upper GI malignancy is present in less than 1 % of people. Even when organic pathology is found, in the majority of cases, this is no more likely to be present in those with dyspepsia compared with individuals without dyspepsia. In addition, despite attempts to classify dyspepsia and GERD separately, one of the commonest findings encountered at upper GI endoscopy performed for dyspepsia is erosive esophagitis, although the prevalence is lower when the Rome criteria are used to define dyspepsia.

Clinical diagnosis

From the above data, it is clear that if the physician could reliably distinguish between individuals with symptoms suggestive of organic pathology, and those who are likely to have functional dyspepsia, then the need for upper GI endoscopy could be obviated in more than 50 % of individuals with dyspepsia. Unfortunately, there is little evidence to suggest that gastroenterologists are capable of achieving this aim. A systematic review and meta-analysis assembled data from all studies examining this issue [63]. There were five articles that reported the accuracy of a

gastroenterologist in diagnosing organic dyspepsia. When data from these studies, containing over 3500 patients, were pooled the positive likelihood ratio for a gastroenterologist's diagnosis of organic dyspepsia being correct was 1.6, while the negative likelihood ratio was 0.4.

These likelihood ratios are extremely modest, but in studies that examined the performance of primary care physicians or computer models the likelihood ratios observed were almost identical. The positive likelihood ratios for gastroenterologists in predicting either peptic ulcer disease or erosive esophagitis were slightly better at around 3.0, but negative likelihood ratios were of similar magnitude to those for predicting an organic cause of dyspepsia. These data emphasize that clinicians perform only modestly in being able to distinguish between organic and functional causes of dyspepsia with any certainty.

Alarm features

When consulting with the dyspeptic patient it is usual for the doctor to elucidate symptoms and signs that may be indicative of underlying upper GI malignancy. These red-flag, or alarm, features include new onset dyspepsia in a patient aged over 50 years, dysphagia, hematemesis, melena, persistent vomiting, unintended weight loss, anemia, family history of gastric cancer, or a palpable epigastric mass. There is evidence to suggest that these features identify only those patients with advanced, and therefore often incurable, disease [64]. Some have therefore proposed that all individuals with dyspepsia should undergo upper GI endoscopy in order to detect early cancers that are amenable to surgical cure, but in a recent study that evaluated this issue in primary care, the cost per case of upper GI cancer detected was estimated at US$83,000 [5]. Despite the concerns that alarm features only identify advanced disease, all current national guidelines for the management of dyspepsia agree that patients with these features require urgent referral for upper GI endoscopy in order to exclude gastroesophageal cancer [65–67]. However, there is conflicting evidence that such items from the clinical history and physical examination are able to predict upper GI malignancy.

Two prospective studies carried out in open-access endoscopy departments in the Netherlands demonstrated that alarm features were present in the majority of individuals with upper GI malignancy [68,69]. After logistic regression, one study confirmed that they were significantly associated with the presence of upper GI malignancy at endoscopy [69]. In a rapid access upper GI cancer service in the United Kingdom, referral criteria for urgent investigation of dyspepsia were examined prospectively in almost 2000 consecutive patients [70], and Kapoor and co-authors established that dysphagia, weight loss, and age over 55 years were all significant positive predictors of malignancy, but uncomplicated dyspepsia in those over 55 years was actually a significant negative predictor. A prediction model was constructed, using these criteria, which was applied to a similar number of subsequent consecutive referrals. Use of these criteria would have reduced referrals by almost one-third, and sensitivity of the model in detecting upper GI cancer was over 90 %. In contrast to these three reports, one multicenter study conducted in the United States, of similar design to that of Kapoor et al., failed to demonstrate a significant association between classical alarm features and upper GI malignancy [71]. Evidence of anemia and bleeding were predictors but, when combined with age in a prediction model, the sensitivity and specificity were both poor.

One problem with reliance on the presence of alarm symptoms is that, because upper GI malignancy is rare, their positive predictive value is actually very low, estimated at between 3 % and 10 % in these studies, meaning that large numbers of individuals would still need to undergo upper GI endoscopy to detect a small number of cancers. A systematic review and meta-analysis published in 2006 identified 15 prospective studies, which evaluated the accuracy of alarm features in the diagnosis of upper GI malignancy in dyspeptic patients [72]. The presence of one or more alarm features had a pooled sensitivity and specificity of 67 % and 66 %, respectively. When the individual alarm features of weight loss, anemia, and dysphagia were examined, pooled sensitivities varied between 13 % and 49 %, whilst specificity ranged from 84 % to 95 %.

Systematically analyzing all available data therefore suggests that more accurate ways of predicting a diagnosis of gastroesophageal malignancy are required. In a study published since this meta-analysis was performed, which was conducted in a Chinese population with a high prevalence of *H. pylori* who were therefore at higher risk of upper GI malignancy, alarm

features still had limited predictive value [73]. Further prospective studies are required to examine the accuracy of combinations of alarm features, or identify those with a very high specificity which, if present, can be used to rule in a diagnosis of upper GI malignancy.

Natural history and mortality

Few studies have described the natural history of dyspepsia. One Swedish study randomly selected over 1000 individuals, and sent them GI symptom questionnaires at baseline, and then again at 1 year, and 7 years [15,16]. The authors demonstrated that the overall prevalence of the condition remained stable with time, but that about 20 % of symptomatic individuals' dyspepsia spontaneously resolved, whilst a similar proportion of asymptomatic people developed new onset of symptoms. Those with mild symptoms, who had not needed to consult their GP, were more likely to experience resolution of symptoms. When those with dyspepsia at baseline were subcategorized into ulcer-like, reflux-like, and dysmotility-like dyspepsia, less than 50 % who were still symptomatic at 1-year follow-up remained in the same symptom subcategory. When the cohort of individuals were followed-up at 7 years, the authors found that most asymptomatic subjects remained symptom-free, those with reflux continued with reflux-type symptoms, whilst those with dyspepsia and IBS moved between both of these symptom groups.

More recently, two studies have been published with an even longer duration of follow-up [18,74]. In two surveys of almost 4000 individuals in the UK community, conducted 10 years apart, one-third of individuals with dyspepsia at study entry experienced symptoms resolution at 10 years, whilst one-quarter still met criteria for dyspepsia [74]. The remaining individuals experienced a flux of their symptoms such that they went on to meet criteria for either IBS or GERD. In a study conducted in the United States, with 12 years of follow-up, 50 % of patients with symptoms of dyspepsia at baseline were asymptomatic at study end [18].

In terms of the effect of dyspepsia on survival, there are few published studies that have examined this issue. In a recent population-based cohort study, with over 30,000 years of follow-up, that reported data from almost 4000 individuals, no association between a diagnosis of dyspepsia at baseline and subsequent survival was demonstrated (hazard ratio = 1.1; 95 % CI 0.6–2.0) after controlling for age, gender, tobacco use, comorbidity, and marital status [75].

Disability, quality of life, and healthcare seeking

Considering the impact of upper GI symptoms on activities of daily living, a large telephone survey of over 20,000 adults in the United States confirmed that upper GI symptoms were associated with significantly higher rates of absenteeism from work, missed leisure time, and reductions in activity around the house [76]. More recently, in a retrospective analysis of payroll data and health insurance claims, the impact of functional dyspepsia on costs and productivity was estimated [77]. Functional dyspepsia among employees was associated with significantly greater medical and prescription medicine costs per year, an additional 0.83 days of sickness absence per year, and led to a 12 % reduction in unit productivity per hour, when compared with employees without functional dyspepsia.

A systematic review performed in 2003 highlighted that, up to that point, there had been few studies of dyspepsia and health-related quality of life conducted among samples of the general population [78]. However, there have been several studies reporting a reduced quality of life in patients with functional dyspepsia attending secondary care compared to non-symptomatic individuals, healthy controls, or the general population [79–82], but it is not clear from these studies whether dyspepsia symptoms cause a reduced quality of life, or whether individuals with a poor quality of life develop dyspepsia symptoms. The negative impact of dyspepsia on quality of life is well recognized [79], although some of this association is thought to be due to other psychological factors [80], and indeed one large study suggested that psychiatric disorders and recent major life events were commoner in those who report the presence of dyspeptic symptoms [83].

It has always been assumed that symptoms of dyspepsia, and other chronic GI conditions, give rise to psychological distress, rather than the reverse. However, in a 10-year follow-up of individuals from the general population, one of the strongest predictors for the development of new-onset dyspepsia at

10 years was a poor quality of life at baseline [17]. There was an almost threefold increase in odds of dyspepsia amongst those with the lowest quality of life at baseline, with the data from this study suggesting that individuals in the lower half of the spectrum of quality of life had a 30 % population attributable risk for new-onset dyspepsia.

Previous studies have estimated that only one in four individuals with dyspepsia will consult a primary care physician as a result of their symptoms [22,84]. Numerous investigators have examined various sociodemographic variables that may influence the decision to consult a physician with dyspepsia. Factors demonstrated to predict consultation behavior include female gender, increasing age, coexistence of other functional GI disorders, concern or anxiety about underlying pathology, frequency, severity, and duration of symptoms, interference of symptoms with activities of daily living, and lower socioeconomic status [9,22,84–92], although the role of psychological factors remains disputed [87,88,91,93].

Prevention

Preventing dyspepsia is problematic as many of the proposed risk factors, such as gender and age, are not modifiable. Minimizing the use of aspirin or NSAIDs, or using prophylactic acid-suppression therapy when these drugs are required may reduce dyspepsia, via a reduction in the incidence of peptic ulcer disease [94–97]. As 5 % of dyspepsia in the general population may be attributable to *H. pylori* [31], screening and treatment programs for the bacterium in the community could be another means of reducing the burden of dyspepsia. There is evidence from large randomized placebo-controlled trials that this leads to a small, but potentially significant, reduction in dyspepsia rates [98,99], and one study demonstrated that the cost savings to the health service were sufficient to cover the initial costs of screening and treatment amongst the *H. pylori*-positive individuals enrolled [100].This did not take into account that there may also be savings from such a strategy, due to a reduction in the costs of managing gastric cancer [101,102].

Areas for further study

Future studies should concentrate on incidence and long-term natural history of the disorder, as well as the effects on life expectancy, as data examining these issues are sparse. We also require better ways of predicting who has organic versus functional dyspepsia, and who has upper GI malignancy, and this means that prospective studies that combine items from the clinical history, as well as physical examination, in order to increase accuracy, are needed. Finally, we need more studies reporting on interventions at a population level that may reduce the burden of dyspepsia in the community, and hence the health service-related costs that arise from managing such a common disorder.

Conclusions

Dyspepsia is common in the general population, with a prevalence of between 5 % and 40 %, depending on the criteria used to define its presence and the population under study. Risk factors are well characterized, though many of these are not modifiable, and their overall importance in the etiology of symptoms is questionable. Most individuals with dyspepsia do not have a structural cause for their symptoms when subjected to upper GI endoscopy, and upper GI malignancy is a rare cause of dyspepsia. However, physicians are not able to discriminate between functional and organic dyspepsia with any great accuracy, and alarm features have poor predictive value for diagnosing upper GI cancer. The natural history of the condition is chronic, with a relapsing and remitting nature. Some individuals become asymptomatic, whilst others experience an alteration of symptoms, such that they go on to meet criteria for another functional GI disorder. Despite its chronicity, there is no evidence, to date, that dyspepsia impacts adversely on life expectancy. Long-term prevention strategies for the disorder remain a hope, rather than a certainty.

Multiple choice questions

1 Which of the following statements concerning the prevalence of dyspepsia is correct:

A It does not vary according to the criteria used to define dyspepsia

B It does not vary according to geographical location

C It does not remain stable with time

D It varies between approximately 5 % and 40 % in cross-sectional surveys

E It is lower in magnitude than the incidence of dyspepsia

2 Proposed risk factors for dyspepsia include:
A Male gender
B *Helicobacter pylori* infection
C Higher socioeconomic status
D Avoiding aspirin use
E Alcohol use

3 Which of the following conditions does not appear to cause symptoms compatible with dyspepsia:
A Erosive esophagitis
B Gastric cancer
C Peptic ulcer disease
D Functional dyspepsia
E Celiac disease

References

1 Colin-Jones DG, Bloom B, Bodemar G, et al. Management of dyspepsia: Report of a working party. *Lancet* 1988;331:576–9.
2 Talley NJ, Colin-Jones DG, Koch KL, et al. Functional dyspepsia: A classification with guidelines for diagnosis and management. *Gastroenterol Int* 1991;4:145–60.
3 Tougas G, Chen MS, Hwang P, et al. Prevalence and impact of upper gastrointestinal symptoms in the Canadian population: Findings from the DIGEST study. *Am J Gastroenterol* 1999;94:2845–54.
4 Johnsen R, Bernersen B, Straume B, et al. Prevalence of endoscopic and histological findings in subjects with and without dyspepsia. *BMJ* 1991;302:749–52.
5 Vakil N, Talley NJ, Veldhuyzen van Zanten S, et al. Cost of detecting malignant lesions by endoscopy in 2741 primary care dyspeptic patients without alarm symptoms. *Clin Gastroenterol Hepatol* 2009;7:756–61.
6 Tack J, Talley NJ, Camilleri M, et al. Functional gastroduodenal disorders. *Gastroenterology* 2006;130:1466–79.
7 Talley NJ, Stanghellini V, Heading RC, et al. Functional gastroduodenal disorders. *Gut* 1999;45(Suppl 2):37–42.
8 Stanghellini V, Tosetti C, Paternico A, et al. Predominant symptoms identify different subgroups in functional dyspepsia. *Am J Gastroenterol* 1999;94:2080–5.
9 Talley NJ, Zinsmeister AR, Schleck CD, et al. Dyspepsia and dyspepsia subgroups: A population-based study. *Gastroenterology* 1992;102:1259–68.
10 Talley NJ, Boyce P, Jones M. Identification of distinct upper and lower gastrointestinal symptom groupings in an urban population. *Gut* 1998;42:690–5.
11 Talley NJ, Holtmann G, Agreus L, et al. Gastrointestinal symptoms and subjects cluster into distinct upper and lower groupings in the community: a four nations study. *Am J Gastroenterol* 2000;95(6):1439–47.
12 Abid S, Siddiqui S, Jafri W. Discriminant value of Rome III questionnaire in dyspeptic patients. *Saudi J Gastroenterol* 2011;17:129–33.
13 Ford AC, Morgan D, Moayyedi P. Rome III criteria for functional gastrointestinal disorders: Too much overlap to be useful? *Gastroenterology* 2011;140(Suppl 1):S725–6.
14 Ford AC, Marwaha A, Lim A, et al. Systematic review and meta-analysis of the prevalence of irritable bowel syndrome in individuals with dyspepsia. *Clin Gastroenterol Hepatol* 2010;8:401–9.
15 Agreus L, Svardsudd K, Nyren O, et al. Irritable bowel syndrome and dyspepsia in the general population: Overlap and lack of stability over time. *Gastroenterology* 1995;109:671–80.
16 Agreus L, Svardsudd K, Talley NJ, et al. Natural history of gastroesophageal reflux disease and functional abdominal disorders. *Am J Gastroenterol* 2001;96:2905–14.
17 Ford AC, Forman D, Bailey AG, et al. Initial poor quality of life and new onset of dyspepsia: Results from a longitudinal 10-year follow-up study. *Gut* 2007;56:321–7.
18 Halder SLS, LockeIII GR, Schleck CD, et al. Natural history of functional gastrointestinal disorders: A 12-year longitudinal population-based study. *Gastroenterology* 2007;133:799–807.
19 Marwaha A, Ford AC, Lim A, et al. Worldwide prevalence of dyspepsia: Systematic review and meta-analysis. *Gastroenterology* 2009;136(Suppl 1):A182.
20 Drossman DA, Li Z, Andruzzi E, et al. U.S. householder survey of functional gastrointestinal disorders. Prevalence, sociodemography, and health impact. *Dig Dis Sci* 1993;38:1569–80.
21 Ho KY, Kang JY, Seow A. Prevalence of gastrointestinal symptoms in a multiracial Asian population, with particular reference to reflux-type symptoms. *Am J Gastroenterol* 1998;93:1816–22.
22 Jones RH, Lydeard SE, Hobbs FDR, et al. Dyspepsia in England and Scotland. *Gut* 1990;31:401–5.
23 Katelaris PH, Tippett GH, Norbu P, et al. Dyspepsia, *Helicobacter pylori*, and peptic ulcer in a randomly selected population in India. *Gut* 1992;33:1462–6.
24 Shah SS, Bhatia SJ, Mistry FP. Epidemiology of dyspepsia in the general population in Mumbai. *Indian J Gastroenterol* 2001;20:103–6.
25 Ihezue CH, Oluwole FS, Onuminya JE, et al. Dyspepsia among the highlanders of Nigeria: An epidemiological survey. *Afr J Med Med Sci* 1996;25:23–9.

26 Farsakh NA, Saadeh A, Rawshdeh M, et al. Dyspepsia in the general population in Jordan. *Indian J Gastroenterol* 2000;19:68–70.
27 Schmulson M, Ortiz O, Santiago-Lomeli M, et al. Frequency of functional bowel disorders among healthy volunteers in Mexico City. *Dig Dis* 2006;24:342–7.
28 Bisschops R, Karamanolis G, Arts J, et al. Relationship between symptoms and ingestion of a meal in functional dyspepsia. *Gut* 2008;57:1495–503.
29 Harvey RF, Lane JA, Murray LJ, et al. Randomised controlled trial of effects of *Helicobacter pylori* infection and its eradication on heartburn and gastro-oesophageal reflux: Bristol *Helicobacter* project. *BMJ* 2004;328:1417–20.
30 Holtmann G, Goebell H, Holtmann M, et al. Dyspepsia in healthy blood donors: Pattern of symptoms and association with *Helicobacter pylori*. *Dig Dis Sci* 1994;5:1090–8.
31 Moayyedi P, Forman D, Braunholtz D, et al. The proportion of upper gastrointestinal symptoms in the community associated with *Helicobacter pylori*, lifestyle factors, and nonsteroidal anti-inflammatory drugs. *Am J Gastroenterol* 2000;95:1448–55.
32 Penston JG, Pounder RE. A survey of dyspepsia in Great Britain. *Aliment Pharmacol Ther* 1996;10:83–9.
33 Rosenstock S, Kay L, Rosenstock C, et al. Relation between *Helicobacter pylori* infection and gastrointestinal symptoms and syndromes. *Gut* 1997;41:169–76.
34 Schlemper RJ, Van der Werf SDJ, Biemond I, et al. Dyspepsia and *Helicobacter pylori* in Japanese employees with and without ulcer history. *J Gastroenterol Hepatol* 1995;10:633–8.
35 Schlemper RJ, Van der Werf SD, Vandenbroucke JP, et al. Nonulcer dyspepsia in a Dutch working population and *Helicobacter pylori*. Ulcer history as an explanation of an apparent association. *Arch Intern Med* 1995;155:82–7.
36 Shaib Y, El-Serag HB. The prevalence and risk factors of functional dyspepsia in a multiethnic population in the United States. *Am J Gastroenterol* 2004;99:2210–6.
37 Talley NJ, Zinsmeister AR, Schleck CD, et al. Smoking, alcohol, and analgesics in dyspepsia and among dyspepsia subgroups: Lack of an association in a community. *Gut* 1994;35:619–24.
38 Wildner-Christensen M, Moller Hansen J, Schaffalitzky De Muckadell O. Rates of dyspepsia one year after *Helicobacter pylori* screening and eradication in a Danish population. *Gastroenterology* 2003;125:372–9.
39 Woodward M, Morrison CE, McColl KEL. The prevalence of dyspepsia and use of antisecretory medication in North Glasgow: Role of *Helicobacter pylor*i vs. lifestyle factors. *Aliment Pharmacol Ther* 1999;13:1505–9.
40 Holtmann G, Gschossman J, Holtmann M, et al. *H. pylori* and functional dyspepsia. Increased serum antibodies as an independent risk factor? *Dig Dis Sci* 2001;46:1550–7.
41 Bytzer P, Howell S, Leemon M, et al. Low socioeconomic class is a risk factor for upper and lower gastrointestinal symptoms: A population based study in 15,000 Australian adults. *Gut* 2001;49:66–72.
42 Marwaha A, Ford AC, Lim A, et al. Risk factors for dyspepsia: Systematic review and meta-analysis. *Gastroenterology* 2009;136(Suppl 1):A58.
43 Marshall JK, Thabane M, Garg AX, et al. Incidence and epidemiology of irritable bowel syndrome after a large waterborne outbreak of bacterial dysentery. *Gastroenterology* 2006;131:445–50.
44 Marshall JK, Thabane M, Borgaonkar MR, et al. Postinfectious irritable bowel syndrome after a foodborne outbreak of gastroenteritis attributed to a viral pathogen. *Clin Gastroenterol Hepatol* 2007;5:457–60.
45 Neal KR, Barker L, Spiller RC. Prognosis in postinfective irritable bowel syndrome: A six-year follow up study. *Gut* 2002;51:410–3.
46 Tack J, Demedts I, Dehondt G, et al. Clinical and pathophysiological characteristics of acute-onset functional dyspepsia. *Gastroenterology* 2002;122:1738–47.
47 Mearin F, Perez-Oliveras M, Perello A, et al. Dyspepsia and irritable bowel syndrome after a *Salmonella* gastroenteritis outbreak: one year follow-up cohort study. *Gastroenterology* 2005;129:98–104.
48 Parry SD, Stansfield R, Jelley D, et al. Does bacterial gastroenteritis predispose people to functional gastrointestinal disorders? A prospective, community-based, case-control study. *Am J Gastroenterol* 2003;98:1970–5.
49 Tornblom H, Holmvall P, Svenungsson B, et al. Gastrointestinal symptoms after infectious diarrhea: A five-year follow-up in a Swedish cohort of adults. *Clin Gastroenterol Hepatol* 2007;5:461–4.
50 Ford AC, Thabane M, Collins SM, et al. Prevalence of uninvestigated dyspepsia 8 years after a large waterborne outbreak of bacterial dysentery: A cohort study. *Gastroenterology* 2010;138:1727–36.
51 Spiller R. Postinfectious functional dyspepsia and postinfectious irritable bowel syndrome: Different symptoms but similar risk factors. *Gastroenterology* 2010;138:1660–3.
52 Locke GR, III, Murray JA, Zinsmeister AR, et al. Celiac disease serology in irritable bowel syndrome and dyspepsia: A population-based case-control study. *Mayo Clin Proc* 2004;79:476–82.
53 Vivas S, Ruiz de Morales JM, Martinez J, et al. Human recombinant anti-transglutaminase antibody testing is

useful in the diagnosis of silent coeliac disease in a selected group of at-risk patients. *Eur J Gastroenterol Hepatol* 2003;15:479–83.
54 Ford AC, Ching E, Moayyedi P. Meta-analysis: Yield of diagnostic tests for coeliac disease in dyspepsia. *Aliment Pharmacol Ther* 2009;30:28–36.
55 Aro P, Ronkainen J, Storskrubb T, et al. Valid symptom reporting at upper endoscopy in a random sample of the Swedish adult general population: The Kalixanda study. *Scand J Gastroenterol* 2004;39:1280–8.
56 Azuma T, Ito Y, Suto H, et al. The effect of *Helicobacter pylori* eradication therapy on dyspepsia symptoms in industrial workers in Japan. *Aliment Pharmacol Ther* 2001;15:805–11.
57 Bernersen B, Johnsen R, Straume B, et al. Towards a true prevalence of peptic ulcer: the Sørreisa gastrointestinal disorder study. *Gut* 1990;31:989–92.
58 Harrison JD, Steele RJC, Morris DL, et al. Screening for gastric cancer: Endoscopic investigation of dyspeptic subjects identified by postal questionnaire. *GI Cancer* 2001;3:335–42.
59 Lu CL, Lang HC, Chang FY, et al. Prevalence and health/social impacts of functional dyspepsia in Taiwan: A study based on the Rome Criteria Questionnaire Survey assisted by endoscopic exclusion among a physical check-up population. *Scand J Gastroenterol* 2005;40:402–11.
60 Zagari RM, Law GR, Fuccio L, et al. Dyspeptic symptoms and endoscopic findings in the community: the Loiano-Monghidoro study. *Am J Gastroenterol* 2010;105:565–71.
61 Zhao Y, Zou D, Wang R, et al. Dyspepsia and irritable bowel syndrome in China: A population-based endoscopy study of prevalence and impact. *Aliment Pharmacol Ther* 2010;32:562–72.
62 Ford AC, Marwaha A, Lim A, et al. What is the prevalence of clinically significant endoscopic findings in subjects with dyspepsia? Systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2010;8:830–7.
63 Moayyedi P, Talley NJ, Fennerty MB, et al. Can the clinical history distinguish between organic and functional dyspepsia? *JAMA* 2006;295:1566–76.
64 Bowrey DJ, Griffin SM, Wayman J, et al. Use of alarm symptoms to select dyspeptics for endoscopy causes patients with curable esophagogastric cancer to be overlooked. *Surg Endosc* 2006;20:1725–8.
65 American Gastroenterological Association. American Gastroenterological Association Technical Review on the Evaluation of Dyspepsia. *Gastroenterology* 2005;129:1756–80.
66 National Institute for Health and Clinical Excellence. Dyspepsia: Managing dyspepsia in adults in primary care. http://www.nice.org.uk/nicemedia/pdf/CG017fullguideline.pdf (accessed May 2013).
67 Talley NJ, Vakil N, the Practice Parameters Committee of the American College of Gastroenterology. Guidelines for the management of dyspepsia. *Am J Gastroenterol* 2005;100:2324–37.
68 Adang RP, Vismans JF-JFE, Talmon JL, et al. Appropriateness of indications for diagnostic upper gastrointestinal endoscopy: Association with relevant endoscopic disease. *Gastrointest Endosc* 1995;42:390–7.
69 Numans ME, van der Graaf Y, de Wit NJ, et al. How useful is selection based on alarm symptoms in requesting gastroscopy? An evaluation of diagnostic determinants for gastro-oesophageal malignancy. *Scand J Gastroenterol* 2001;36:437–43.
70 Kapoor N, Bassi A, Sturgess R, et al. Predictive value of alarm features in a rapid access upper gastrointestinal cancer service. *Gut* 2005;54:40–5.
71 Wallace MB, Durkalski VL, Vaughan J, et al. Age and alarm symptoms do not predict endoscopic findings among patients with dyspepsia: A multicentre database study. *Gut* 2001;49:29–34.
72 Vakil N, Moayyedi P, Fennerty MB, et al. Limited value of alarm features in the diagnosis of upper gastrointestinal malignancy: Systematic review and meta-analysis. *Gastroenterology* 2006;131:390–401.
73 Bai Y, Li ZS, Zou DW, et al. Alarm features and age for predicting upper gastrointestinal malignancy in Chinese patients with dyspepsia with high background prevalence of *Helicobacter pylori* infection and upper gastrointestinal malignancy: An endoscopic database review of 102,665 patients from 1996 to 2006. *Gut* 2010;59:722–8.
74 Ford AC, Forman D, Bailey AG, et al. Fluctuation of gastrointestinal symptoms in the community: A 10-year longitudinal follow-up study. *Aliment Pharmacol Ther* 2008;28:1013–20.
75 Chang JY, LockeIII GR, McNally MA, et al. Impact of functional gastrointestinal disorders on survival in the community. *Am J Gastroenterol* 2010;105:822–32.
76 Camilleri M, Dubois D, Coulie B, et al. Prevalence and socioeconomic impact of upper gastrointestinal disorders in the United States: Results of the US Upper Gastrointestinal Study. *Clin Gastroenterol Hepatol* 2005;3:543–52.
77 Brook RA, Kleinman NL, Choung RS, et al. Functional dyspepsia impacts absenteeism and direct and indirect costs. *Clin Gastroenterol Hepatol* 2010;8:498–503.
78 El-Serag HB, Talley NJ. Systematic review: Health-related quality of life in functional dyspepsia. *Aliment Pharmacol Ther* 2003;18:387–93.

79 Enck P, Dubois D, Marquis P. Quality of life in patients with upper gastrointestinal symptoms: Results from the Domestic/International Gastroenterology Surveillance Study (DIGEST). *Scand J Gastroenterol* 1999;34(Suppl 231):48–54.
80 Halder SLS, Locke GR, Talley NJ, et al. Impact of functional gastrointestinal disorders on health-related quality of life: A population-based case-control study. *Aliment Pharmacol Ther* 2004;19:233–42.
81 Mones J, Adan A, Segu JL, et al. Quality of life in functional dyspepsia. *Dig Dis Sci* 2002;47:20–6.
82 Talley NJ, Weaver AL, Zinsmeister AR. Impact of functional dyspepsia on quality of life. *Dig Dis Sci* 1995;40:584–9.
83 Stanghellini V. Relationship between upper gastrointestinal symptoms and lifestyle, psychosocial factors and comorbidity in the general population: Results from the Domestic/International Gastroenterology Surveillance Study (DIGEST). *Scand J Gastroenterol* 1999;34(Suppl 231):29–37.
84 Jones R, Lydeard S. Prevalence of symptoms of dyspepsia in the community. *BMJ* 1989;298:30–2.
85 Ahlawat SK, Locke GR, Weaver AL, et al. Dyspepsia consulters and patterns of management: A population-based study. *Aliment Pharmacol Ther* 2005;22: 251–9.
86 Ford AC, Forman D, Bailey AG, et al. Who consults with dyspepsia? Results from a longitudinal ten-year follow-up study. *Am J Gastroenterol* 2007;102:957–65.
87 Howell S, Talley NJ. Does fear of serious disease predict consulting behaviour amongst patients with dyspepsia in general practice? *Eur J Gastroenterol Hepatol* 1999;11:881–6.
88 Koloski NA, Talley NJ, Boyce PM. Epidemiology and health care seeking in the functional GI disorders: A population-based study. *Am J Gastroenterol* 2002;97:2290–9.
89 Koloski NA, Talley NJ, Huskic SS, et al. Predictors of conventional and alternative health care seeking for irritable bowel syndrome and functional dyspepsia. *Aliment Pharmacol Ther* 2003;17:841–51.
90 Lydeard S, Jones R. Factors affecting the decision to consult with dyspepsia: Comparison of consulters and non-consulters. *J R Coll Gen Pract* 1989;39: 495–8.
91 Talley NJ, Boyce PM, Jones M. Dyspepsia and health care seeking in a community: How important are psychological factors? *Dig Dis Sci* 1998;43:1016–22.
92 Westbrook JI, Mcintosh J, Talley NJ. Factors associated with consulting medical or non-medical practitioners for dyspepsia: An Australian population-based study. *Aliment Pharmacol Ther* 2000;14:1581–8.
93 Koloski NA, Boyce PM, Talley NJ. Is health care seeking for irritable bowel syndrome and functional dyspepsia a socially learned response to illness? *Dig Dis Sci* 2005;50:153–62.
94 Taha AS, McCloskey C, Prasad R, et al. Famotidine for the prevention of peptic ulcers and oesophagitis in patients taking low-dose aspirin (FAMOUS): A phase III, randomised, double-blind, placebo-controlled trial. *Lancet* 2009;374:119–25.
95 Yeomans N, Lanas A, Labenz J, et al. Efficacy of esomeprazole (20 mg once daily) for reducing the risk of gastroduodenal ulcers associated with continuous use of low-dose aspirin. *Am J Gastroenterol* 2008;103:2465–73.
96 Taha AS, Hudson N, Hawkey CJ, et al. Famotidine for the prevention of gastric and duodenal ulcers caused by nonsteroidal antiinflammatory drugs. *New Engl J Med* 1996;334:1435–9.
97 Bianchi PG, Lazzaroni M, Imbesi V, et al. Efficacy of pantoprazole in the prevention of peptic ulcers, induced by non-steroidal anti-inflammatory drugs: A prospective, placebo-controlled, double-blind, parallel-group study. *Dig Liver Dis* 2000;32:201–8.
98 Lane JA, Murray LJ, Noble S, et al. Impact of *Helicobacter pylori* eradication on dyspepsia, health resource use, and quality of life in the Bristol *Helicobacter* project: Randomised controlled trial. *BMJ* 2006;332:199–202.
99 Moayyedi P, Feltbower R, Brown J, et al. Effect of population screening and treatment for *Helicobacter pylori* on dyspepsia and quality of life in the community: A randomised controlled trial. *Lancet* 2000;355:1665–9.
100 Ford AC, Forman D, Bailey AG, et al. A community screening program for *Helicobacter pylori* saves money: Ten-year follow-up of a randomised controlled trial. *Gastroenterology* 2005;129:1910–7.
101 Parsonnet J, Harris RA, Hack HM, et al. Modelling cost-effectiveness of *Helicobacter pylori* screening to prevent gastric cancer: A mandate for clinical trials. *Lancet* 1996;348:150–4.
102 Sonnenberg A, Inadomi JM. Medical decision models of *Helicobacter pylori* therapy to prevent gastric cancer. *Aliment Pharmacol Ther* 1998;12(Suppl 1):111–21.

Answers to multiple choice questions

1. D

The prevalence is somewhere between 5 % and 40 % in cross-sectional surveys. Prevalence varies according to both the criteria used to define its presence and geographical location, and usually remains remarkably stable during follow-up. Due to its chronic

nature, prevalence of dyspepsia is greater than incidence.

2. B

Proposed risk factors for dyspepsia include *Helicobacter pylori* infection, female gender, lower socioeconomic status, use of aspirin or nonsteroidal anti-inflammatory drugs, and acute enteric infection.

3. E

In a meta-analysis of cross-sectional surveys and case-control studies there was no association between celiac disease and symptoms compatible with dyspepsia. All the other conditions should be part of the differential diagnosis when a patient presents with dyspepsia.

16 Epidemiology of upper gastrointestinal bleeding

Colin J. Crooks[1], *Joseph Sung*[2] *& Timothy R. Card*[1]
[1]Division of Epidemiology and Public Health, Nottingham City Hospital, University of Nottingham, Nottingham, UK
[2]Department of Medicine & Therapeutics, Prince of Wales Hospital, Shatin, The Chinese University of Hong Kong, NT, Hong Kong

Key points

- Nonvariceal upper gastrointestinal bleeding has become a disease of the elderly with comorbid illnesses. This trend keeps the mortality of the condition relatively high despite advances in endoscopic and pharmacologic therapies.
- Although *Helicobacter pylori*-related peptic ulcer is declining, increased usage of combinations of preventative medications for ischemic heart disease and cerebral vascular disease have become an increasingly important cause of nonvariceal upper gastrointestinal bleeding. These require a careful assessment of the evidence for risks and benefits when prescribing for different populations.
- The most important risk factors predicting death in nonvariceal upper gastrointestinal bleeding are old age, comorbidities, severe bleeding as manifested by shock at presentation or fresh hematemesis, continued or recurrent bleeding, onset of bleeding while hospitalized for other causes, and major stigmata of bleeding.
- The incidence of variceal hemorrhage does not appear to be increasing despite an increase in cirrhosis, suggesting improving primary and secondary prevention strategies.
- The most important risk factor predicting death in variceal upper gastrointestinal hemorrhage apart from age is the underlying severity of the cirrhosis.

Clinical summary

Upper gastrointestinal hemorrhage defined as acute bleeding into the lumen of the gastrointestinal tract above the ligament of Trietz, leads typically to presentation with hematemesis or melena. It is the commonest emergency medical admission for gastroenterology [1], has an overall 28-day case fatality in the range 2–14 % [2,3] and is associated with a significant burden on healthcare resources. Upper gastrointestinal bleeding is commonly categorized as variceal (from esophageal or gastric varices) or nonvariceal bleeding. Nonvariceal bleeding is more common and can be further subdivided. The proportions of upper gastrointestinal bleeding admissions in each category are shown in Table 16.1. Variceal bleeding is reported in lower proportions in larger population-based studies than hospital-derived case series. However, comparisons between studies are difficult as many hospital studies only report cases that had an endoscopy performed, therefore excluding a large proportion of patients who remain undiagnosed.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 16.1 Diagnoses of patients presenting with upper gastrointestinal hemorrhage

Country	Year	Restricted to endoscoped cases?	Mallory–Weiss syndrome	Erosive inflammation	Varices	Ulceration	Malignancy	Other	Unspecified diagnosis	*n* =
Hospital-based studies										
Cameroon [103]	1990	Yes		22 %	14 %	47 %				172
Israel [104]	1994	Yes		20 %	13 %	46 %		21 %		321
Kenya [105]	1994	Yes			35 %	36 %			7 %	97
Canada [106]	2004	Yes		25 %		50 %			5 %	2484
Zambia [107]	2008	Yes	1 %	18 %	26 %	29 %	8 %	3 %	15 %	179
Italy [108]	2008	Yes	5 %	13 %		66 %	6 %	2 %	7 %	1844
Togo [109]	2010	Yes	11 %	16 %	18 %	41 %				44
Europe [13]	2011	Yes		33 %		35 %				2655
England* [98]	2011		5 %	59 %	11 %	36 %	4 %	3 %	17 %	6750
Population-based studies										
Scotland* [53]	1993		7 %	47 %	6 %	28 %	2 %	7 %	29 %	1882
England [76]	1993		5 %	24 %	4 %	31 %	4 %	6 %	25 %	4137
Crete [110]	1999			34 %	4 %	48 %	3 %	3 %	7 %	353
Netherlands [7]	2000			20 %	7 %	46 %	5 %	8 %	14 %	769
USA [3]	2006			12 %	9 %	34 %		4 %	41 %	N/A
England* [2]	2007		6 %	25 %	3 %	22 %	2 %		56 %	516,153
Wales [54]	2007		6 %	24 %	3 %	22 %	1 %		44 %	22,299
Italy [8]	2009		3 %	14 %	12 %	50 %	5 %	9 %	5 %	539

*Multiple diagnoses allowed.

Table 16.2 Variations in incidence of upper gastrointestinal bleeding between studies

Year of estimate	Country	Number of bleeds for estimate	Crude incidence per 100,000 person-years	Indirect age standardized incidence	95 % confidence interval	Study type
2007 [4]	Israel	864	17*	17	(16–18)	National admissions database
2000 [7]	Netherlands	769	48	45	(43–47)	10 hospitals in Amsterdam region
1991 [11]	USA	3294	36**	71	(68–73)	139 military facilities
2003 [9]	Canada	13,017	53***	50	(49–51)	National admissions database
2006 [111]	Spain	291	66	55	(49–62)	Single hospital
2004 [8]	Italy	21	74	59	(36–90)	Single hospital
2006 [3]	USA	N/A	82****	89	(88–90)	National inpatient sample
2007 [112]	England	34,482	84	84	(84–84)	National admissions database
1996 [113]	France	2,133	84	73	(70–76)	29 hospitals in 1 region
1993 [76]	England	3,508	89	77	(74–79)	74 hospitals in 4 regions
2005 [10]	Greece	353	98	85	(76–94)	3 hospitals in 1 region
2002 [27]	Scotland	211	99	83	(72–95)	Single hospital
1999–2007 [54]	Wales	22,299	119*****	99	(98–101)	National admissions database
1999 [110]	Crete	21	149	137	(84–209)	All hospitals in 1 region
1993 [53]	Scotland	1720	157	135	(129–142)	19 hospitals in 1 region

*No hematemesis or melena codes.
**More restrictive definition requiring combinations of codes for non-ulcer codes.
***Military population.
****Estimates extrapolated from 20 % national sample, vulnerable to sampling biases.
*****Estimate only available as aggregate for all years.

This chapter will examine the occurrence, causes, and outcomes of upper gastrointestinal hemorrhage for both variceal and nonvariceal bleeding.

Incidence of acute upper gastrointestinal bleeding

The reported incidence of upper gastrointestinal bleeding varies widely as can be seen from Table 16.2. Recent large European and North American studies suggest figures in the region of 50–100/100,000 person-years. Though some of the differences in incidence around the world are doubtless genuine, some of the variation in the figures may be a consequence of different case definitions, management systems, timing of studies, and study methodology. For example, the highest incidence estimates were reduced when indirectly standardized for age, and two of the lower incidence studies used restrictive definitions of upper gastrointestinal bleeds [4,5]. Differences in clinical management may also have an effect if the case definition depends on hospitalization, for example within the United States the proportion of patients managed without a hospital admission varies by over twofold between states (19–45 %) [6]. It is unclear to what extent changes over time are similarly explained, and to what extent they reflect disease changes such as reduction in *H. pylori* carriage but over the last two decades countries including the United States, Canada, Israel, Netherlands, Greece, and Italy, have reported reductions in overall upper gastrointestinal bleeding admissions of between 10 and 40 % [3,4,7–10]. Finally, some studies report specific subsections of the population with differing risk. For example, a low incidence has been reported from a military population [11]. There is little literature on the occurrence of variceal hemorrhage separately from nonvariceal bleeding. The proportions of variceal hemorrhage reported in the larger population-based

studies are between 3 and 9 % (Table 16.1) suggesting an incidence of between 2.1 and 8.1 per 100,000 person-years.

Causes and their trends

Nonvariceal upper gastrointestinal hemorrhage

Peptic ulceration and erosion is the most frequently identified cause of upper gastrointestinal hemorrhage. Its incidence has been variously described as declining over the last two decades (e.g. Sweden (64–35/100,000) [12], Spain (55–26/100,000) [13], USA (71–57/100,000) [14]) or as decreasing among young people and increasing in the elderly [7,8,15–17]. Changes in the occurrence of peptic ulcer bleeding will reflect trends in underlying risk factors. For peptic ulceration the risk factors with the highest attributable fractions are NSAIDs and antiplatelet medications [18] (Table 16.3). In addition *H. pylori* is associated with a fivefold increase in bleeding episodes independently of these medications [19].

Helicobacter pylori

H. pylori was historically the most important cause of peptic ulceration. It is generally acquired during childhood, and prevalence is reducing with generations [20] and among peptic ulcer bleeding admissions [14,21]. However, a recent systematic review suggested that *H. pylori* prevalence in peptic ulcer bleeding is underestimated and that the mean prevalence remains high at 72 % in study populations [22]. The lowest prevalence estimates in studies included were reported from the United Kingdom, Italy, the Netherlands, and Denmark (<50 %), though country was not found to be a significant predictor of *H. pylori* prevalence in multivariate analysis. *H. pylori* does not appear to further potentiate the individual risks of medications such as NSAIDS, rather the increased risk from *H. pylori* is merely additive to that from medications [23].

Table 16.3 Estimated adjusted attributable fractions for peptic ulcer bleeding [18]

	Attributable fraction
Previous peptic ulcer	19 %
Smoking	2 %
Heart failure	5 %
Diabetes	4 %
Steroids	3 %
Anticoagulants	3 %
NSAIDs	22 %
Aspirin	11 %

Source: Data from Weil et al. 2000 [18].

Medications

As stated earlier, NSAIDs and antiplatelet agents are important risk factors for upper gastrointestinal bleeding. NSAID use carries a relative risk of gastrointestinal bleeding events of 3.8 (3.6–4.1) [24] which is removed by cessation, and this translates for nonselective NSAIDS users in clinical trials into an incidence of upper gastrointestinal bleeding of up to 560 per 100,000 person-years [25]. Selective cyclo-oxygenase-2 inhibitors are associated with lower risks than nonselective NSAIDs [26], but although there has been an increase in their prescription over the last decade, there has been minimal change in the overall prescription of NSAIDS and it is unlikely the changes account for any overall trends in bleeding incidence [10,27,28].

One percent of patients on low-dose aspirin (the most commonly used antiplatelet agent) have a gastrointestinal bleed within 28 months (number needed to harm per year = 248 [29]). With increasing use of these drugs the contribution of aspirin to bleeding is rising as shown by the near doubling over 6 years of the rate of bleeding admissions prescribed aspirin or anticoagulants in the northeast of Scotland [27]. Prescribing decisions are therefore a balance between the risks and benefits of these drugs. For example, low-dose aspirin given for low-risk primary prevention (1 % cardiovascular risk over 5 years) prevents 1–4 myocardial infarctions a year and causes 2–4 gastrointestinal bleeding events with no improvement in mortality [30]. For patients with high cardiovascular risk or for secondary prevention antiplatelet and anticoagulants are increasingly given in combinations and this can further increase risk. A recent meta-analysis shows low-dose aspirin increases the risk of bleeding from the gastrointestinal tract by 31 %, a further 81 % when combined with clopidogrel, and a further 91 % when combined with warfarin [31].

Proton pump inhibitors consistently reduce the risk of bleeding associated with NSAIDS by 67 % [32] and their use has a demonstratable cost benefit [33]. In

patients on low-dose aspirin the risk of bleeding is similarly reduced [31]; however, there has been some concern about proton pump inhibitors reducing the efficacy of clopidogrel when coprescribed. A large cohort study reassuringly did not find an increased cardiovascular risk and estimated that only if the cardiovascular risk was increased by more than 19 % would the risks of proton pump inhibitors outweigh their benefits [34]. A randomized controlled trial of proton pump inhibitors for patients on dual antiplatelet therapy found a reduction in upper gastrointestinal bleeding (HR 0.13 (0.03–0.56)) with no difference in cardiovascular outcomes (HR 0.99 (0.68–1.44)) [35].

Other drug associations with bleeding that have been reported include an up to threefold increased risk from SSRIs [36,37,38], twofold increased risk from spironolactone [39,40], 2.5-fold increased risk from iron supplementation [41], two- to fourfold increased risk from corticosteriods [42], and threefold increased risk from bisphosphonates [43,44].

Comorbidities

It is difficult to ascertain with certainty the role of comorbidities in causing gastrointestinal bleeding independent of their therapies. It is widely assumed that the 1–3 % incidence of gastrointestinal bleeding during the month following an acute coronary syndrome (ACS) [35,45,46] is largely related to therapies. However, this is not necessarily the case for all comorbidities. Acute renal failure, for example, has a high incidence of upper gastrointestinal bleeding 13 % [47] with a subsequent increase in mortality (adjusted OR 2.6 (1.3–5.1)), and following surgical procedures at two university hospitals ($n = 25{,}845$), gastrointestinal bleeding incidence was reported in 0.39 % of patients, with an associated mortality of 31 %. Most of this bleeding was due to erosive gastritis (70 %) or ulceration (18 %) and occurred in patients with sepsis, multiorgan dysfunction, as well as in those who were prescribed NSAIDS during the admission. It is possible therefore that an aging population with increasing comorbidity may contribute to trends in gastrointestinal bleeding directly.

Other

There are a number of other risk factors for gastrointestinal hemorrhage. Higher alcohol intake, for example, is associated with a higher risk [48]. Ex-drinkers, however, remain at a slightly lower yet still elevated risk (after adjusting for smoking, previous ulcers, aspirin, and NSAIDS) suggesting that there is an underlying confounder associated with alcohol excess [49]. Smoking is also a risk factor [18,50], and its effect may be mediated through altering the ulcerative effects of *H. pylori* [51]. It is possible likewise that smoking to some extent mediates a steep socioeconomic gradient long shown to exist for peptic ulcer disease [52] and more recently for upper gastrointestinal hemorrhage also [53,54]. This gradient exists for all causes of upper gastrointestinal bleeding [2], and may also be contributed to by differences in prescribing practices, alcohol consumption or *H. pylori* prevalence. Finally high altitudes are associated with an increased incidence of gastrointestinal hemorrhage among migrant workers [55], as well as among acclimatized people [56]. This is possibly as part of the syndrome of both acute and chronic altitude sickness [57], though interestingly in the latter bleeding can actually be therapeutic in avoiding complications of high blood cell counts.

Variceal hemorrhage

Esophageal and gastric varices are a complication of portal hypertension usually due to cirrhosis. Among cirrhotic patients admitted with upper gastrointestinal hemorrhage 78–87 % are due to bleeding varices [58,59]. The predictors of variceal hemorrhage therefore are the causes of cirrhosis and its progression, and the subsequent development of portal hypertension. Acute precipitants of variceal hemorrhage in patients with known varices include excess alcohol consumption the week before admission, constipation, and vomiting [60]. However, whilst the prevalence and incidence of cirrhosis is increasing [61,62], the occurrence of variceal hemorrhage is not [2,3,63]. This could be because of improved primary prevention with increased use of banding and beta-blockers, or because cirrhosis is being diagnosed earlier.

Natural history and risk stratification

The natural history of a condition is the course it would take without intervention, and for a frequently mortal condition such as upper gastrointestinal

hemorrhage we cannot aim to study this. What we can do is to look at the outcome of the condition with treatment, and how changes in therapy have altered this.

At the beginning of the twentieth century, hospital mortality from hematemesis and melena due to peptic ulcers was reported to be be over 20 % for patients over 40 years old [64]. Mortality was higher in older patients and in those in whom bleeding recurred. The first advance in bleeding management was the use of generous blood transfusions guided by measured hemoglobin concentration given in a controlled intravenous drip [65]. Surgery was advocated following resuscitation when bleeding continued or reoccurred for those who were diagnosed with peptic ulceration, though the selection of patients and reported mortality varied widely and was controversial [66,64]. Indeed generous early eating regimes apparently demonstrated a strikingly low hospital mortality [67].

However, comparisons of the mortality from these early case series are difficult, as cases and deaths not thought to be due directly to bleeding, such as those from malignancy or cirrhosis, were often excluded [64]. Concerning this Lewin and Truelove commented "... it is noteworthy that the literature shows that most series with a low fatality rate have come from interested single physicians presenting their own cases, whereas studies of gross hospital figures commonly indicate a much less favourable prognosis.... We believe that mass hospital figures are more truly representative of the dangers of haematemesis than are the results obtained by a few specialists, provided that the data are handled with an appreciation of possible fallacies" [66]. Lewin and Truelove's case series in 1949 (median age about 50 years) of all presentations with hematemesis and melena in Oxford estimated a high mortality of 19 % following chronic ulcers, 7 % following acute ulcers, 24 % following other diagnoses, and 33 % where no diagnosis was made.

By the 1960–70s medical management was similar to that developed during the 1930s with early feeding and generous blood transfusions guided by hemoglobin measurement. Following medical management over 70 % of peptic ulcer bleeds and 44 % of variceal bleeds resolved with no further bleeding [68, 69]. Surgery was mostly reserved for those with unstable ulcer bleeding, whereas other causes such as varices and gastric cancer were not amenable to emergency treatment. Gastroscopy was recommended acutely for early diagnosis where a barium meal was inconclusive [70]. In 1967–8 the overall mortality in Aberdeen was reported to be 14 % for all admissions over 12 years old with hematemesis and melena (median age about 60 years), but this increased to 29 % if further bleeding occurred [69]. Age and comorbidity were consistently predictors of further bleeding, and for specific diagnoses mortality for peptic ulcer bleeding was 5 %, for variceal bleeding was 24 %, for other causes was 47 %, and for undiagnosed bleeding was 12 %.

Over the last few decades improvements in endoscopic therapy have been shown to reduce risks of rebleeding, for example the increased use of combination therapies [71], variceal banding [72] and the use of proton pump inhibitors to reduce stomach pH and promote clot stability [73], (a similar approach to that originally intended by early feeding). For variceal hemorrhage the use of antibiotics and glypression at the time of variceal bleeding has also been shown to reduce mortality [74,75]. At the same time that these improvements have been developed, the age of those admitted has risen. The median age of patients being admitted with bleeding from nonvariceal causes during the last two decades is around 70 years old [76,77], and for variceal patients around 55 years [2].

Nonvariceal hemorrhage

The consistent tendency noted at the start of the last century for comorbidity and advanced age to predict worse outcome has been extended by a number of authors to develop risk stratification strategies to aid in selecting the appropriate level of care. Well-validated scores include the Rockall and Blatchford scores [78,79] which allow selection of the lowest risk patients for early discharge [80,81,82]. Major risk factors predicting death include old age, comorbidities, shock at presentation, continued or recurrent bleeding, and onset of bleeding while hospitalized for other causes. Ulcers with active bleeding, or stigmata of recent bleeding such as a visible vessel or an adherent clot also predict re-bleeding and mortality risk.

At present, however, the evidence to suggest that adoption of newer management strategies has changed outcomes overall is disappointingly limited [83]. In fact, although there has been a wide variation of

Table 16.4 Short-term and inpatient mortality from variceal hemorrhage from studies $n > 1000$

Year	Country	Size	Mortality
1970–2000 [93]	Many*	1475	55–40 %
1981–1991 [114]	USA (Veteran Affairs)	4975	30–21 %
1988–2004 [115]	USA (NIS)	N/A	18–12 %
2004 [116]	USA (NIS)	6000	11 %
1998–2005 [92]	USA (NIS)	36,734	11 %
1999–2007 [2]	England	14,682	25–21 %

*Control groups in randomized controlled trials.

overall short-term mortality from nonvariceal hemorrhage with low estimates from the United States and some of Europe and higher estimates from elsewhere in Europe (see Table 16.4), mortality in a large cohort of upper gastrointestinal bleeds in the United Kingdom remains at about 14 % [2]. Causes of death, however, have changed. Papers published in the 1930s–1960s suggested that about 50 % of patients were dying from exsanguination before treatment or from re-bleeding. However, following endoscopic therapy more recent studies have found only 18–30 % of deaths were bleeding-related [84–87], indeed a recent trial has demonstrated that cardiovascular mortality is so important that patients with gastrointestinal hemorrhage benefit from an early reintroduction of aspirin [88]. The high proportion of nonbleeding-related deaths over the last decade is consistent with routinely collected English hospital data that demonstrates a gradual reduction in mortality that has been improving whether or not an endoscopy has been performed [2]. The reductions occurred for all the underlying causes of bleeding and it seems likely that improvements in the management of coexisting illnesses has caused the more recent reductions in mortality, rather than changes in specific bleed management. Following hospital discharge for upper gastrointestinal bleeding from peptic ulcers, mortality remains elevated twofold for up to 6 years compared to the general population [89–91]. Much of this long-term increase in mortality is related to comorbidity, particularly cancer and cardiovascular disease [91], and up to 50 % is due to smoking-related diseases [89]. As discussed earlier this suggests that upper gastrointestinal bleeding might itself be a marker for sicker patients with more comorbidity.

Variceal hemorrhage

The inpatient mortality of variceal hemorrhage remains on average higher than that of nonvariceal bleeding with large studies suggesting a mortality of 11–40 % [92,93]. Estimates of short-term mortality are generally limited by small sample sizes; however, studies with more than 1000 patients show a persistently higher mortality than for nonvariceal hemorrhage that is reducing over time (Table 16.5). Most deaths occur within the first 2 weeks [59]. Variceal bleeding is itself recognized as a prognostic indicator of the progression of cirrhosis [94]. The outcomes following variceal bleeding are generally related to the underlying severity of cirrhosis as demonstrated by the fact that general prognostic scores for cirrhosis, such as MELD or Child-Pugh, are useful predictors of mortality and rebleeding following variceal hemorrhage [95–97].

Healthcare costs

Healthcare costs vary between countries, but upper gastrointestinal bleeding is a consistently large consumer of them. Nonvariceal hemorrhage is associated with a median length of stay of 4–5 days [2,6,98] and variceal hemorrhage 7–9 days [2,99]. Using the National Inpatient Sample from the United States (restricted to patients who survived to discharge), the costs for an uncomplicated nonvariceal bleed were US$3402, and when associated with complications US$5632 [100]. For variceal hemorrhage the costs were US$6612 and US$23,207, respectively. However, within the USA a higher proportion of upper gastrointestinal hemorrhage admissions are managed in an ITU setting [101]. In contrast, lower estimates were derived from Canada for nonvariceal hemorrhage at $1883, and these costs increased with age and decreased with previous history of peptic ulcer disease. In Ireland the average cost for a nonvariceal hemorrhage admission is €2537; however, interestingly 75 % of the expenditure is on patients with a Rockall score ≤ 3 [102]. Overal costs, based upon the incidence figures quoted above, suggest an expenditure of between

Table 16.5 Thirty-day or inhospital mortality for nonvariceal upper gastrointestinal hemorrhage reported from population-based studies with $n > 1000$

Year	Country	Size	NVGIB in patient mortality (28 day)
1993–2003 [9]	Canada	95,905	4 %*
1993 [76]	England	4486	14 %
1999–2007 [2]	England	501,471	15–13 %
1996 [113]	France	2133	14 %
2005 [117]	France	1665	11 %
1996 2000 [118]	France	1165	12–7 %
1996–2007 [4]	Israel	12,074	8–7 %**
1993–2000 [7]	Netherlands	1582	14–13 %
1997 [53]	Scotland	1882	7 %
2004 [6]	USA (Medicare)	5617 (5 % stratified sample)	8 %***
1998–2006 [3]	USA (NIS)	N/A (20 % stratified sample)	4–3 %
1999–2007 [54]	Wales	24,421	10 %
1983–2004 [8]	Italy	1126	16–9 %

*Excluded melena; gastrointestinal bleeding, unspecified; hemorrhage of esophagus.
**Excluded hematemesis; melena; nonspecific GI bleeding.
***Excluded melena; nonspecific GI bleeding.

US$170,000 and US$563,000 per annum per 100,000 population in the United States.

Multiple choice questions

1 The incidence of nonvariceal upper gastrointestinal hemorrhage in large recent population-based studies is
- **A** Increasing rapidly
- **B** 50 per 10,000 per annum
- **C** 50 to 100 per 1000,000 per annum
- **D** Declining rapidly
- **E** 50 to 100 per 100,000 per annum

2 Mortality rates from nonvariceal gastrointestinal hemorrhage are
- **A** Very low
- **B** Rapidly rising
- **C** About 25 % in most series currently
- **D** Little changed over four decades
- **E** Less than 2 % for hospitalized cases

3 Important predictors of mortality from nonvariceal gastrointestinal hemorrhage include
- **A** Age
- **B** Hemoglobin concentration at admission
- **C** Gender
- **D** A low resting pulse
- **E** Only liver disease among comorbidities

4 Regarding variceal hemorrhage
- **A** Its incidence is increasing rapidly due to the increase in cirrhosis
- **B** Costs are less per admission than nonvariceal hemorrhage
- **C** Case fatality continues to increase over time
- **D** Patients are older than nonvariceal hemorrhage patients
- **E** Outcomes can be predicted from the severity of the underlying cirrhosis

5 Variation in incidence and mortality estimates worldwide
- **A** Is increased when the age of the study populations is adjusted for
- **B** Is affected by length of stay if only inpatient deaths are recorded
- **C** Is unaffected by different national and regional management strategies
- **D** Can all be explained by random error
- **E** Shows evidence of selection bias in population-based studies

References

1 Williams JG, Roberts, Ali MF, et al. Gastroenterology services in the UK. The burden of disease, and the organisation and delivery of services for gastrointestinal and liver disorders: a review of the evidence. *Gut* 2007;56(Suppl 1):1–113.

2 Crooks C, Card T, West J. Reductions in 28-day mortality following hospital admission for upper gastrointestinal hemorrhage. *Gastroenterology* 2011;141(1): 62–70.

3 Zhao Y, Encinosa W. Hospitalizations for gastrointestinal bleeding in 1998 and 2006. *HCUP Statistical Brief* 2008;65:1–12.

4 Hershcovici T, Haklai Z, Gordon ES, Zimmerman J. Trends in acute non-variceal bleeding in Israel in 1996–2007: A significant decrease in the rates of bleeding peptic ulcers. *Digest Liver Dis* 2010;42(7):477–81.

5 Targownik LE, Al-Mamfud A. The prevalence of risk factors for gastrointestinal complications and use of gastroprotection among persons hospitalized for cardiovascular disease. *Aliment Pharmacol Ther* 2006;23(6):743–9.

6 Cooper GS, Kou TD, Wong RC. Outpatient management of nonvariceal upper gastrointestinal hemorrhage: unexpected mortality in Medicare beneficiaries. *Gastroenterology* 2009;136(1):108–14.

7 van Leerdam ME, Vreeburg EM, Rauws EAJ, et al. Acute upper GI bleeding: did anything change? : Time trend analysis of incidence and outcome of acute upper GI bleeding between 1993/1994 and 2000. *Am J Gastroenterol* 2003;98(7):1494–9.

8 Loperfido S, Baldo V, Piovesana E, et al. Changing trends in acute upper-GI bleeding: a population-based study. *Gastrointest Endosc* 2009;70(2): 212–24.

9 Targownik LE, Nabalamba A. Trends in management and outcomes of acute nonvariceal upper gastrointestinal bleeding: 1993–2003. *Clin Gastroenterol Hepatol* 2006;4(12):1459–66.

10 Theocharis GJ, Thomopoulos KC, Sakellaropoulos G, et al. Changing trends in the epidemiology and clinical outcome of acute upper gastrointestinal bleeding in a defined geographical area in Greece. *J Clin Gastroenterol* 2008;42(2):128–33.

11 Yavorski RT, Wong RK, Maydonovitch C, et al. Analysis of 3,294 cases of upper gastrointestinal bleeding in military medical facilities. *Am J Gastroenterol* 1995;90(4):568–73.

12 Ahsberg K, Ye W, Lu Y, et al. Hospitalisation of and mortality from bleeding peptic ulcer in Sweden: a nationwide time-trend analysis. *Aliment Pharmacol Therap* 2011;33(5):578–84.

13 Lanas A, Aabakken L, Fonseca J, et al. Clinical predictors of poor outcomes among patients with nonvariceal upper gastrointestinal bleeding in Europe. *Aliment Pharmacol Ther* 2011;33(11):1225–33.

14 Feinstein LB, Holman RC, Christensen KLY, et al. Trends in hospitalizations for peptic ulcer disease, United States, 1998–2005. *Emerg Infect Dis* 2010;16(9):1410–18.

15 Higham J, Kang JY, Majeed A. Recent trends in admissions and mortality due to peptic ulcer in England: increasing frequency of haemorrhage among older subjects. *Gut* 2002;50(4):460–4.

16 Kang JY, Elders A, Majeed A, et al. Recent trends in hospital admissions and mortality rates for peptic ulcer in Scotland 1982–2002. *Aliment Pharmacol Ther* 2006;24(1):65–79.

17 Ohmann C, Imhof M, Ruppert C, et al. Time-trends in the epidemiology of peptic ulcer bleeding. *Scand J Gastroenterol* 2005;40(8):914–20.

18 Weil J, Langman MJS, Wainwright P, et al. Peptic ulcer bleeding: accessory risk factors and interactions with non-steroidal anti-inflammatory drugs. *Gut* 2000;46(1):27–31.

19 Lanas A, Fuentes J, Benito R, et al. *Helicobacter pylori* increases the risk of upper gastrointestinal bleeding in patients taking low-dose aspirin. *Alimentary Pharmacology and Therapeutics* 2002;16(4):779–86.

20 Morris Brown, Linda. *Helicobacter Pylori*: Epidemiology and routes of transmission. *Epidemiologic Reviews* 2000;22(2):283–97.

21 Bakkevold KE. Time trends in incidence of peptic ulcer bleeding and associated risk factors in Norway 1985–2008. *Clin Exp Gastroenterol* 2010;3(3):71–7.

22 Sánchez-Delgado J, Gené E, Suárez D, et al. Has *H. pylori* prevalence in bleeding peptic ulcer been underestimated? A meta-regression. *Am J Gastroenterol* 2011;106(3):398–405.

23 Cullen DJE, Hawkey GM, Greenwood DC, et al. Peptic ulcer bleeding in the elderly: relative roles of *Helicobacter pylori* and non-steroidal anti-inflammatory drugs. *Gut* 1997;41(4):459–62.

24 Hernández-Díaz S, García Rodríguez LA. Association between nonsteroidal anti-inflammatory drugs and upper gastrointestinal tract bleeding/perforation: An overview of epidemiologic studies published in the 1990s. *Arch Intern Med* 2000;160(14):2093–9.

25 Salvo F, Fourrier-Réglat A, Bazin F, et al. Cardiovascular and gastrointestinal safety of NSAIDs: a systematic review of meta-analyses of randomized clinical trials. *Clin Pharmacol Ther* 2011;89(6):855–66.

26 Chang CH, Chen HC, Lin JW, et al. Risk of hospitalization for upper gastrointestinal adverse events associated with nonsteroidal anti-inflammatory drugs:

a nationwide case-crossover study in Taiwan. *Pharmacoepidemiol Drug Saf* 2011;20(7):763–71.
27 Taha AS, Angerson WJ, Knill-Jones RP, Blatchford O. Upper gastrointestinal haemorrhage associated with low-dose aspirin and anti-thrombotic drugs – a 6-year analysis and comparison with non-steroidal anti-inflammatory drugs. *Aliment Pharmacol Ther* 2005;22(4):285–9.
28 Cai S, García Rodríguez LA, Massó-González EL, Hernández-Díaz S. Uncomplicated peptic ulcer in the UK: trends from 1997 to 2005. *Aliment Pharmacol Ther* 2009;30(10):1039–48.
29 Derry S, Loke YK. Risk of gastrointestinal haemorrhage with long term use of aspirin: meta-analysis. *BMJ* 2000;321(7270):1183.
30 Hayden M, Pignone M, Phillips C, Mulrow C. Aspirin for the primary prevention of cardiovascular events: A summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002;136(2):161–72.
31 Lanas A, Wu P, Medin J, Mills EJ. Low doses of acetylsalicylic acid increase risk of gastrointestinal bleeding in a meta-analysis. *Clin Gastroenterol Hepatol* 2011;9(9):762–8.
32 Vonkeman HE, Fernandes RW, van der Palen J, et al. Proton-pump inhibitors are associated with a reduced risk for bleeding and perforated gastroduodenal ulcers attributable to non-steroidal anti-inflammatory drugs: a nested case-control study. *Arthritis Res Ther* 2007;9(3):R52.
33 Abraham NS, Hartman C, Hasche J. Reduced hospitalization cost for upper gastrointestinal events that occur among elderly veterans who are gastroprotected. *Clin Gastroenterol Hepatol* 2010;8(4):350–6; quiz e45.
34 Ray WA, Murray KT, Griffin MR, et al. Outcomes with concurrent use of clopidogrel and proton-pump inhibitors: a cohort study. *Ann Intern Med* 2010;152(6):337–45.
35 Bhatt DL, Cryer BL, Contant CF, et al. Clopidogrel with or without omeprazole in coronary artery disease. *New Engl J Med* 2010;363(20):1909–17.
36 Oksbjerg Dalton S, Johansen C, Mellemkjaer L, et al. Use of selective serotonin reuptake inhibitors and risk of upper gastrointestinal tract bleeding: a population-based cohort study. *Arch Int Med* 2003;163(1):59–64.
37 van Walraven C, Mamdani MM, Wells PS, Williams JI. Inhibition of serotonin reuptake by antidepressants and upper gastrointestinal bleeding in elderly patients: retrospective cohort study. *BMJ (Clin Res Ed)* 2001;323(7314):655–8.
38 Abajo FJ, García Rodríguez LA, Montero D. Association between selective serotonin reuptake inhibitors and upper gastrointestinal bleeding: population based case-control study. *BMJ Int Ed* 1999;319(7217):1106–9.
39 Gulmez SE, Lassen AT, Aalykke C, et al. Spironolactone use and the risk of upper gastrointestinal bleeding: a population-based case-control study. *Br J Clin Pharmacol* 2008;66(2):294–9.
40 Russo A, Autelitano M, Bisanti L. Spironolactone and gastrointestinal bleeding: a population based study. *Pharmacoepidemiol Drug Saf* 2008;17(5):495–500.
41 Gallerani M, Simonato M, Manfredini R, et al. Risk of hospitalization for upper gastrointestinal tract bleeding. *J Clin Epidemiol* 2004;57(1):103–10.
42 Nielsen G. Risk of hospitalization resulting from upper gastrointestinal bleeding among patients taking corticosteroids: a register-based cohort study. *Am J Med* 2001;111(7):541–5.
43 Donahue JG, Chan KA, Andrade SE, et al. Gastric and duodenal safety of daily alendronate. *Arch Int Med* 2002;162(8):936–42.
44 Marshall JK, Rainsford KD, James C, Hunt RH. A randomized controlled trial to assess alendronate-associated injury of the upper gastrointestinal tract. *Aliment Pharmacol Therap* 2000;14(11):1451–7.
45 Al-Mallah M, Bazari R, Jankowski M, Hudson M. Predictors and outcomes associated with gastrointestinal bleeding in patients with acute coronary syndromes. *J Thromb Thrombolysis* 2007;23(1):51–5.
46 Ng FH, Wong SY, Lam KF, et al. Gastrointestinal bleeding in patients receiving a combination of aspirin, clopidogrel, and enoxaparin in acute coronary syndrome. *Am J Gastroenterol* 2008;103(4):865–71.
47 Fiaccadori E, Maggiore U, Clima B, et al. Incidence, risk factors, and prognosis of gastrointestinal hemorrhage complicating acute renal failure. *Kidney Int* 2001;59(4):1510–19.
48 Kelly JP, Kaufman DW, Koff RS, et al. Alcohol consumption and the risk of major upper gastrointestinal bleeding. *Am J Gastroenterol* 1995;90(7):1058–64.
49 Watanabe H, Kamijima Y, Sato T, et al. Ex-drinking may be a surrogate for unmeasured risk factors for upper gastrointestinal bleeding: reappraisal and an additional survey of subjects from a case-control study in Japan. *Eur J Epidemiol* 2009;24(3):143–7.
50 Kaplan RC, Heckbert SR, Koepsell TD, et al. and Investigators for the Cardiovascular Health Study. Risk factors for hospitalized gastrointestinal bleeding among older persons. *J Am Geriatr Soc* 2001;49(2):126–33.
51 Stack WA, Atherton JC, Hawkey GM, et al. Interactions between *Helicobacter pylori* and other risk factors for peptic ulcer bleeding. *Aliment Pharmacol Therap* 2002;16(3):497–506.
52 Pulvertaft CN. Comments on the incidence and natural history of gastric and duodenal ulcer. *Postgrad Med J* 1968;44(514):597–602.

53 Blatchford O, Davidson LA, Murray WR, et al. Acute upper gastrointestinal haemorrhage in west of Scotland: case ascertainment study. *BMJ* 1997;315(7107):510–14.
54 Button LA, Roberts SE, Evans PA, et al. Hospitalized incidence and case fatality for upper gastrointestinal bleeding from 1999 to 2007: a record linkage study. *Aliment Pharmacol Therap* 2011;33(1): 64–76.
55 Wu TY, Ding SQ, Liu JL, et al. High-altitude gastrointestinal bleeding: an observation in Qinghai-Tibetan railroad construction workers on Mountain Tanggula. *World J Gastroenterol* 2007;13(5):774.
56 Villanueva Palacios J, López de Guimaraes D, Avila Polo F. [Upper digestive tract hemorrhage in the Peruvian Andes: report of 115 cases observed in Huaraz.] *Rev Gastroenterol Perú* 1996;16(2):99–104.
57 Berríos J, Sedano O, Calle E, et al. [Upper digestive hemorrhage in the inhabitants of high altitudes in Peru.] *Rev Gastroenterol Perú* 1996;16(1):13–8.
58 D'Amico G, de Franchis R. Upper digestive bleeding in cirrhosis. Post-therapeutic outcome and prognostic indicators. *Hepatology* 2003;38(3): 599–612.
59 Pinto HC, Abrantes A, Esteves AV, et al. Long-term prognosis of patients with cirrhosis of the liver and upper gastrointestinal bleeding. *Am J Gastroenterol* 1989;84(10):1239–43.
60 Liao WC, Hou MC, Chang CJ, et al. Potential precipitating factors of esophageal variceal bleeding: a case-control study. *Am J Gastroenterol* 2010;106(1):96–103.
61 Fleming KM, Aithal GP, Solaymani-Dodaran M, et al. Incidence and prevalence of cirrhosis in the United Kingdom, 1992–2001: a general population-based study. *J Hepatol* 2008;49(5):732–8.
62 Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. *Lancet* 2006;367(9504):52–6.
63 Stokkeland K, Brandt L, Ekbom A, Hultcrantz R. Improved prognosis for patients hospitalized with esophageal varices in Sweden 1969–2002. *Hepatology* 2006;43(3):500–5.
64 Jones FA. Haematomesis and melaena with special reference to bleeding peptic ulcer. *BMJ* 1947;2(4525):477–82.
65 Marriott HL, Kekwick A. Continuous drip blood transfusion: (Section of Surgery). *Proc R Soc Med* 1936:29(4):337–8.
66 Lewin DC, Truelove S. Haematemesis. *BMJ* 1949;1(4600):383–6.
67 Meulengracht E. Medical treatment of peptic ulcer. *BMJ* 1939:2(4101):321–4.
68 Ward-McQuaid JN, Pease JC, Smith AM, Twort RJ. Surgery in bleeding peptic ulcers. *Gut* 1960;1(3):258–65.
69 Jones PF, Johnston SJ, McEwan AB, et al. Further haemorrhage after admission to hospital for gastrointestinal haemorrhage. *BMJ* 1973;3(5882):660–4.
70 Chandler GN, Cameron AD, Nunn AH, Street DF. Early investigations of haematemesis. *Gut* 1960;1(1):6–13.
71 Calvet X, Vergara M, Brullet E, et al. Addition of a second endoscopic treatment following epinephrine injection improves outcome in high-risk bleeding ulcers. *Gastroenterology* 2004;126(2):441–50.
72 Gonzalez R, Zamora J, Gomez-Camarero J, et al. Meta-analysis: combination endoscopic and drug therapy to prevent variceal rebleeding in cirrhosis. *Ann Intern Med* 2008;149(2):109–22.
73 Leontiadis GI, Sharma VK, Howden CW. Proton pump inhibitor treatment for acute peptic ulcer bleeding. *Cochrane Database Syst Rev* 2006;25(1):CD002094.
74 Chavez-Tapia NC, Barrientos-Gutierrez T, Tellez-Avila F, et al. Meta-analysis: antibiotic prophylaxis for cirrhotic patients with upper gastrointestinal bleeding – an updated Cochrane review. *Aliment Pharmacol Therap* 2011;34(5):509–18.
75 Ioannou G, Doust J, Rockey, DC. Terlipressin for acute esophageal variceal hemorrhage. *Cochrane Database Syst Rev (Online)* 2003;(1):CD002147.
76 Rockall TA, Logan RF, Devlin HB, Northfield TC. Incidence of and mortality from acute upper gastrointestinal haemorrhage in the United Kingdom. Steering Committee and members of the National Audit of Acute Upper Gastrointestinal Haemorrhage. *BMJ* 1995;311(6999):222–6.
77 Hearnshaw SA, Lowe D, Logan RF, et al. UK comparative audit of upper gastrointestinal bleeding and the use of blood. *BSG, RCP, Blood and Transplant,* 2007.
78 Rockall TA, Logan RF, Devlin HB, Northfield TC. Risk assessment after acute upper gastrointestinal haemorrhage. *Gut* 1996;38(3):316–21.
79 Blatchford O, Murray WR, Blatchford M. A risk score to predict need for treatment for upper-gastrointestinal haemorrhage. *Lancet* 2000;356(9238): 1318–21.
80 Longstreth GF, Feitelberg SP. Outpatient care of selected patients with acute non-variceal upper gastrointestinal haemorrhage. *Lancet* 1995;345(8942):108–11.
81 Packham CJ, Rockall TA, Logan RFA. Outpatient care for selected patients with acute upper gastrointestinal bleeding. *Lancet* 1995;345(8950):659–60.
82 Stanley AJ, Ashley D, Dalton HR, et al. Outpatient management of patients with low-risk upper-gastrointestinal haemorrhage: multicentre validation

and prospective evaluation. *Lancet* 2009;373(9657): 42–7.
83 Lanas A. Editorial: Upper GI bleeding-associated mortality: challenges to improving a resistant outcome. *Am J Gastroenterol* 2010;105(1):90–2.
84 Sung JJY, Tsoi KKF, Ma TKW, et al. Causes of mortality in patients with peptic ulcer bleeding: a prospective cohort study of 10,428 cases. *Am J Gastroenterol* 2009;105(1):84–9.
85 Nahon S, Pariente A, Nalet B, et al. Causes of mortality related to peptic ulcer bleeding in a prospective cohort of 965 French patients: a plea for primary prevention. *Am J Gastroenterol* 2010;105(8):1902–3.
86 Vreeburg EM, Snel P, de Bruijne JW, et al. Acute upper gastrointestinal bleeding in the Amsterdam area: incidence, diagnosis, and clinical outcome. *Am J Gastroenterol* 1997;92(2):236–43.
87 Hasselgren G, Blomqvist A, Eriksson S, et al. Short and long term course of elderly patients with peptic ulcer bleeding – analysis of factors influencing fatal outcome. *Eur J Surg* 1998;164(9):685–91.
88 Sung JJY, Lau JYW, Ching JYL, et al. Continuation of low-dose aspirin therapy in peptic ulcer bleeding. *Ann Intern Med* 2010;152(1):1.
89 Hudson N, Faulkner G, Smith SJ, et al. Late mortality in elderly patients surviving acute peptic ulcer bleeding. *Gut* 1995;37(2):177–81.
90 Kubba AK, Choudari C, Rajgopal C, et al. Reduced long-term survival following major peptic ulcer haemorrhage. *BJS* 1997;84(2):265–8.
91 Ruigomez A, García Rodríguez LA, Hasselgren G, et al. Overall mortality among patients surviving an episode of peptic ulcer bleeding. *J Epidemiol Community Health* 2000;54(2):130–3.
92 Myers RP, Kaplan GG, Shaheen AM. The effect of weekend versus weekday admission on outcomes of esophageal variceal hemorrhage. *Can J Gastroenterol*, 23(7):495–501, July 2009.
93 McCormick PA, O'Keefe C. Improving prognosis following a first variceal haemorrhage over four decades. *Gut* 2001;49(5):682–5.
94 de Franchis R. Evolving consensus in portal hypertension. Report of the Baveno IV consensus workshop on methodology of diagnosis and therapy in portal hypertension. *J Hepatol* 2005;43(1):167–76.
95 del Olmo JA, Pena A, Serra MA, et al. Predictors of morbidity and mortality after the first episode of upper gastrointestinal bleeding in liver cirrhosis. *J Hepatol* 2000;32(1):19–24.
96 Chen WT, Lin CY, Sheen IS, et al. MELD score can predict early mortality in patients with rebleeding after band ligation for variceal bleeding. *World J Gastroenterol* 2011;17(16):2120–5.
97 Bambha K, Kim WR, Pedersen R, et al. Predictors of early re-bleeding and mortality after acute variceal haemorrhage in patients with cirrhosis. *Gut* 2008;57(6):814–20.
98 Hearnshaw SA, Logan RFA, Lowe D, et al. Acute upper gastrointestinal bleeding in the UK: patient characteristics, diagnoses and outcomes in the 2007 UK audit. *Gut* 2011;60(10):1327–35.
99 Hobolth L, Krag A, Bendtsen F. The recent reduction in mortality from bleeding oesophageal varices is primarily observed from Days 1 to 5. *Liver Int* 2010;30(3):455–62.
100 Viviane A, Alan BN. Estimates of costs of hospital stay for variceal and nonvariceal upper gastrointestinal bleeding in the United States. *Value Health* 2008;11(1):1–3.
101 Targownik LE, Gralnek IM, Dulai GS, et al. Management of acute nonvariceal upper gastrointestinal hemorrhage: comparison of an American and a Canadian medical centre. *Can J Gastroenterol* 2003;17(8):489–95.
102 Gleeson F, Clarke E, Lennon J, et al. Outcome of accident and emergency room triaged patients with low risk non-variceal upper gastrointestinal haemorrhage. *Irish Med J* 2006;99(4):114–7.
103 Ndjitoyap Ndam EC, Koki Ndombo PO, Fouda OA, et al. [Upper digestive system hemorrhages in Cameroon (apropos of 172 cases examined via endoscopy).] *Méd Trop* 1990;50(2):181–4.
104 Zimmerman J, Meroz Y, Siguencia J, et al. Upper gastrointestinal hemorrhage. Comparison of the causes and prognosis in primary and secondary bleeders. *Scand J Gastroenterol* 1994;29(9):795–8.
105 Lule GN, Obiero ET, Ogutu EO. Factors that influence the short term outcome of upper gastrointestinal bleeding at Kenyatta National Hospital. *East Afr Med J* 1994;71(4):240–5.
106 Barkun A, Sabbah S, Enns R, et al. The Canadian Registry on Nonvariceal Upper Gastrointestinal Bleeding and Endoscopy (RUGBE): Endoscopic hemostasis and proton pump inhibition are associated with improved outcomes in a real-life setting. *Am J Gastroenterol* 2004;99(7):1238–46.
107 Kelly P, Katema M, Amadi B, et al. Gastrointestinal pathology in the University Teaching Hospital, Lusaka, Zambia: review of endoscopic and pathology records. *Trans R Soc Trop Med Hyg* 2008;102(2): 194–9.
108 Marmo R, Koch M, Cipolletta L, et al. Predictive factors of mortality from nonvariceal upper gastrointestinal hemorrhage: a multicenter study. *Am J Gastroenterol* 2008;103(7):1639–47; quiz 1648.

109 Djibril AM, Tomta K, Balaka K, et al. [Hematemesis in Togo: findings of a 12-month study in an intensive care unit.] *Méd Trop* 2010;70(3):311–2.
110 Paspatis GA, Matrella E, Kapsoritakis A, et al. An epidemiological study of acute upper gastrointestinal bleeding in Crete, Greece. *Eur J Gastroenterol Hepatol* 2000;12(11):1215–20.
111 Jurado Hernandez AM, de Teresa Galvan J, Ruiz-Cabello Jimenez M, Pinel Julian, LM. [Evolution in the epidemiology of non-variceal upper digestive hemorrhage from 1985 to 2006.] *Rev Esp Enferm Dig* 2008;100(5):273–7.
112 Crooks CJ, West J, Card TR. Upper gastrointestinal haemorrhage and deprivation: a nationwide cohort study of health inequality in hospital admissions. *Gut* 2012;61(4):514–20.
113 Czernichow P, Hochain P, Nousbaum JB, et al. Epidemiology and course of acute upper gastro-intestinal haemorrhage in four French geographical areas. *Eur J Gastroenterol Hepatol* 2000;12(2):175–81.
114 El-Serag HB, Everhart JE. Improved survival after variceal hemorrhage over an 11-year period in the Department of Veterans Affairs. *Am J Gastroenterol* 2000;95(12):3566–73.
115 Jamal MM, Samarasena JB, Hashemzadeh M. Decreasing in-hospital mortality for oesophageal variceal hemorrhage in the USA. *Eur J Gastroenterol Hepatol* 2008;20(10):947–55.
116 Ananthakrishnan AN, McGinley EL, Saeian K. Outcomes of weekend admissions for upper gastrointestinal hemorrhage: a nationwide analysis. *Clin Gastroenterol Hepatol* 2009;7(3):296–302e1.
117 Lecleire S, Di Fiore F, Merle V, et al. Acute upper gastrointestinal bleeding in patients with liver cirrhosis and in noncirrhotic patients: epidemiology and predictive factors of mortality in a prospective multicenter population-based study. *J Clin Gastroenterol* 2005;39(4):321–7.
118 Di Fiore F, Lecleire S, Merle V, et al. Changes in characteristics and outcome of acute upper gastrointestinal haemorrhage: a comparison of epidemiology and practices between 1996 and 2000 in a multicentre French study. *Eur J Gastroenterol Hepatol* 2005;17(6):641–7.

Answers to multiple choice questions

1. E
2. D
3. A,B
4. E
5. B

17 Epidemiology of celiac disease

Alberto Rubio-Tapia[1], *Jonas F. Ludvigsson*[2], *& Joseph A. Murray*[1]

[1]Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

[2]Senior Researcher in Epidemiology, Orebro University Hospital, Orebro, Sweden

Key points

- Celiac disease (CD) is a global health problem.
- At present, "nonclassical" or screen-detected cases of CD are the most frequent clinical presentations of the disease.
- Often greatly delayed diagnosis results in many years of symptoms that could have been improved by treatment of CD.
- Greater knowledge of CD epidemiology may aid in the early diagnosis of CD.

Clinical summary

Celiac disease (CD) is defined as a permanent intolerance to ingested gluten (the storage protein components of wheat, barley, and rye) that damages the small intestine by inducing crypt hyperplasia and villous atrophy; and which resolves or at least improves with removal of gluten from the diet [1]. CD results from the interaction between the environment (gluten intake) and genetic susceptibility, such as the presence of human leukocyte antigens (HLA) haplotypes DQ2 or DQ8, which drives an immune response in the gut. The inflammation and perpetuation of the immune-mediated process is induced by tissue transglutaminase-mediated gliadin deamidation (resulting in enhanced antigen presentation) and complex interactions between the enterocyte, cytokines, and inflammatory cells (principally T-cell lymphocytes).

Clinical features vary by type of presentation (classical vs. nonclassical), severity (mild vs. severe), and patient age at diagnosis (child vs. adult). The constellation of symptoms and signs includes steatorrhea, weight loss or failure to thrive, as well as less-specific gastrointestinal complaints – such as bloating, abdominal pain, diarrhea, constipation, flatulence, secondary lactose intolerance, and dyspepsia – and nongastrointestinal complaints – such as fatigue, depression, arthralgias, osteomalacia or osteoporosis, abnormal liver function tests, and iron-deficiency anemia. Furthermore, CD can be "asymptomatic". Thus, a high index of suspicion is necessary to establish the diagnosis in a wide variety of conditions and settings [2].

The detection of CD most often begins with serologic testing (detection of CD-specific antibodies) [3,4]. Confirmation of the disease requires demonstration of intestinal lesions on duodenal biopsies and ultimately a positive objective response to a gluten-free diet (GFD). In selected cases, HLA genotyping may provide adjunctive information, especially in patients who do not respond to a gluten-free diet or in patients where histologic or serologic determination has been rendered insensitive by prior treatment with a gluten-free diet.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

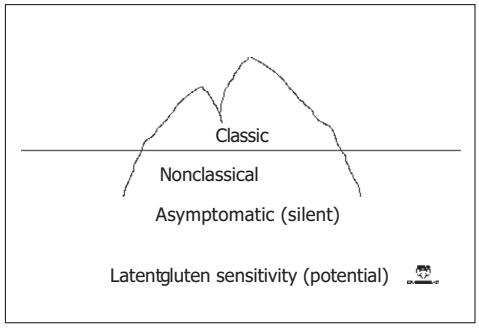


Figure 17.1 Concept of the celiac disease "iceberg."

The absence of the susceptibility-associated genotype has a high negative predictive value. Empiric treatment with a GFD is not recommended and renders most of the other tests inaccurate [5]. In most cases, a presumptive diagnosis can be made when serology and histology are both consistent with the ultimate proof occurring with the measurable objective response to a GFD.

Disease definition

Wide clinical presentation in CD has been likened to the iceberg model of disease (Figure 17.1). The tip of the iceberg represents the most obvious part of the clinical spectrum (classic malabsorption). If the patient's symptoms are characteristic of the malabsorption syndrome (diarrhea, steatorrhea, weight loss, fatigue) then the adjective "classical" is used. There is also "nonclassical" CD, these adjectives being applied when patients have nonspecific symptoms such as abdominal discomfort, bloating, indigestion or nongastrointestinal manifestations.

The definitely submerged portion of the iceberg consists of "silent" patients who are clinically asymptomatic (but have histologic evidence of CD if biopsied). Finally, there is an additional group of patients who are genetically susceptible to CD, but without symptoms or histologic evidence of CD. Some of these will ultimately go on to develop CD (latent CD) [5]. Such individuals are typically identified by repeat testing those with persistently positive autoantibodies, patients with dermatitis herpetiformis who initially have a normal small intestine biopsy, or asymptomatic family members of individuals with CD. Unfortunately, most cases with CD remain undiagnosed below the waterline.

Prevalence and incidence

CD is common in Western countries including North America. Data from the US population demonstrated an estimated prevalence of 0.8 % [6]. However, this figure may underestimate the true prevalence of CD due to lack of testing in all those with nonclassical or asymptomatic or latent CD. Indeed, more recent estimates using sequential serologic testing in a North American population suggest a prevalence of CD of 1 % in Caucasians [7,8]. The total prevalence (clinically diagnosed and unrecognized cases) of CD was 1.8 % in a population-based study in Sweden that used parallel serology and histopathology for detection of CD [9].The total prevalence of CD was 2 % in a Finland population in 2000–2001 [10]. The prevalence of CD in a select group of subjects born during a well-described Swedish celiac epidemic had risen from 1 % at the age of 2 years to slightly less than 3 % by 12 years of age [11].

Other reported prevalence studies (mainly based on serologic testing) indicate that CD could be a common disease around the world, suggesting that it may be a global health problem (Table 17.1).

The incidence of CD varies internationally. In several countries, there has been a significant increase in the overall incidence and, hence, prevalence of CD. However, this change has not been uniform across the age spectra. In the United Kingdom and the Republic of Ireland, although childhood CD reached epidemic proportions in the late 1960s and early 1970s, a substantial decrease in childhood CD was observed in the latter half of the 1970s. This decrease was ascribed to a public health campaign to delay the introduction of solids and to encourage breastfeeding in newborns. Sweden also had a dramatic increase in CD incidence in childhood through the 1980s (200–240 cases per 100,000 person-years) and into the 1990s followed by an equally abrupt decline in incidence of symptomatic CD (50–60 cases per 100,000 person-years) after 1995 [12]. This decrease was attributed to a change in public policy whereby the quantity of gluten in infant foodstuffs was reduced and a national recommendation was made to encourage breastfeeding during the period when gluten-containing foods are introduced

Table 17.1 Prevalence of CD in different countries based on selected serologic studies

Country	Type of study	Antibodies tested	Intestinal biopsy	Positive/tested	Prevalence/1000
Algeria [13]	PB	EMA	No	56/959	57
Tunisia[14]	BD	EMA, tTG	Yes	2/1418	1.4
USA [15]	BD	AGA, EMA	No	8/2000	4
England [16]	PB	EMA, tTG	No	87/7527	10
Brazil [17]	BD	AGA, EMA	Yes	3/2045	1.4
Italy [18]	PB	EMA	Yes	17/3483	5
Israel [19]	BD	AGA, tTG, EMA	Yes	10/1571	6.3
Argentina [20]	PB	AGA, EMA	Yes	12/2000	6
Finland [21]	PB	tTG, EMA	Yes	27/3654	7.3
The Netherlands [22]	BD	EMA	Yes	3/1000	3
Egypt [23]	PB	tTG, EMA	Yes	8/1500	5
Mexico [24]	BD	tTG	No	27/1009	27
India [25]	PB	tTG	Yes	14/4347	3
Iran [26]	PB	tTG, EMA	Yes	9/2799*	3

BD, blood donors; PB, population-based study; EMA, anti-endomysium antibodies; AGA, antigliadin antibodies; tTG, anti-tissue transglutaminase antibodies.
*Nine individuals had positive tTG or EMA and Marsh score II–III.

into the diet. In contrast, the overall annual incidence of CD in North America has shown a gradual increase. A study from Olmsted County showed that the incidence was 0.9 per 100,000 in 1950–1989, 3.3 per 100,000 in the 1990s, and 9.1 per 100,000 in 2000 and 2001 [27]. Serology prompted biopsy in a substantial proportion of recently diagnosed subjects suggesting that the increase in this population was due in part to an increased detection rate arising from increased physician awareness of the disorder and thus higher rates of screening for CD, although a true increase in incidence may have also occurred [27]. The prevalence of CD doubled in Finland over two decades and quadruplicated in the United States over a 50-year period [7,10]. These changes in prevalence cannot be attributed to better detection rate alone. This phenomenon could be present in other countries; for example, a national prospective study in the Netherlands showed a significant continual increase in reported incidence of CD (0.1–0.4/1000 live births from 1975 to 1990, to 0.81/1000 live births for 1993 to 2000) [28]. The reasons underlying the observed increasing incidence (and prevalence) of CD will require further study but are likely environmental and multifactorial. Thus higher incidence of CD is due to a true increase and along with increased testing for CD has driven the dramatic increase in case detection.

Risk factors for disease

Gender

Females predominate in clinically detected cases (by about 2:1) [29–32]. Interestingly, a high male preponderance (3:1) was found among the new cases of CD in members of nuclear families with two affected children [33], while the female-to-male ratio among patients with positive CD serology but normal small intestinal mucosa was 3:2 [34]. In population screening studies there is closer to parity. Additionally, in a US cohort of biopsy-proven adult patients with CD, men show indirect evidence of greater malabsorption than females and have female-predominant associated autoimmune diseases [35].

Geography

CD is present in every continent although the prevalence may vary across countries [36]. The extent of CD mirrors the coincidence of consumption of wheat and a high frequency of the genetic susceptibility genotype

DQ2. Ironically, wheat cultivation started in the fertile crescent (a historical region in the Middle East related to the origins of agriculture) wherein Caucasians also originated, and the areas of the world that have the highest prevalence of the CD susceptibility genotype have relied on wheat and similar grains as major staples to enable population growth and civilization.

Socioeconomic factors

There appears to be little association between CD and specific socioeconomic factors. CD may be associated with more severe nutritional consequences in developing countries, as exemplified by lower height-for-age and hemoglobin levels among CD-affected Saharawi children [37]. It has been suggested that socioeconomic factors may modify the risk of CD; specifically, lower economic status and consequent inferior hygiene environment may decrease the risk of CD [38]. However, extensive evidence of high prevalence of CD in several developing countries argues against the "hygiene hypothesis" as sole explanation for the observed differences in prevalence of CD among countries.

Familial aggregation/genetics

CD occurs commonly in families. In a population-based study, the prevalence of CD was 11 % among first-degree relatives of biopsy-proven CD cases, siblings had the greatest risk [39]. The risk is much higher in relatives of affected sibling pairs (17 %) [33,40], monozygotic twins (75 %), and HLA-identical siblings (40 %) [41] These clinical studies strongly support the role of genetics in CD pathogenesis.

The inheritance pattern is complex and determined by the effects of several genes and the environment. CD is strongly associated with the HLA class II genes *DQA1*05, DQB1*02* that encode the molecule DQ2, and less frequently *DQA1*0301, DQB1*0302* that encode DQ8. Such is the strength of the association that these HLA haplotypes are virtually essential for the disease to occur and they are a valuable tool for diagnosis in selected cases [42]. Furthermore, homozygosity for the *DQB1*0201* allele has been associated with a more severe form of CD characterized by total villous atrophy on small bowel biopsy, younger age of disease onset, more severe diarrhea, and a lower level of blood hemoglobin at the time of diagnosis [43]. This allele has also been associated with a slower recovery of villous atrophy after commencing a GFD [43]. Other non-HLA genes have been reported to be associated with CD and can improve identification of high-risk individuals [44].

However, despite the fact that the majority of family members will carry the at-risk HLA haplotype, far fewer of these will actually develop the disease. This indicates that genes other than HLA, or environmental factors have a major effect on causation of the disease in family members. This has important implications for HLA testing in suspected CD. Few of those with positive HLA have CD [45], and HLA-testing should only be used to rule out CD (high negative predictive value).

Other diseases

Several other diseases are associated with a high prevalence of CD (Table 17.2).

CD is strongly associated with type 1 diabetes mellitus [46–48], thyroid disease [49], Addison's disease [50], osteopenic bone disease [51–53], and Down syndrome [54]; but also with less common conditions such as autoimmune heart disease [55–57]. The prevalence of CD among osteoporotic individuals (3.4 %) is higher than that among nonosteoporotic individuals (0.2 %) [58]. Furthermore, female patients aged >50 years with CD demonstrated a higher risk of

Table 17.2 Associated conditions and consequences of CD (partial list)

Associated conditions	Consequences
Isolated hypertransaminasemia	Hyposplenism
Autoimmune thyroiditis	Arthralgia or arthropathy
Microscopic colitides	Ataxia
Autoimmune hepatitis	Dental enamel hypoplasia
IgA deficiency	Folate or iron deficiency anemia
Psoriasis	Recurrent pancreatitis
Primary biliary cirrhosis	Oral aphthous ulcers
Dermatitis herpetiformis	Lymphoma
Down syndrome	Osteoporosis
Type 1 diabetes mellitus	Bone fractures
Turner syndrome	Vitamin deficiencies

fracture [59]. Excess fracture risk in CD has been confirmed in population-based studies [51,60]. Thus, active investigation and management of bone disease is advisable in CD.

Natural history and mortality

The natural history of CD recognizes that, at certain points in time, the disease is not associated with clinical manifestations. There may be a long latent phase followed by a "silent" phase. At some point, intestinal and/or extraintestinal symptoms develop and the diagnosis is made by demonstrating the villous atrophy [61,62] and strongly positive anti-tissue transglutaminase (tTG) and anti-endomysial IgA autoantibodies[63]. Celiac disease is a chronic disease and one that will persist unless treated. Many patients may remain undiagnosed and the ultimate outcome in these individuals remains unknown.

The Denver studies have followed a birth cohort of individuals who had HLA typing performed at birth. Using tTG antibodies, this cohort was followed on a yearly basis up to the age of 7 years. One percent of these children, most of whom had the at-risk HLA haplotype, developed evidence of CD [64], but most of these had minimal or no symptoms. This is consistent with the theory that CD starts in the first decade, although the majority of patients are not diagnosed until later (Figure 17.2). There is emerging evidence that loss of tolerance may occur at any age by finding "de novo" adult-onset CD in individuals with a prior seronegative study [65,66]. In Finland, the prevalence of biopsy-proven CD increased from 2.13 % to 2.34 % within 3 years in individuals over 55 years of age, 5 new cases were found among previously seronegative subjects [65]. Nine previously seronegative subjects had evidence of CD 15 years after original testing for CD in an American population (CLUE cohort). The prevalence of CD increased twofold in the CLUE cohort over a 15-year period [66].

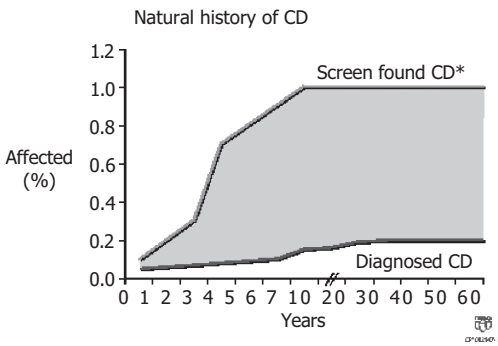


Figure 17.2 Natural history of CD.

In 2009, a Swedish cohort study of some 46,000 patients undergoing small intestinal biopsy with various degrees of histopathologic changes, showed an increased risk of death in CD [67]. This risk was similar in patients with Marsh 3 and individuals with Marsh 0 but positive CD-related antibodies (+30–40 % increased risk of death in both groups), while excess mortality was more than 70 % in individuals with Marsh 2 (inflammation) [67]. One potential reason for the higher relative risk of death in patients with Marsh 2 rather than Marsh 3 may be that patients with Marsh 2 have traditionally not been assigned a gluten-free diet.

The absolute mortality rate was 10.4 (95 % CI 10.0–10.8) per 1000 person-years in CD with an excess mortality of 2.9 per 1000 person-years. Importantly, the excess risk decreased with follow-up after biopsy, but nevertheless remained almost 30 % increased even 5 years after CD diagnosis [67]. The most common causes of death in patients with diagnosed CD were cardiovascular disease and cancer [67].

Enteropathy-associated T-cell lymphoma (EATL) is a rare form of high-grade, T-cell non-Hodgkin lymphoma of the upper small intestine that is specifically associated with CD [68–70]. EATL is a rare neoplasm and a wide variety of other most common histologic subtypes of lymphoma have been associated to CD [71]. Refractory celiac disease, a rare but severe complication of CD characterized by persistent or recurrent symptoms despite strict adherence to GFD, is associated with a greatly increased mortality risk especially when a T-cell clone is present in the intestine (refractory celiac disease Type II) [72]. Simultaneously the lower body mass index in patients with CD, may protect against breast cancer in women with CD [31,73,74].

Studies of mortality in undetected CD (hence untreated) compared with the general population give contradictory results, suggesting either increased all-cause mortality risk [7,75] or similar risk compared with the general population [8,76–78]. The possibility

of increased mortality risk of undetected CD affecting young adults is worrisome [7].

Excess mortality may be especially high in patients with malabsorption syndromes [79]. It is unclear if patients with minor symptoms or identified through antibody screening are at increased risk of dying than the general population [79].

Both Swedish and British population-based data study demonstrate that the increased risk of malignancy (other than lymphoproliferative disease) is primarily in the first year after diagnosis [31,80]. Adenocarcinoma of the small intestine, nasopharyngeal, melanoma and esophageal cancer are also more common in patients with CD than the general population [81,82]. The risk of most GI cancers are only increased in the first year after CD diagnosis, and it cannot be ruled out that this increase is due to ascertainment bias [82].

Fatal pneumococcal septicemia has been reported in celiac patients with hyposplenism, and prophylactic vaccination may be appropriate in this clinical scenario [83]. Also tuberculosis [84] and influenza [85] may be more common in patients with CD.

Disability and quality of life

As a chronic condition, symptomatic CD impairs health-related quality of life. Before treatment onset, 63 % of patients report that their perceived quality of life is less than good ("bad" or "fair") [86]. Several studies have shown that quality of life improves after treatment with a GFD [87,88]. Longstanding clinical experience reports that a GFD will generally result in a dramatic improvement in what are often severe GI symptoms, including symptoms other than the typical ones of diarrhea, steatorrhea, and weight loss [89]. It would be expected that such improvement in symptoms could result in improved quality of life. However, because (i) the disease is chronic and (ii) the treatment is lifelong, the rather restrictive GFD may have negative effects on quality of life. Careful studies, incorporating patients who have minimal or no GI symptoms at the onset of treatment, need to be performed to identify the degree of ultimate benefit for overall quality of life of the early detection and treatment of asymptomatic CD.

Adherence to a GFD for at least 1 year causes 82 % of classic CD patients to consider that they reached a "well" or "very well" feeling of wellbeing as assessed by a modified version of the Zung Self-Related Depression Scale [87]. A Finnish study demonstrated that after 1 year of following a GFD, quality of life for patients with screen-detected CD significantly improved as measured by a generic quality of life questionnaire (Psychological General Well-Being Questionnaire) [88]. In a Spanish population, it was found that CD impaired the perceived health of affected individuals, and that their health improved when on a GFD reaching levels comparable with the general population, as assessed by administering two generic health-related quality of life questionnaires: EuroQol-5D and Gastrointestinal Quality of Life [90]. Another study found that, using the SF-36, adherence to a GFD causes patients to perceive a health-related quality of life comparable to that of the general population [91]. Patients with untreated CD also suffer from fatigue [92], and this will affect their quality of life negatively. It cannot be ruled out that the lower quality of life contributes to the small excess risk of suicide in CD [93].

It must be noted that no CD-specific quality-of-life measure exists to date, and that this remains an open area for development and research.

Economics of celiac disease

While it is self-evident that long delays of the diagnosis of CD may result in several years of suffering for symptomatic patients and that following a GFD can be expensive, there has been very little to address the health-related cost of care in CD. Two studies have shown a reduction in cost of health care that results from the diagnosis of CD [94,95]. Healthcare costs were greater in the years before the diagnosis of CD as compared to age- and gender-matched controls in the same community [95].

Prevention

Protective factors, such as breastfeeding and delayed introduction of large amounts of gluten into the infant diet, seem to reduce the likelihood of developing CD at an early age. In Sweden, the prevalence of symptomatic CD (clinically detected) declined after a national change in infant feeding recommendations

was proposed in 1996: a slow introduction to gluten during weaning was stressed, and the recommendation was to maintain breast-feeding during the period when smaller quantities of gluten are introduced into the diet, beginning at the age of 4 months instead of the introduction of larger amounts of gluten at the age of 6 months. However, no difference was found in undiagnosed CD between the screened children born before and after 1996. Thus, a slow introduction of gluten in infancy could protect some children against developing symptomatic CD, but it may not protect them from subclinical or silent forms of this disease in childhood [96]. At the same time another Swedish study prospectively collecting dietary data on breast-feeding duration found no association with future childhood CD [97]. A large European CD research study (PreventCD) is underway to find new strategies for the prevention of CD [98].

Serologic testing in high-risk populations, such as type 1 diabetics, could be a good approach for early detection of the disease [99]. However, the optimal screening and management strategy for CD in children with type 1 diabetes mellitus is unknown [100]. Although screening for CD was widely accepted in a general population from Wyoming during an annual health fair [101], the benefits and limitations of a mass screening program and treatment of asymptomatic patients require further study. Mass screening would need to be carefully evaluated in terms of advantages, caveats, risk, and cost before it is ready for introduction into routine practice.

Issues and gaps in epidemiology

There are several notable issues and unresolved questions in the epidemiology of CD. These include:

- What are the benefits and drawbacks of screening high-risk populations for CD? Are screen-detected individuals at the same risk of complications as patients detected through symptoms?
- Which populations are not affected by CD (Asians etc.)? Who do we not need to test?!
- What is the clinical significance of CD detected by mass screening?
- Why is the prevalence of undiagnosed CD changing in the developed world and the developing world?
- What are the most appropriate CD-specific quality-of-life measures?
- Are there benefits for an individual with undiagnosed celiac disease? (Patients with *diagnosed* CD seem to be at a lower risk of breast cancer.)
- What environmental factors other than infant feeding influence the development of CD?
- Does CD always develop in infant age (then remaining undiagnosed until testing), or can it develop in later age?

Recommendations for future studies

Much remains unknown about the natural history of undiagnosed and, hence, untreated CD. Studies that could address this topic, possibly examining historical cohorts from patients with silent CD who were not treated or patients in whom the diagnosis could be established retrospectively in stored sera, may give some helpful insights into the natural history of untreated CD. Inherent in this evaluation of outcome of untreated disease is the possibility that in some circumstances, such as societies where excess calories are a major cause of morbidity, it may be reasonable to consider the potential positive effects of subtle malabsorption on cardiovascular risk, especially because research in this area is contradictory (increased risk of cardiovascular disease [67,102,103]; lower risk of cardiovascular disease [104]). It is imperative to understand the benefits and limitations of mass screening for CD because most patients with CD remain undetected for years using the current diagnostic strategies (e.g. case finding).

Conclusions

Over the last decade, the epidemiology of CD has changed because of the emergence of a new generation of serologic tools, permitting a better understanding of the true incidence and prevalence of CD around the world. It has also become apparent that CD can have many faces on presentation, so clinicians must be familiar with all of them in order to obtain an early diagnosis of the disease before complications appear or become irreversible. As a global health problem that may affect millions around the world, CD warrants additional study to allow better detection of disease, as well as better prevention of the development of disease and its complications.

Multiple choice questions

1 What is the prevalence of celiac disease in Western Europe and North America?

A 1

B 10 %

C 0.1 %

2 Which of the following is likely to increase the risk of celiac disease?

A High alcohol consumption

B Poor socioeconomic status

C Short breastfeeding

3 Which of the following statements are correct?

A Genetics play an important role suggesting that first-degree relatives should be tested for celiac disease

B All celiac patients have diarrhea at onset

C IgE against oats is a fundamental part of celiac disease investigation

References

1 Farrell RJ, Kelly CP. Celiac sprue. *New Engl J Med* 2002;346(3):180–8.

2 Green PH, Jabri B. Celiac disease. *Annu Rev Med* 2006;57:207–21.

3 Dieterich W, Ehnis T, Bauer M, et al. Identification of tissue transglutaminase as the autoantigen of celiac disease [see comments]. *Nat Med* 1997;3(7):797–801.

4 Sulkanen S, Halttunen T, Laurila K, et al. Tissue transglutaminase autoantibody enzyme-linked immunosorbent assay in detecting celiac disease [see comments]. *Gastroenterology* 1998;115(6):1322–8.

5 Rostom A, Murray JA, Kagnoff MF. American Gastroenterological Association (AGA) Institute technical review on the diagnosis and management of celiac disease. *Gastroenterology* 2006;131(6):1981–2002.

6 Fasano A, Berti I, Gerarduzzi T, et al. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Arch Intern Med* 2003;163(3):286–92.

7 Rubio-Tapia A, Kyle RA, Kaplan EL, et al. Increased prevalence and mortality in undiagnosed celiac disease. *Gastroenterology* 2009;137(1):88–93.

8 Godfrey JD, Brantner TL, Brinjikji W, et al. Morbidity and mortality among older individuals with undiagnosed celiac disease. *Gastroenterology* 2010;139(3):763–9.

9 Walker MM, Murray JA, Ronkainen J, et al. Detection of celiac disease and lymphocytic enteropathy by parallel serology and histopathology in a population-based study. *Gastroenterology* 2010;139(1):112–9.

10 Lohi S, Mustalahti K, Kaukinen K, et al. Increasing prevalence of coeliac disease over time. *Aliment Pharmacol Ther* 2007;26(9):1217–25.

11 Myleus A, Ivarsson A, Webb C, et al. Celiac disease revealed in 3 % of Swedish 12-year-olds born during an epidemic. *J Pediatr Gastroenterol Nutr* 2009;49(2):170–6.

12 Ivarsson A, Persson LA, Nystrom L, et al. Epidemic of coeliac disease in Swedish children [see comments]. *Acta Paediatr* 2000;89(2):165–71.

13 Catassi C, Ratsch IM, Gandolfi L, et al. Why is coeliac disease endemic in the people of the Sahara? *Lancet* 1999;354(9179):647–8.

14 Bdioui F, Sakly N, Hassine M, Saffar H. Prevalence of celiac disease in Tunisian blood donors. *Gastroenterol Clin Biol* 2006;30(1):33–6.

15 Not T, Horvath K, Hill ID, et al. Celiac disease risk in the USA: high prevalence of antiendomysium antibodies in healthy blood donors. *Scand J Gastroenterol* 1998;33(5):494–8.

16 West J, Logan RF, Hill PG, et al. Seroprevalence, correlates, and characteristics of undetected coeliac disease in England. *Gut* 2003;52(7):960–5.

17 Gandolfi L, Pratesi R, Cordoba JC, et al. Prevalence of celiac disease among blood donors in Brazil. *Am J Gastroenterol* 2000;95(3):689–92.

18 Volta U, Bellentani S, Bianchi FB, et al. High prevalence of celiac disease in Italian general population. *Dig Dis Sci* 2001;46(7):1500–5.

19 Shamir R, Lerner A, Shinar E, et al. The use of a single serological marker underestimates the prevalence of celiac disease in Israel: a study of blood donors. *Am J Gastroenterol* 2002;97(10):2589–94.

20 Gomez JC, Selvaggio GS, Viola M, et al. Prevalence of celiac disease in Argentina: screening of an adult population in the La Plata area. *Am J Gastroenterol* 2001;96(9):2700–4.

21 Maki M, Mustalahti K, Kokkonen J, et al. Prevalence of celiac disease among children in Finland. *New Engl J Med* 2003;348(25):2517–24.

22 Rostami K, Mulder CJ, Werre JM, et al. High prevalence of celiac disease in apparently healthy blood donors suggests a high prevalence of undiagnosed celiac disease in the Dutch population. *Scand J Gastroenterol* 1999;34(3):276–9.

23 Abu-Zekry M, Kryszak D, Diab M, et al. Prevalence of celiac disease in Egyptian children disputes the east-west

agriculture-dependent spread of the disease. *J Pediatr Gastroenterol Nutr* 2008;47(2):136–40.
24 Remes-Troche JM, Ramirez-Iglesias MT, Rubio-Tapia A, et al. Celiac disease could be a frequent disease in Mexico: prevalence of tissue transglutaminase antibody in healthy blood donors. *J Clin Gastroenterol* 2006;40(8):697–700.
25 Sood A, Midha V, Sood N, et al. Prevalence of celiac disease among school children in Punjab, North India. *J Gastroenterol Hepatol* 2006;21(10): 1622–5.
26 Akbari MR, Mohammadkhani A, Fakheri H, et al. Screening of the adult population in Iran for coeliac disease: comparison of the tissue-transglutaminase antibody and anti-endomysial antibody tests. *Eur J Gastroenterol Hepatol* 2006;18(11):1181–6.
27 Murray JA, Van Dyke C, Plevak MF, et al. Trends in the identification and clinical features of celiac disease in a North American community, 1950–2001. *Clin Gastroenterol Hepatol* 2003;1(1):19–27.
28 Steens RF, Csizmadia CG, George EK, et al. A national prospective study on childhood celiac disease in the Netherlands 1993–2000: an increasing recognition and a changing clinical picture. *J Pediatr* 2005;147(2): 239–43.
29 Ciacci C, Cirillo M, Sollazzo R, et al. Gender and clinical presentation in adult celiac disease. *Scand J Gastroenterol* 1995;30(11):1077–81.
30 Ivarsson A, Hernell O, Nystrom L, Persson LA. Children born in the summer have increased risk for coeliac disease. *J Epidemiol Community Health* 2003; 57(1):36–9.
31 West J, Logan RF, Smith CJ, et al. Malignancy and mortality in people with coeliac disease: population-based cohort study. *BMJ* 2004;329(7468):716–9.
32 Ludvigsson JF, Brandt L, Montgomery SM, et al. Validation study of villous atrophy and small intestinal inflammation in Swedish biopsy registers. *BMC Gastroenterol* 2009;9(1):19.
33 Gudjonsdottir AH, Nilsson S, Ek J, et al. The risk of celiac disease in 107 families with at least two affected siblings. *J Pediatr Gastroenterol Nutr* 2004;38(3): 338–42.
34 Ludvigsson JF, Brandt L, Montgomery SM. Symptoms and signs in individuals with serology positive for celiac disease but normal mucosa. *BMC Gastroenterol* 2009;9:57.
35 Bai D, Brar P, Holleran S, et al. Effect of gender on the manifestations of celiac disease: evidence for greater malabsorption in men. *Scand J Gastroenterol* 2005;40(2):183–7.
36 Mustalahti K, Catassi C, Reunanen A, et al. The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Ann Med* 2010;42(8):587–95.
37 Ratsch IM, Catassi C. Coeliac disease: a potentially treatable health problem of Saharawi refugee children. *Bull World Health Organ* 2001;79(6):541–5.
38 Kondrashova A, Mustalahti K, Kaukinen K, et al. Lower economic status and inferior hygienic environment may protect against celiac disease. *Ann Med* 2008;40(3):223–31.
39 Rubio-Tapia A, Van Dyke CT, Lahr BD, et al. Predictors of family risk for celiac disease: a population-based study. *Clin Gastroenterol Hepatol* 2008;6(9):983–7.
40 Book L, Hart A, Black J, et al. Prevalence and clinical characteristics of celiac disease in Down's syndrome in a US study. *Am J Med Genet* 2001;98(1):70–4.
41 Greco L, Romino R, Coto I, et al. The first large population-based twin study of coeliac disease. *Gut* 2002;50(5):624–8.
42 van Heel DA, Hunt K, Greco L, Wijmenga C. Genetics in coeliac disease. *Best Pract Res Clin Gastroenterol* 2005;19(3):323–39.
43 Karinen H, Karkkainen P, Pihlajamaki J, et al. Gene dose effect of the DQB1*0201 allele contributes to severity of coeliac disease. *Scand J Gastroenterol* 2006;41(2):191–9.
44 Romanos J, van Diemen CC, Nolte IM, et al. Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology* 2009;137(3):834–40, 840.e1–3.
45 Hadithi M, von Blomberg BM, Crusius JB, et al. Accuracy of serologic tests and HLA-DQ typing for diagnosing celiac disease. *Ann Intern Med* 2007;147(5):294–302.
46 Cronin CC, Shanahan F. Insulin-dependent diabetes mellitus and coeliac disease. *Lancet* 1997;349(9058): 1096–7.
47 Bao F, Yu L, Babu S, et al. One third of HLA DQ2 homozygous patients with type 1 diabetes express celiac disease-associated transglutaminase autoantibodies. *J Autoimmun* 1999;13(1):143–8.
48 Ludvigsson JF, Ludvigsson J, Ekbom A, Montgomery SM. Celiac disease and risk of subsequent type 1 diabetes: A general population cohort study of children and adolescents. *Diabetes Care* 2006;29(11):2483–8.
49 Elfstrom P, Montgomery SM, Kampe O, et al. Risk of thyroid disease in individuals with celiac disease. *J Clin Endocrinol Metab* 2008;93(10):3915–21.
50 Elfstrom P, Montgomery SM, Kampe O, et al. Risk of primary adrenal insufficiency in patients with celiac disease. *J Clin Endocrinol Metab* 2007;92(9):3595–8.
51 West J, Logan RF, Card TR, et al. Fracture risk in people with celiac disease: a population-based cohort study. *Gastroenterology* 2003;125(2):429–36.

52 Ludvigsson JF, Michaelsson K, Ekbom A, Montgomery SM. Coeliac disease and the risk of fractures: a general population-based cohort study. *Aliment Pharmacol Ther* 2007;25(3):273–85.
53 Thomason K, West J, Logan RF, et al. Fracture experience of patients with coeliac disease: a population based survey. *Gut* 2003;52(4):518–22.
54 George EK, Mearin ML, Bouquet J, et al. High frequency of celiac disease in Down syndrome. *J Pediatr* 1996;128(4):555–7.
55 Frustaci A, Cuoco L, Chimenti C, et al. Celiac disease associated with autoimmune myocarditis. *Circulation* 2002;105(22):2611–8.
56 Fonager K, Sorensen HT, Norgard B, Thulstrup AM. Cardiomyopathy in Danish patients with coeliac disease. *Lancet* 1999;354(9189):1561.
57 Elfstrom P, Hamsten A, Montgomery SM, et al. Cardiomyopathy, pericarditis and myocarditis in a population-based cohort of inpatients with coeliac disease. *J Intern Med* 2007;262(5):545–54.
58 Stenson WF, Newberry R, Lorenz R, et al. Increased prevalence of celiac disease and need for routine screening among patients with osteoporosis. *Arch Intern Med* 2005;165(4):393–9.
59 Davie MW, Gaywood I, George E, et al. Excess non-spine fractures in women over 50 years with celiac disease: a cross-sectional, questionnaire-based study. *Osteoporos Int* 2005;16(9):1150–5.
60 Jafri MR, Nordstrom CW, Murray JA, et al. Long-term fracture risk in patients with celiac disease: A population-based study in Olmsted County, Minnesota. *Dig Dis Sci* 2007;53(4):964–71.
61 Dickson BC, Streutker CJ, Chetty R. Coeliac disease: an update for pathologists. *J Clin Pathol* 2006;59(10):1008–16.
62 Walker MM, Murray JA. An update in the diagnosis of coeliac disease. *Histopathology* 2010.
63 Rewers M. Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? *Gastroenterology* 2005;128(4 Suppl 1):S47–51.
64 Hoffenberg EJ, MacKenzie T, Barriga KJ, et al. A prospective study of the incidence of childhood celiac disease. *J Pediatr* 2003;143(3):308–14.
65 Vilppula A, Kaukinen K, Luostarinen L, et al. Increasing prevalence and high incidence of celiac disease in elderly people: a population-based study. *BMC Gastroenterol* 2009;9:49.
66 Catassi C, Kryszak D, Bhatti B, et al. Natural history of celiac disease autoimmunity in a USA cohort followed since 1974. *Ann Med* 2010;42(7):530–8.
67 Ludvigsson JF, Montgomery SM, Ekbom A, et al. Small-intestinal histopathology and mortality risk in celiac disease. *JAMA* 2009;302(11):1171–8.
68 Viljamaa M, Kaukinen K, Pukkala E, et al. Malignancies and mortality in patients with coeliac disease and dermatitis herpetiformis: 30-year population-based study. *Dig Liver Dis* 2006;38(6):374–80.
69 Catassi C, Bearzi I, Holmes GK. Association of celiac disease and intestinal lymphomas and other cancers. *Gastroenterology* 2005;128(4 Suppl 1):S79–86.
70 Mearin ML, Catassi C, Brousse N, et al. European multi-centre study on coeliac disease and non-Hodgkin lymphoma. *Eur J Gastroenterol Hepatol* 2006;18(2):187–94.
71 Halfdanarson TR, Rubio-Tapia A, Ristow KM, et al. Patients with celiac disease and B-cell lymphoma have a better prognosis than those with T-cell lymphoma. *Clin Gastroenterol Hepatol* 2010;8(12):1042–7.
72 Rubio-Tapia A, Murray JA. Classification and management of refractory coeliac disease. *Gut* 2010;59(4):547–57.
73 Askling J, Linet M, Gridley G, et al. Cancer incidence in a population-based cohort of individuals hospitalized with celiac disease or dermatitis herpetiformis. *Gastroenterology* 2002;123(5):1428–35.
74 Ludvigsson JF, West J, Ekbom A, Stephansson O. Reduced risk of breast, endometrial, and ovarian cancer in women with celiac disease. *Int J Cancer* 2011 (accepted for publication).
75 Metzger MH, Heier M, Maki M, et al. Mortality excess in individuals with elevated IgA anti-transglutaminase antibodies: the KORA/MONICA Augsburg cohort study 1989–1998. *Eur J Epidemiol* 2006;21(5):359–65.
76 Johnston SD, Watson RG, McMillan SA, et al. Coeliac disease detected by screening is not silent – simply unrecognized. *QJM* 1998;91(12):853–60.
77 Lohi S, Maki M, Rissanen H, et al. Prognosis of unrecognized coeliac disease as regards mortality: a population-based cohort study. *Ann Med* 2009;41(7):508–15.
78 Canavan C, Logan RF, Khaw KT, West J. No difference in mortality in undetected coeliac disease compared with the general population: a UK cohort study. *Aliment Pharmacol Ther* 2011.
79 Corrao G, Corazza GR, Bagnardi V, et al. Mortality in patients with coeliac disease and their relatives: a cohort study. *Lancet* 2001;358(9279):356–61.
80 Elfstrom P, Granath F, Ekstrom Smedby K, et al. Risk of lymphoproliferative malignancy in relation to small intestinal histopathology among patients with celiac disease. *J Natl Cancer Inst* 2011;103(5):436–44.
81 Green PH, Fleischauer AT, Bhagat G, et al. Risk of malignancy in patients with celiac disease. *Am J Med* 2003;115(3):191–5.
82 Elfstrom P, Granath F, Ye W, Ludvigsson JF. Low risk of gastrointestinal cancer among patients with celiac

disease, inflammation, or latent celiac disease. *Clin Gastroenterol Hepatol* 2011.
83 Johnston SD, Robinson J. Fatal pneumococcal septicaemia in a coeliac patient. *Eur J Gastroenterol Hepatol* 1998;10(4):353–4.
84 Ludvigsson JF, Sanders DS, Maeurer M, et al. Risk of tuberculosis in a large sample of patients with coeliac disease: a nationwide cohort study. *Aliment Pharmacol Ther* 2011;33(6):689–96.
85 Marild K, Fredlund H, Ludvigsson JF. Increased risk of hospital admission for influenza in patients with celiac disease: a nationwide cohort study in Sweden. *Am J Gastroenterol* 2010;105(11):2465–73.
86 Green PHR, Stavropoulos SN, Panagi SG, et al. Characteristics of adult celiac disease in the USA: results of a national survey. *Am J Gastroenterol* 2001;96(1): 126–31.
87 Ciacci C, D'Agate C, De Rosa A, et al. Self-rated quality of life in celiac disease. *Dig Dis Sci* 2003;48(11): 2216–20.
88 Mustalahti K, Lohiniemi S, Collin P, et al. Gluten-free diet and quality of life in patients with screen-detected celiac disease. *Eff Clin Pract* 2002;5(3):105–13.
89 Murray JA, Watson T, Clearman B, Mitros F. Effect of a gluten-free diet on gastrointestinal symptoms in celiac disease. *Am J Clin Nutr* 2004;79(4):669–73.
90 Casellas F, Lopez Vivancos J, Malagelada JR. Perceived health status in celiac disease. *Rev Esp Enferm Dig* 2005;97(11):794–804.
91 O'Leary C, Wieneke P, Healy M, et al. Celiac disease and the transition from childhood to adulthood: a 28-year follow-up. *Am J Gastroenterol* 2004;99(12): 2437–41.
92 Siniscalchi M, Iovino P, Tortora R, et al. Fatigue in adult coeliac disease. *Aliment Pharmacol Ther* 2005; 22(5):489–94.
93 Ludvigsson JF, Sellgren C, Runeson B, et al. Increased suicide risk in coeliac disease: a Swedish nationwide cohort study. *Dig Liver Dis* 2011;43(8):616–22.
94 Green PH, Neugut AI, Naiyer AJ, et al. Economic benefits of increased diagnosis of celiac disease in a national managed care population in the United States. *J Insur Med* 2008;40(3–4):218–28.
95 Long KH, Rubio-Tapia A, Wagie AE, et al. The economics of coeliac disease: a population-based study. *Aliment Pharmacol Ther* 2010;32(2):261–9.
96 Carlsson A, Agardh D, Borulf S, et al. Prevalence of celiac disease: before and after a national change in feeding recommendations. *Scand J Gastroenterol* 2006; 41(5):553–8.
97 Welander A, Tjernberg AR, Montgomery SM, et al. Infectious disease and risk of later celiac disease in childhood. *Pediatrics* 2010;125(3):e530–6.
98 Hogen Esch CE, Rosen A, Auricchio R, et al. The PreventCD Study design: towards new strategies for the prevention of coeliac disease. *Eur J Gastroenterol Hepatol* 2010;22(12):1424–30.
99 Catassi C, Kryszak D, Louis-Jacques O, et al. Detection of celiac disease in primary care: a multicenter case-finding study in North America. *Am J Gastroenterol* 2007;102(7):1454–60.
100 Simmons JH, Klingensmith GJ, McFann K, et al. Celiac autoimmunity in children with type 1 diabetes: a two-year follow-up. *J Pediatr* 2011;158(2):276–81, e1.
101 Katz KD, Rashtak S, Lahr BD, et al. Screening for celiac disease in a North American population: sequential serology and gastrointestinal symptoms. *Am J Gastroenterol* 2011;106(7):1333–9.
102 Wei L, Spiers E, Reynolds N, et al. Association between coeliac disease and cardiovascular disease. *Aliment Pharmacol Ther* 2007;27(6):514–9.
103 Ludvigsson JF, James S, Askling J, et al. Nationwide cohort study of risk of ischemic heart disease in patients with celiac disease. *Circulation* 2011;123(5):483–90.
104 West J, Logan RF, Card TR, et al. Risk of vascular disease in adults with diagnosed coeliac disease: a population-based study. *Aliment Pharmacol Ther* 2004; 20(1):73–9.

Answers to multiple choice questions

1. A
2. C
3. A

18 Measuring utilization of endoscopy in clinical practice

Frances Tse[1] *& Alan Barkun*[2]

[1]Division of Gastroenterology, McMaster University, McMaster University Medical Centre, Hamilton, ON, Canada

[2]Division of Gastroenterology, Montreal General Hospital Site, The McGill University, Health Centre, Montreal, QC, Canada

Key points

- Research using administrative databases, clinical registries, and electronic health record (EHR) databases can complement randomized controlled trials in providing real-life outcome data for enhancing quality.
- While administrative databases play a central role in epidemiologic evaluation because of their wide availability, low cost, easy accessibility of data, and ability to measure large samples of the population, their usefulness for measuring quality is limited.
- Clinical registries are designed for both research and quality improvement, but significant investment is required for data collection and quality control.
- EHR databases can collect all the data needed for outcome research and quality assessment, but are subject to the same caveats as using any large databases.
- Given the various limitations, these databases should only be viewed as hypothesis-generating tools.

Introduction

Gastrointestinal (GI) endoscopy has become a widely available and routine technique for the screening, diagnosis, and management of a variety of GI conditions. The demand for colonoscopy has greatly increased in North America and worldwide over the past decade, largely in response to national colorectal cancer (CRC) screening programs. In the United States, an estimated 17 million lower GI procedures (colonoscopy and flexible sigmoidoscopy) are performed annually [1]. The safety and effectiveness of colonoscopy depend critically on the quality of examination, and there is increasing evidence to suggest that the quality of colonoscopy varies. For example, cecal intubation rates vary substantially from 59 % to 98 % [2]. Miss rates were estimated to be 2–6 % for CRC [2–5], and 2–26 % for polyps [4,6–8]. Despite having evidence-based guidelines, colonoscopies are frequently performed earlier than suggested [9–15]. These wide variations in practice underscore the importance of improving the quality of colonoscopy for enhancing the safety and cost-effectiveness of the examination.

In 2002, the US Multi-Society Task Force on Colorectal Cancer (MSTF-CRC) published

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

recommendations to improve the quality of colonoscopy [16] with key quality indicators for use in the continuous quality improvement (CQI) process for colonoscopy [17–19]. Suggested quality indicators include preprocedural, intraprocedural, and postprocedural metrics (Table 18.1). These indicators, however, should be considered as surrogate markers that may correlate with the only true clinically relevant endpoint in colonoscopy screening – CRC incidence and mortality. Nevertheless, accurate, complete, and standardized reporting of all these quality indicators within and across practices is essential for the CQI process of colonoscopy and benchmarking of quality-of-care outcomes. To facilitate measuring and monitoring of these quality indicators, the Quality Assurance Task Force of the National Colorectal Cancer Roundtable (NCCRT) developed a standardized Colonoscopy Reporting and Data System (CO-RADS) [20]. The primary goals of this reporting system were twofold: (i) to improve communication of test results and coordination of care by ensuring appropriate documentation of endoscopic findings and recommendations; and (ii) to provide endoscopists with a CQI tool to routinely monitor quality indicators in their practice and benchmark their performance against aggregate data from other physicians or groups. Quality improvement measures can then be instituted using these data.

Table 18.1 Colonoscopy quality indicators (adapted from [17–19])

1 Appropriate indication
2 Informed consent
3 History and physical examination recorded
4 Risk stratification employed and documented
5 Sedation plan
6 Anticoagulants recorded
7 Team pre-procedure pause
8 Patient monitoring employed and documented
9 Medications are documented
10 Reversal agent used
11 Use of recommended postpolypectomy and post-cancer resection surveillance intervals
12 Use of recommended ulcerative colitis/Crohn's disease surveillance intervals
13 Documentation of bowel preparation quality
14 Cecal intubation rates (documentation of cecal landmarks with photo documentation)
15 Adenoma detection rate in asymptomatic individuals
16 Withdrawal time: mean withdrawal time should be >6 min in colonoscopies with normal results
17 Biopsy specimens obtained in patients with chronic diarrhea
18 Appropriate number and distribution of biopsy samples in ulcerative colitis and Crohn's colitis surveillance. approximately 32 specimens per case of pancolitis
19 Mucosally based pedunculated polyps and sessile polyps <2 cm in size are endoscopically resected or documentation of unresectability is provided
20 Discharge criteria employed
21 Written procedure specific discharge instructions are provided
22 Plan for post-procedure resumption of anticoagulants provided
23 Procedure report complete
24 Pathology reviewed and communicated
25 Communication with referring provider(s)
26 Incidence of perforation by procedure type
27 Incidence of postpolypectomy bleeding
28 Postpolypectomy bleeding managed nonoperatively
29 Patient satisfaction surveyed

With increasing emphasis placed on measurement of quality and patient-centered outcomes for medical procedures including endoscopy, there is a growing interest in collecting real-life effectiveness data. While randomized controlled trials (RCTs) are the "gold standard" for establishing efficacy, they are not a panacea to answer all clinical questions. By operating in a tightly controlled environment with strict inclusion and exclusion criteria, RCTs can only measure efficacy in limited populations. As such, they cannot provide a true indication of effectiveness in real-world settings. Furthermore, in many situations, RCTs are not feasible because of practical, legal, or ethical reasons. Therefore, real-life effectiveness data obtained by other methods, such as databases, surveys, chart reviews, cohort studies, and pragmatic clinical trials, are often used. Databases, in particular, can be very powerful tools for outcomes research and quality improvement.

The purpose of this chapter is to explore the use of databases in assessing the effectiveness and quality of endoscopy, look for current examples of their use, and consider the strengths and limitations of each. Specifically, data sources for quality assessment in endoscopy including administrative databases, clinical registries, and electronic health records (EHRs) that capture clinical data at the point-of-care will be reviewed.

Administrative databases

Administrative databases, also called claims data, are computerized records of encounters that are used to support the routine administration of healthcare programs by recording billing activities or healthcare utilization. These data are often compiled by government agencies or third-party payers responsible for funding health care. Core data elements usually include demographic variables, diagnosis codes (e.g. *International Classification of Diseases, Ninth Revision, Clinical Modification* [ICD-9-CM]) and procedure codes (e.g. *Current Procedure Terminology*, [CPT-4]).

Strengths and limitations

Administrative data can be used for assessing the effectiveness of interventions in real-world settings. Recent examples in endoscopy have included the use of a national American database, the Nationwide Inpatient Sample (NIS). The NIS includes inpatient discharge data collected via federal–state partnerships, as part of the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project. As of the year 2009, the NIS totaled administrative data on approximately 8 million hospital stays each year from 1050 hospitals within 40 states, approximating 20 % of community hospitals within the United States, including public hospitals and academic medical centers. It is the sole hospital database in the United States with charge information on all patients regardless of payer, including persons covered by Medicare, Medicaid, private insurance, and the uninsured. Groups have used this database to demonstrate a weekend effect on mortality with regards to the use of endoscopy and patients presenting with upper GI bleeding [21]. The advantages of using such databases for outcomes research are their availability, low cost, easy accessibility of data in electronic formats, and ability to measure large samples of the population over a broad geography. On the other hand, there are many limitations of administrative data which could potentially compromise their validity and reliability for research and quality assessment. First, because the data are maintained primarily for administrative and financial purposes, they are often limited in clinical details to inform quality assessment. In fact, most administrative databases explicitly aim to collect the minimum amount of information required to perform the relevant function [22]. Second, diagnosis (ICD-9-CM) and procedure (CPT-4) codes are often used to identify patient populations and as surrogates for clinical outcomes, but are infrequently validated [23]. When administrative data are used to identify target populations of interest, studies have shown fair specificity, but low sensitivity [24–27]. In a recent such example by Wyse et al., the authors collected data on 689 patients who underwent colonoscopy during the study period. The sensitivity of physician claims for polypectomy in the administrative database was 84.7 % (95 % confidence interval (CI) 78.6–89.4 %), the specificity was 99.0 % (95 % CI 97.5–99.6 %), concordance was 95.1 % (95 % CI 93.1–96.5 %), and the kappa value was 0.87 (95 % CI 0.83–0.91) [28]. Conversely, when quality assessment is based on administrative data alone, under-detection of quality indicators has been reported [28–31]. This could be because the algorithms for the identification of target populations or clinical outcomes depend on the accuracy and completeness of coding. Unfortunately, coding inaccuracy can occur due to administrative errors, imprecise or ambiguous definitions, or financial incentives to "upcode" conditions to maximize reimbursement. Therefore, validation of administrative data by comparison with other data sources (e.g. medical charts, clinical registries, self-reported surveys) is an important prerequisite for the use of claims-based data. Third, the timing of diagnoses may be impossible to ascertain. Being able to distinguish between a condition that arises during treatment (a complication) and a pre-existing condition that was present before treatment (comorbidity) is essential for assessing the quality of care.

Using administrative data for assessing colonoscopy outcomes and quality: population-based studies

A series of population-based studies on screening colonoscopy conducted in Canada highlight some of the strengths and limitations of using administrative databases [5,32–35]. In a study that assessed colonoscopy completion rate and factors associated with incomplete procedure in Ontario, administrative databases were used to identify average risk screening individuals who underwent colonoscopy [35]. Patient, endoscopist and setting factors were evaluated for

their association with incomplete colonoscopy. Of the 331,608 individuals who had a colonoscopy, 43,483 (13.1 %) were incomplete [35]. This is greater than the quality targets of ≤10 % for all colonoscopies and ≤5 % for screening colonoscopies set by the MSTF-CRC [16]. The analysis also found that increased patient age (OR 1.20 per 10-year increment; 95 % CI 1.19–1.22), female sex (OR 1.35; 95 % CI 1.30–1.39), and having the procedure done in a private office (OR 3.57; 95 % CI 2.55–4.98) were factors strongly associated with incomplete colonoscopy [35]. In another study, the pooled rates of colonoscopy-related bleeding and perforation were found to be 1.64/1000 and 0.85/1000, respectively [33]. These rates were within the target of ≤1/1000 for all colonoscopies set by the MSTF-CRC [16]. However, older age, male sex, having a polypectomy, and having the colonoscopy performed by low-volume endoscopists were associated with increased odds of these complications [33]. In another study, the rate of new or missed CRC after colonoscopy was found to be between 2 % and 6 %, depending on the site of the cancer [5]. The analysis also identified right-sided or transverse CRC as one of the risk factors for new or missed CRC [5]. Concerns were, therefore, raised about the lower protective effects of screening colonoscopy for right-sided CRC compared to left-sided CRC.

Because such type of information can only be determined from large unselected cohorts, these studies thus provide important insights about the quality of colonoscopy in usual clinical practice. Indeed, the major strength of these studies is the use of large-scale population-based cohorts because they more accurately reflect usual practice than cohorts from single centers. However, the results must be interpreted in light of the strengths and limitations related to the use of administrative databases. First, the procedure codes have not been validated in those settings (although they have elsewhere, such as in Quebec) [28], and coding errors may have resulted in misclassification. It is possible that some procedures coded as complete were in fact incomplete. Therefore, the completion rate could have been overestimated. This may also be one of the reasons why colonoscopy appeared to be less protective for right-sided CRC. Second, the indications for colonoscopy were unknown. Although inclusion and exclusion criteria were used to approximate screening colonoscopy cohorts, the investigators could not distinguish screening from diagnostic colonoscopies based on the limited clinical information. If rectal bleeding was a common indication, the likelihood of finding a left-sided lesion would be presumably higher. This in turn may lead to an apparent stronger protective effect of colonoscopy for left-sided CRC. Third, the investigators were not able to evaluate additional clinical factors such as comorbid illnesses, medication use (e.g. aspirin, nonsteroidal anti-inflammatory agents, warfarin), and quality of bowel preparation – factors that could be associated with the outcomes of interest. Quality of bowel preparation, in particular, can have a significant impact on the completion rate and may explain the lower protective effects of colonoscopy for right-sided CRC.

Clinical registries

Clinical registries are databases of specific clinical conditions, procedures, therapies, or populations. They are set up to efficiently monitor patterns of care and progression of disease, evaluate effectiveness and safety, and improve outcomes and quality in real-world settings. Detailed but selected information on clinically important events is systematically captured from clinical records, and entered into computer databases in a structured manner on an ongoing basis. Such registries not only allow clinicians to efficiently monitor and treat patients with specific conditions, but also expedite the identification of potential participants for research studies. In addition, the data can facilitate effectiveness research. Important observations of associations can be generated and utilized as basis for prospective studies. By continuously capturing data, registries have the potential to identify variations in practice and drive quality improvement by creating a continuous feedback loop.

Strengths and limitations

Clinical registries are ideal for assessing effectiveness and quality of interventions in real-world settings. They can enhance knowledge of clinical service patterns, processes and patient outcomes and can capture valuable, real-time patient data that are not present in administrative databases. Data derived from clinical registries are also devoid of many of the limitations of

Figure 18.1 RUGBE participating sites. Image courtesy of Dr Alan Barkun.

administrative data. However, significant investment is required for setting up and organizing the registry, collecting the data and instituting quality control measures. There are also many potential limitations of registry data that can threaten their internal and external validity. These include accuracy and completeness of data. Also, ambiguous data definitions can lead to variability in data interpretation and misclassification. To assure data quality, all registry personnel should be trained in a standardized fashion for data abstraction, validation and verification. Range and consistency checks as well as external validation of data using alternative sources are required for quality control. Because selection bias can arise with selected or incomplete patient sampling, explicit inclusion and inclusion criteria and sampling rules are required to ensure accurate representation of the targeted population in the registry. Finally, unmeasured confounding factors may still be present despite comprehensive and detailed data collection, which in turn can lead to erroneous conclusions. Hence, any demonstrated associations based on observational data should only be considered exploratory and require prospective confirmation.

Using clinical registry for assessing effectiveness of interventions: the Canadian registry on nonvariceal upper GI bleeding undergoing endoscopy (RUGBE)

The Canadian Registry on Nonvariceal Upper GI Bleeding undergoing Endoscopy (RUGBE) initiative demonstrates how observational data from a national clinical registry can be useful adjuncts to RCTs in determining whether efficacy under controlled conditions in representative centers would translate into effective treatment in routine practice (Figure 18.1). The RUGBE is a multicenter retrospective registry that collected descriptive data on 1869 randomly selected patients with nonvariceal upper gastrointestinal bleeding (NVUGIB) managed at 12 Canadian university-affiliated and 6 community centers between 1999 and 2002 [36]. Extensive information was collected for each patient including demographics, past medical history, medication intake, physical examination findings, laboratory values, resuscitative efforts, endoscopic diagnosis and treatment, management, and outcomes.

The RUGBE registry data provided important information about the management and outcomes of patients with NVUGIB in real-world settings. As

shown previously in RCTs, the RUGBE data confirmed that both endoscopic therapy and PPI use decreased rebleeding in patients with high-risk stigmata [37,38]. The data also suggested a mortality benefit attributable to endoscopy in this subgroup as reported by previous meta-analyses [38]. In addition, the data provided some new findings. First, acute use of PPI use was associated with decreased rebleeding in all patients regardless of endoscopic stigmata (OR: 0.53; 95 % CI 0.37–0.77) [36]. The reasons for this may include adequate statistical power to demonstrate a small but significant benefit in low-risk patients. Alternatively, high-risk stigmata may have been incorrectly diagnosed as low risk (ascertainment bias). Second, PPI use (OR: 0.18; 95 % CI 0.04–0.80) was independently associated with decreased mortality in patients with high-risk stigmata [36]. In fact, the mortality benefit with PPI use was supported at the time only by observed trends from smaller randomized trials, but was later confirmed by a large Cochrane systematic review [39]. Over the years, this database has been used to drive many prospective studies and clinical trials, as well as inform nonvariceal upper GI bleeding patient care with issues such as early discharge, process-related factors, and the predictive role of inhospital onset of bleeding or that of the presenting INR, and the development of international guidelines on the management of NVUGIB [36,40–45].

Using clinical registry for assessing quality of interventions: the GI Quality Improvement Consortium (GIQuIC) and the Global Rating Scale (GRS)

A key agenda for health systems worldwide is the improvement of safety and quality of health care provided to patients. As a result, a high priority has been placed on measuring quality and using those measurements to promote improvements in the delivery of care, to influence payment for services, and to increase transparency [46]. Clinical registries provide one of the most accurate and efficient means for measuring quality, particularly for measuring outcomes of endoscopy. There are now many national quality improvement programs and benchmarking initiatives around the world. For example, the GI Quality Improvement Consortium (GIQuIC) was established by the American College of Gastroenterology (ACG) and the American Society for Gastrointestinal Endoscopy (ASGE) to support quality improvement efforts by creating a national GI endoscopy data repository of quality measures. GIQuIC collects quality indicators electronically that have been abstracted either manually or electronically, and provides benchmarking reports to participating physicians and facilities to support their quality improvement initiatives. The data can also be used for conducting outcomes research and quality initiatives. The colonoscopy module with 84 quality indicators was launched in July 2010. It is anticipated that modules for endoscopy, endoscopic retrograde cholangiopancreatography (ERCP) and endoscopic ultrasound (EUS) procedures will be available soon. The results of documented performance improvement and quality participation could set the stage for improved reimbursements through Medicare and other third-party payers. In the United Kingdom, a patient-centered quality assurance program known as the Global Rating Scale (GRS) was developed in 2004 to provide an assessment of the quality of endoscopy services and to guide improvement efforts [47]. Use of the GRS is facilitated by a dedicated website for data entry of quality measures. Acceptance of the GRS in the UK has been high, and improvements in quality and reductions in waiting times have been achieved [47]. It is important to note that such scales may need to be adapted for a given country, such as the case for the GRS in Canada [48].

Electronic health record databases

Electronic health record (EHR) databases are computer-based health records that are used to document all clinical activities related to endoscopy. They are designed primarily to support clinical care by recording endoscopic findings, storing images, and generating reports. Unfortunately, many stand-alone reporting systems suffer from poor accessibility of data for research and limited abilities for extracting or exporting data. A structured and comprehensive electronic endoscopic database, however, can potentially be used for research and quality improvement.

Using electronic health record databases for assessing outcomes and quality of interventions: the Clinical Outcomes Research Initiative (CORI)

The Clinical Outcomes Research Initiative (CORI) experience is an example of how a national

multicenter electronic database may be used to advance the quality of patient care and facilitate outcomes research. Under the auspices of the ASGE, the CORI database was established in 1995 to study utilization and outcomes of endoscopy in a wide variety of practice settings distributed throughout the United States. The practice sites include community or private practices (80 %), academic centers (10 %), and Veterans Affairs (VA) medical centers (10 %) [49]. All participating sites must use a structured, computerized endoscopic report generator to complete all endoscopic reports and comply with quality control requirements. Based on guidelines from the ASGE regarding the standard elements of an endoscopy report [50,51], data including demographics, comorbidity using the American Society of Anesthesiologists (ASA) classification, indications, sedation, procedure completeness, endoscopic findings, details of endoscopic interventions, and immediate complications are entered into all procedure reports in a structured fashion. Encrypted data of each procedure is transmitted electronically from each practice site to a central data repository – the National Endoscopic Database (NED). The data are then subjected to a series of computerized quality control measures. After completion of quality control checks, the data from all sites are merged in the NED for analysis.

CORI receives about 250,000 endoscopic reports annually from more than 65 adult and 12 pediatric practice sites in 24 states with approximately 500 participating physicians (representing approximately 1 % of all endoscopies performed nationally in Medicare patients) [52,53]. To date, CORI is the largest national clinical data repository for GI endoscopy in the USA with close to two million reports. The database has been validated, and the data derived have been shown to provide a representative picture of endoscopic practice patterns in the United States [53].

CORI has been a valuable resource for promoting endoscopic clinical research with the goal of measuring utilization and outcomes to improve the practice of endoscopy. Historically, most research studies are conducted in academic or tertiary care centers which limit the generalizability of their findings to broader populations. The obvious strength of CORI is that it is representative of real-world practice patterns across a wide variety of practice settings throughout the United States [53]. For these reasons, the CORI database not only provides a unique opportunity to study and monitor the epidemiology of GI diseases, but also provides essential data to characterize existing and evolving endoscopic practice patterns to allow benchmarking of quality-of-care outcomes. Research goals of CORI include analyses of practice over time to identify areas for quality improvement, and to develop interventions to improve quality (Table 18.2). As of 2011, these analyses have resulted in publications of 61 manuscripts in peer-reviewed journals covering a wide range of endoscopic topics [49,53–100]. These include studies of: (i) *variations* in trends and endoscopic practice patterns, (ii) *effectiveness and safety* of interventions, (iii) *appropriateness* of procedures, and (iv) *quality indicators* for endoscopic procedures. As an example of how CORI database can be used for outcome research and quality improvement, we will examine the utilization of screening and surveillance colonoscopy. Examples of other endoscopic databases that have led to determinations of utilization or facilitated quality improvement initiatives can also be found using adaptations of commercially available software [101,102].

Table 18.2 Research goals of CORI

- Perform descriptive analyses of endoscopic utilization, frequency and severity of endoscopic findings, endoscopic treatment and medical management.
- Observe the natural history of chronic GI diseases for which endoscopic surveillance is used.
- Determine the success and effectiveness of endoscopic therapies
- Determine the impact of endoscopic diagnosis and therapies on patient outcomes, such as morbidity, mortality, quality of life, functional status, and healthcare utilization.
- Evaluate the frequency of endoscopic complications, and risk factors for complications.
- Prospectively monitor the results of new endoscopic innovations.
- Assess the validation of quality indicators by studying the relationship between specific indicators and key endoscopic endpoints.
- Identify subjects for prospective research projects.

Source: Adapted from http://www.cori.org.

CORI: Identifying variations in trends and endoscopic practice patterns

Studies that have evaluated colonoscopy practice trends in the United States have shown a significant rise in the utilization of colonoscopy over time largely because of increased rates of CRC screening [69,86]. Since the introduction of Medicare coverage for average-risk screening colonoscopy in 2001, the proportion of procedures performed for average-risk screening has increased dramatically from 4.6 % (before July 2001) to 14.2 % (after July 2001) [69].

Although CRC screening rates have increased over time, there is evidence that CRC screening still remains under-utilized [103]. National data on CRC screening uptake showed that only 47 % of men and 43 % of women over 50 years of age received CRC screening [104], although more recent estimates now suggest that 62.9 % of Americans aged 50–75 years are up to date with CRC screening, however leaving more than 22 million adults untested [105]. Furthermore, there still exist significant racial and ethnic disparities in screening utilization [106]. For example, the CORI data has shown that Blacks and Hispanics are underrepresented relative to their proportion in the population [86]. Notably, asymptomatic screening was performed more commonly in Whites than non-Whites (36.2 % vs. 34.0 %, $P < 0.001$), while non-white patients were more likely to have colonoscopy to evaluate symptoms (36.4 % vs. 27.1 %, $P < 0.001$) [86]. These findings may suggest that non-white patients do not receive colonoscopy until they develop symptoms suggestive of serious pathology. Indeed, potential precancerous lesions (polyps sized >9 mm) were found more commonly in Blacks than Whites (7.7 % vs. 6.2 %, $P < 0.001$) with greater disparities seen in black women (OR 1.62; 95 % CI 1.39–1.89) than in black men (OR 1.16; 95 % CI 1.01–1.32) [87]. Similarly, Blacks undergoing screening colonoscopy were found to have higher odds of tumors (OR 1.78; 95 % CI 1.14–2.77) and proximal tumors than Whites (OR 4.37; 95 % CI 1.16–16.42) [107]. The reasons for higher incidence rates of precancerous and cancerous lesions in Blacks are unclear; however, genetics, dietary and nutritional factors, physical inactivity, smoking, and under-utilization of screening colonoscopy have been most commonly implicated [85,87,108,109]. Apart from black race, multivariate analysis also suggested that increasing age and male gender were associated with increased risk of mass or polyps >9 mm [85]. In other studies, individuals older than 60 years, female sex, and patients with a family history of CRC were more likely to have proximal lesions [87,93]. The clinical implications of these results would therefore push for complete colonoscopy particularly for these patients because of the proximal pathology.

These various examples illustrate how the CORI database can provide data on incidence, prevalence, and demographic factors that may yield clues to the etiology of disease or help identify associations and risk factors not previously known. Also, by examining the variation in existing and evolving practice patterns, these data can help identify potential gaps in clinical practice that can be utilized to guide decision making and target resources and policies to populations that would benefit the most.

CORI: Assessing the effectiveness and safety of interventions

Concurrent with the rise in the utilization of colonoscopy, detection of significant colonic lesions (masses and polyps greater than 9 mm) has declined from 4.9 % (before July 2001) to 3.8 % (after July 2001) [69]. There was also a sharp decline in the endoscopic diagnosis of CRC from 109.9 (98.3–122.8) to 72.2 (67.4–77.2) per 1000 colonoscopies between 2000 and 2003 [55]. These findings likely reflect a general increase in public awareness of CRC screening due to celebrity endorsement [57] and various promotional campaigns as well as an expansion of Medicare coverage. Additionally, these results may also support the therapeutic benefit of polypectomy in reducing the risk of CRC development. It is reassuring to note that the risk of serious complications after screening and surveillance colonoscopy was found to be low, with an incidence of perforations of 0.19 per 1000 examinations and GI bleeding requiring hospitalization in 1.59 per 1000 examinations [77]. Overall, screening colonoscopy appears to be a safe and effective intervention in real-world settings.

There are no RCTs of screening colonoscopy. Although several case-control studies have demonstrated the benefits of CRC screening [110–112], few studies have examined the effectiveness of CRC screening as it is used in real-world settings. Based on the association between the utilization of colonoscopy

and CRC diagnosis, the CORI data suggests that screening colonoscopy may effectively and safely reduce the incidence of CRC. Additional recent observational data from other groups appear to confirm this finding although the effectiveness of colonoscopy in improving the prognosis from right-sided colonic neoplasias remains controversial [111,113,114]. It is important to remember that observational data cannot prove causality due to possible unmeasured confounding factors, and these results should therefore be interpreted with caution and should only be used to supplement RCTs.

CORI: Assessing the appropriateness of procedures

The US Preventative Services Task Force and many existing clinical practice guidelines strongly recommend routine CRC screening for individuals aged >/=50 years as an effective strategy for reducing CRC incidence and mortality [115–119]. Despite these recommendations and Medicare coverage, rates of CRC screening remain low [103,120]. Nevertheless, there is evidence to suggest both under- and over-utilization of colonoscopy in clinical practice. For example, analysis of the CORI data found that patients less than 50 years of age accounted for 20 % of colonoscopy procedures performed [86]. Appropriate indications were present in only 35 % of examinations in these patients (positive family history of CRC, surveillance of polyps/CRC, surveillance or evaluation of inflammatory of bowel disease, anemia or iron deficiency) [86]. In the remaining patients less than 50 years of age (65 %), the procedure indications may be controversial [86]. Specifically, colonoscopy was often performed for irritable bowel symptoms (23.8 %), hematochezia (33.6 %), or average risk screening (12.8 %), for which benefits are uncertain in younger patients [86]. Overall, surveillance after polypectomy and cancer resection was the most common indication for colonoscopy in clinical practice, accounting for 24 % of all procedures [83].

To date, there have been no RCTs examining the incremental benefits of surveillance examinations after initial screening colonoscopy with polypectomy. According to microsimulation modeling of the US National Polyp Study data, initial polypectomy as opposed to subsequent surveillance examinations accounted for 80 % of the mortality reduction resulting from colonoscopy [121,122]. Additionally, population-based case-control studies have indicated low risks for CRC to at least 5 years after polypectomy [123] and to beyond 10 years after a negative initial colonoscopy [124]. Consequently, recent guidelines have extended surveillance intervals based on risk stratification for recurrent advanced adenomas. Yet, poor adherence to these guidelines continues to be a major barrier for cost-effective CRC prevention strategies [10–12]. Indeed, survey studies based on physicians' self-reported practice patterns have suggested that postpolypectomy surveillance is often done inappropriately at more frequent intervals than guidelines recommend [9,13–15], a pattern which has been confirmed in clinical studies as well [10–12].

Analysis of the CORI data, therefore, suggested a potential pattern of over-utilization of colonoscopy which could expose patients to unnecessary risks without providing benefits and divert scarce resources that could be used for more cost-effective CRC prevention strategies. These examples illustrate how the CORI database can provide data on the appropriateness of resource utilization that may lead to further studies to measure optimal use of endoscopic procedures and provide a basis for aligning resources with more cost-effective practice. In the case of CRC screening and surveillance, reductions in over-utilization in low-risk populations can free up resources to improve the under-utilization of cancer screening in high-risk populations.

CORI: Monitoring quality indicators for endoscopic procedures

The CORI database highlights some possible ways in which a computerized reporting software may be used to measure and improve quality. As most of the quality indicators proposed in CO-RADS are captured by the CORI software [20], analysis of the CORI data can provide a snapshot of colonoscopy quality across practices. In a study on the quality of colonoscopy reporting, the CORI database was queried to determine if specific quality indicators were recorded in the procedure reports [52]. In addition, quality indicators for the technical performance of colonoscopy, including cecal intubation rates, polyp detection rates and unplanned interventions for adverse events were compared within and among practices in the CORI consortium [52]. Of the 438,521 reports analyzed, 13.9 % did not include bowel preparation quality, and cecal landmarks were not recorded in 14 % of

reports [52]. Without these key elements in the report, it would be impossible to determine the adequacy of the examination. Also, polyp descriptors (size, morphology) were not recorded in 4.9 % and 14.7 % of reports [52]. This information is especially important for determining surveillance intervals for repeat colonoscopy. Inadequate reporting may in turn lead to unnecessary repeated examinations. Overall, the CORI data demonstrated significant variation in quality of colonoscopy reports across diverse practices even with the use of a structured reporting system.

In terms of technical performance of colonoscopy, analysis of the CORI data revealed that the cecal intubation rate was 96.3 % when this was intended [52]. Among average-risk individuals, the detection rate of polyp (>9 mm) adjusted for age, gender, and race, was between 4 % and 10 % in 81 % of practices [52]. These results were generally consistent with other clinical studies in which the detection rates of advanced neoplasia, defined by polyp size (≥10 mm) or histology (villous or high-grade dysplasia), were reported to be between 4 % and 10 % [125–128]. In this study, polyp >9 mm was used as a surrogate endpoint for advanced neoplasm [52]. This endpoint was validated by a previous CORI study which demonstrated that a polyp >9 mm harbored an estimated 82 % chance of being an adenoma and a 30.6 % chance of having advanced histology [82]. Reporting of unplanned interventions for adverse events during the procedure varied from 0 % to 6.5 % [52]. These differences among sites may be because of true variation in adverse-event rates or may reflect different reporting thresholds [52].

The CORI project demonstrated how measurement of quality can be facilitated by using a computerized reporting system. Endoscopists using a database like CORI can measure and monitor their practice to determine if quality improves. In addition, CORI can produce aggregate data from multiple sites for benchmarking purposes. Discovery of quality or lack of quality can result in hypothesis and promote quality improvement measures. The hypothesis that performance can improve by measuring and monitoring quality indicators was tested in a prospective study in which the impact of feedback on three key performance indicators: cecal intubation rate, insertion time, and withdrawal time, was assessed [129]. In this study, anonymous feedback was provided to all endoscopists by emails every 3 months. Individual endoscopists can also compare their performance with the aggregate data. Following the introduction of feedback, there was a relative decline of 19 % in incomplete procedures, while median insertion times declined from 10.6 mins to 9.5 mins ($P = 0.02$) without any effect on median withdrawal times. These findings support the utility of feedback in enhancing colonoscopy performance.

Strengths and limitations of CORI

The main advantage of using EHR databases such as CORI for research and quality measurement is the easy accessibility of comprehensive data in electronic formats. The databases can also be customized to capture structured data to answer any research questions. By providing more than one way of triangulating the data, EHR databases have been shown to be more accurate in identifying target population than administrative databases [130,131]. Although the CORI database facilitates large-scale observational studies leading to an improved understanding of endoscopic practice patterns and patient outcomes, there are several notable limitations that are inherent in this type of database. Similar to other EHR systems, missing or incomplete data (except for mandatory data fields) represents a limitation of CORI. Furthermore, the data are only as accurate as the data provided by the endoscopists. For example, clinical and demographic data are entered at the discretion of the endoscopists and therefore may be subject to misclassification bias. In particular, information pertaining to race/ethnicity is often based on the endoscopist's impression rather than the patient's history. The recorded indications may also be highly susceptible to misclassification bias. This is due to the fact that endoscopists may be more likely to enter an indication that is more acceptable for third-party reimbursement. The most important limitation of CORI is the lack of detailed clinical information of patients such as specific comorbidities, patient characteristics, medication use, prior diagnostic tests or treatments. As well, follow-up of patients and assessment of complications may be incomplete. In the case of surveillance colonoscopy, data on the pathology of the baseline lesions or adequacy of prior exams are not recorded in CORI. These are factors which could strongly influence the decision on surveillance and screening intervals. Additionally, endoscopic findings are often based on

subjective interpretation rather than objective confirmation by histopathology data. Hence, studies often resort to the use of surrogate endpoints (e.g. key performance indicators, polyp size) based on the assumption that they correlate with clinically relevant outcomes. The CORI system attempts to standardize endoscopic reports with controlled terminology, but does allow for unstructured free-text entry. As a result, some useful information may be locked in free-text and could be difficult to extract for analysis. There are also no explicit definitions for some of the variables in the database (e.g. bowel preparation quality). Finally, although the CORI consortium was designed to represent the diverse practice settings in the United States, the possibility of selection bias cannot be excluded. Physicians who participate in the CORI consortium are comfortable using an electronic endoscopic reporting system, and are likely to be highly motivated clinicians interested in improving quality and efficiency of patient care. Based on national survey studies, it is evident that EHR systems have only been adopted by a small minority of US physicians [132–135]. Thus, the CORI consortium may not be truly representative of the "average" endoscopic practice in the United States.

In summary, the use of CORI data is subject to the same caveats as for conducting research by using any large databases. While observational studies may provide a powerful alternative for outcome research, especially in situations where large RCTs may not be a feasible option, it is important to remember that observational data cannot prove cause-and-effect relationships. Investigators must be cautious that any proposed associations may be due to confounding variables that are not captured in the database. RCTs remain the gold standard for evaluating the effects of any interventions or treatment strategies. There have been many instances whereby the results of large RCTs not only failed to confirm findings of observational studies, but in some cases refuted them [136].

Conclusions

Administrative databases, clinical registries and EHR databases are important tools for clinical outcomes research and quality improvement initiatives. While RCTs remain the accepted gold standard for determining the efficacy of interventions, they are costly and cannot provide information about the effectiveness and quality of interventions in real-world settings. Research using administrative databases, clinical registries and EHR databases can complement RCTs in providing outcome data for enhancing quality. Administrative databases play a central role in epidemiologic evaluation because of their wide availability, low cost, easy accessibility of data, and ability to measure large samples of the population. However, their usefulness for measuring quality of care is limited due to lack of clinical details and coding inaccuracy. Clinical registries are designed for both research and quality improvement, but significant investment is required for data collection and quality control. EHR databases can be customized to collect all the data needed for outcome research and quality assessment, but are subject to the same caveats as for conducting research by using any large databases including quality of data, lack of clinical details, and use of unstructured free-text. Given these various limitations, investigators should view these databases only as hypothesis-generating tools that can: (i) stimulate further investigations with prospective studies; (ii) confirm findings of RCTs; and (iii) assess implementation and dissemination of evidence in real-world settings. In the future, integration of clinical registries with EHR systems and linking of clinical registries with administrative databases may provide tremendous opportunities for research and quality improvement.

Multiple choice questions

1 Administrative databases include the following data elements except:

A Diagnosis codes
B Procedure codes
C Demographic variables
D Comorbid illnesses
E Admission and discharge dates

2 Clinical registries are ideal for assessing effectiveness and quality of interventions in real-world settings because they:

A Allow clinicians to efficiently monitor and treat patients with specific conditions
B Expedite the identification of potential participants for research studies
C Facilitate effectiveness research

D Identify variations in practice and drive quality improvement by creating a continuous feedback loop

E All of the above

3 The following is *NOT* a limitation of electronic health record databases:

A Coding inaccuracy

B Missing or incomplete data

C Lack of detailed clinical information

D Data inaccuracy

E Information may be locked in free-text and could be difficult to extract for analysis

References

1 Seeff LC, Richards TB, Shapiro JA, et al. How many endoscopies are performed for colorectal cancer screening? Results from CDC's survey of endoscopic capacity. *Gastroenterology* 2004;127(6):1670–7.

2 Singh H, Penfold RB, DeCoster C, et al. Colonoscopy and its complications across a Canadian regional health authority. *Gastrointest Endosc* 2009;69(3 Pt 2):665–71.

3 Rex DK, Rahmani EY, Haseman JH, et al. Relative sensitivity of colonoscopy and barium enema for detection of colorectal cancer in clinical practice. *Gastroenterology* 1997;112(1):17–23.

4 Rex DK, Cutler CS, Lemmel GT, et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology* 1997;112(1):24–8.

5 Bressler B, Paszat LF, Chen Z, et al. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. *Gastroenterology* 2007;132(1):96–102.

6 Sanchez W, Harewood GC, Petersen BT. Evaluation of polyp detection in relation to procedure time of screening or surveillance colonoscopy. *Am J Gastroenterol* 2004;99(10):1941–5.

7 van Rijn JC, Reitsma JB, Stoker J, et al. Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol* 2006;101(2):343–50.

8 Rex DK. Colonoscopic withdrawal technique is associated with adenoma miss rates. *Gastrointest Endosc* 2000;51(1):33–6.

9 Boolchand V, Olds G, Singh J, et al. Colorectal screening after polypectomy: a national survey study of primary care physicians. *Ann Intern Med* 2006;145(9):654–9.

10 Krist AH, Jones RM, Woolf SH, et al. Timing of repeat colonoscopy: disparity between guidelines and endoscopists' recommendation. *Am J Prev Med* 2007;33(6):471–8.

11 Schoen RE, Pinsky PF, Weissfeld JL, et al. Utilization of surveillance colonoscopy in community practice. *Gastroenterology* 2010;138(1):73–81.

12 Laiyemo AO, Pinsky PF, Marcus PM, et al. Utilization and yield of surveillance colonoscopy in the continued follow-up study of the polyp prevention trial. *Clin Gastroenterol Hepatol* 2009;7(5):562–7.

13 Mysliwiec PA, Brown ML, Klabunde CN, Ransohoff DF. Are physicians doing too much colonoscopy? A national survey of colorectal surveillance after polypectomy. *Ann Intern Med* 2004;141(4):264–71.

14 Rossi F, Sosa JA, Aslanian HR. Screening colonoscopy and fecal occult blood testing practice patterns: a population-based survey of gastroenterologists. *J Clin Gastroenterol* 2008;42(10):1089–94.

15 Saini SD, Nayak RS, Kuhn L, Schoenfeld P. Why don't gastroenterologists follow colon polyp surveillance guidelines?: results of a national survey. *J Clin Gastroenterol* 2009;43(6):554–8.

16 Rex DK, Bond JH, Winawer S, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the U.S. Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol* 2002;97(6):1296–308.

17 Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. *Gastrointest Endosc* 2006;63(4 Suppl):S16–S28.

18 Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. *Am J Gastroenterol* 2006;101(4):873–85.

19 Faigel DO, Pike IM, Baron TH, et al. Quality indicators for gastrointestinal endoscopic procedures: an introduction. *Gastrointest Endosc* 2006;63(4 Suppl):S3–S9.

20 Lieberman D, Nadel M, Smith RA, et al. Standardized colonoscopy reporting and data system: report of the Quality Assurance Task Group of the National Colorectal Cancer Roundtable. *Gastrointest Endosc* 2007;65(6):757–66.

21 Ananthakrishnan AN, McGinley EL, Saeian K. Outcomes of weekend admissions for upper gastrointestinal hemorrhage: a nationwide analysis. *Clin Gastroenterol Hepatol* 2009;7(3):296–302e1.

22 Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997;127(8 Pt 2):666–74.

23 van WC, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. *J Clin Epidemiol* 2011;64(10):1054–9.

24 Atreja A, Achkar JP, Jain AK, et al. Using technology to promote gastrointestinal outcomes research: a

case for electronic health records. *Am J Gastroenterol* 2008;103(9):2171–8.
25 Raiford DS, Perez GS, García Rodríguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. *Epidemiology* 1996;7(1):101–4.
26 Lopushinsky SR, Covarrubia KA, Rabeneck L, et al. Accuracy of administrative health data for the diagnosis of upper gastrointestinal diseases. *Surg Endosc* 2007;21(10):1733–7.
27 Hawker GA, Coyte PC, Wright JG, et al. Accuracy of administrative data for assessing outcomes after knee replacement surgery. *J Clin Epidemiol* 1997;50(3):265–73.
28 Wyse JM, Joseph L, Barkun AN, Sewitch MJ. Accuracy of administrative claims data for polypectomy. *CMAJ* 2011;183(11):E743–E747.
29 Jollis JG, Ancukiewicz M, DeLong ER, et al. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. *Ann Intern Med* 1993;119(8):844–50.
30 Keating NL, Landrum MB, Landon BE, et al. Measuring the quality of diabetes care using administrative data: is there bias? *Health Serv Res* 2003;38(6 Pt 1):1529–45.
31 Ko CW, Dominitz JA, Green P, et al. Accuracy of Medicare claims for identifying findings and procedures performed during colonoscopy. *Gastrointest Endosc* 2011;73(3):447–53.
32 Lakoff J, Paszat LF, Saskin R, Rabeneck L. Risk of developing proximal versus distal colorectal cancer after a negative colonoscopy: a population-based study. *Clin Gastroenterol Hepatol* 2008;6(10):1117–21.
33 Rabeneck L, Paszat LF, Hilsden RJ, et al. Bleeding and perforation after outpatient colonoscopy and their risk factors in usual clinical practice. *Gastroenterology* 2008;135(6):1899–906.
34 Rizek R, Paszat LF, Stukel TA, et al. Rates of complete colonic evaluation after incomplete colonoscopy and their associated factors: a population-based study. *Med Care* 2009;47(1):48–52.
35 Shah HA, Paszat LF, Saskin R, et al. Factors associated with incomplete colonoscopy: a population-based study. *Gastroenterology* 2007;132(7):2297–303.
36 Barkun A, Sabbah S, Enns R, et al. The Canadian Registry on Nonvariceal Upper Gastrointestinal Bleeding and Endoscopy (RUGBE): Endoscopic hemostasis and proton pump inhibition are associated with improved outcomes in a real-life setting. *Am J Gastroenterol* 2004;99(7):1238–46.
37 Lau JY, Sung JJ, Lee KK, et al. Effect of intravenous omeprazole on recurrent bleeding after endoscopic treatment of bleeding peptic ulcers. *New Engl J Med* 2000;343(5):310–6.
38 Cook DJ, Guyatt GH, Salena BJ, Laine LA. Endoscopic therapy for acute nonvariceal upper gastrointestinal hemorrhage: a meta-analysis. *Gastroenterology* 1992;102(1):139–48.
39 Leontiadis GI, Sharma VK, Howden CW. Proton pump inhibitor treatment for acute peptic ulcer bleeding. *Cochrane Database Syst Rev* 2006;(1):CD002094.
40 Barkun A, Fallone CA, Chiba N, et al. A Canadian clinical practice algorithm for the management of patients with nonvariceal upper gastrointestinal bleeding. *Can J Gastroenterol* 2004;18(10):605–9.
41 Barkun AN. The role of intravenous proton pump inhibitors in the modern management of nonvariceal upper gastrointestinal bleeding. *Drugs Today (Barc)* 2003;39(Suppl A):3–10.
42 da Silveira EB, Lam E, Martel M, et al. The importance of process issues as predictors of time to endoscopy in patients with acute upper-GI bleeding using the RUGBE data. *Gastrointest Endosc* 2006;64(3):299–309.
43 Enns RA, Gagnon YM, Barkun AN, et al. Validation of the Rockall scoring system for outcomes from nonvariceal upper gastrointestinal bleeding in a Canadian setting. *World J Gastroenterol* 2006;12(48):7779–85.
44 Muller T, Barkun AN, Martel M. Non-variceal upper GI bleeding in patients already hospitalized for another condition. *Am J Gastroenterol* 2009;104(2):330–9.
45 Shingina A, Barkun AN, Razzaghi A, et al. Systematic review: the presenting international normalised ratio (INR) as a predictor of outcome in patients with upper nonvariceal gastrointestinal bleeding. *Aliment Pharmacol Ther* 2011;33(9):1010–8.
46 Chassin MR, Loeb JM, Schmaltz SP, Wachter RM. Accountability measures – using measurement to promote quality improvement. *New Engl J Med* 2010; 363(7):683–8.
47 Global rating scale (UK). [Online source] 2010. http://www.globalratingscale.com/ (last accessed June 13, 2013).
48 Macintosh D, Dubé C, Hollingworth R, et al. The endoscopy Global Rating Scale – Canada: Development and implementation of a quality improvement tool. *Can J Gastroenterol* 2013;27(2):74–82.
49 Diamond SJ, Enestvedt BK, Jiang Z, et al. Adenoma detection rate increases with each decade of life after 50 years of age. *Gastrointest Endosc* 2011;74(1):135–40.
50 Computer Committee of the American Society for Gastrointestinal Endoscopy (ASGE) (1992) Standard Format and Content of the Endoscopic Procedure Report, ASGE, Oak Brook, IL.
51 Committee of the American Society for Gastrointestinal Endoscopy. Quality improvement of gastrointestinal

endoscopy: guidelines for clinical application. *Gastrointest Endosc* 1999;49(6):842–4.
52 Lieberman DA, Faigel DO, Logan JR, et al. Assessment of the quality of colonoscopy reports: results from a multicenter consortium. *Gastrointest Endosc* 2009;69(3 Pt 2):645–53.
53 Sonnenberg A, Amorosi SL, Lacey MJ, Lieberman DA. Patterns of endoscopy in the United States: analysis of data from the Centers for Medicare and Medicaid Services and the National Endoscopic Database. *Gastrointest Endosc* 2008;67(3):489–96.
54 Auslander JN, Lieberman DA, Sonnenberg A. Lack of seasonal variation in the endoscopic diagnoses of Crohn's disease and ulcerative colitis. *Am J Gastroenterol* 2005;100(10):2233–8.
55 Auslander JN, Lieberman DA, Sonnenberg A. Endoscopic procedures and diagnoses are not influenced by seasonal variations. *Gastrointest Endosc* 2006;63(2):267–72.
56 Boonpongmanee S, Fleischer DE, Pezzullo JC, et al. The frequency of peptic ulcer as a cause of upper-GI bleeding is exaggerated. *Gastrointest Endosc* 2004;59(7):788–94.
57 Cram P, Fendrick AM, Inadomi J, et al. The impact of a celebrity promotional campaign on the use of colon cancer screening: the Katie Couric effect. *Arch Intern Med* 2003;163(13):1601–5.
58 Dekel R, Pearson T, Wendel C, et al. Assessment of oesophageal motor function in patients with dysphagia or chest pain: the Clinical Outcomes Research Initiative experience. *Aliment Pharmacol Ther* 2003;18(11–12):1083–9.
59 Dickman R, Mattek N, Holub J, et al. Prevalence of upper gastrointestinal tract findings in patients with noncardiac chest pain versus those with gastroesophageal reflux disease (GERD)-related symptoms: results from a national endoscopic database. *Am J Gastroenterol* 2007;102(6):1173–9.
60 Diggs NG, Holub JL, Lieberman DA, et al. Factors that contribute to blood loss in patients with colonic angiodysplasia from a population-based study. *Clin Gastroenterol Hepatol* 2011;9(5):415–20.
61 El-Serag HB, Gilger MA, Shub MD, et al. The prevalence of suspected Barrett's esophagus in children and adolescents: a multicenter endoscopic study. *Gastrointest Endosc* 2006;64(5):671–5.
62 Enestvedt BK, Gralnek IM, Mattek N, et al. An evaluation of endoscopic indications and findings related to nonvariceal upper-GI hemorrhage in a large multicenter consortium. *Gastrointest Endosc* 2008;67(3):422–9.
63 Enestvedt BK, Williams JL, Sonnenberg A. Epidemiology and practice patterns of achalasia in a large multi-centre database. *Aliment Pharmacol Ther* 2011;33(11):1209–14.
64 Gerson LB, Michaels L, Ullah N, et al. Adverse events associated with anticoagulation therapy in the periendoscopic period. *Gastrointest Endosc* 2010;71(7):1211–7.
65 Gilger MA, Spearman RS, Dietrich CL, et al. Safety and effectiveness of ketamine as a sedative agent for pediatric GI endoscopy. *Gastrointest Endosc* 2004;59(6):659–63.
66 Gupta M, Holub J, Knigge K, Eisen G. Constipation is not associated with an increased rate of findings on colonoscopy: results from a national endoscopy consortium. *Endoscopy* 2010;42(3):208–12.
67 Harewood GC, Lieberman DA. Prevalence of advanced neoplasia at screening colonoscopy in men in private practice versus academic and Veterans Affairs medical centers. *Am J Gastroenterol* 2003;98(10):2312–6.
68 Harewood GC, Sharma VK, de GP. Impact of colonoscopy preparation quality on detection of suspected colonic neoplasia. *Gastrointest Endosc* 2003;58(1):76–9.
69 Harewood GC, Lieberman DA. Colonoscopy practice patterns since introduction of medicare coverage for average-risk screening. *Clin Gastroenterol Hepatol* 2004;2(1):72–7.
70 Harewood GC, Holub JL, Lieberman DA. Variation in small bowel biopsy performance among diverse endoscopy settings: results from a national endoscopic database. *Am J Gastroenterol* 2004;99(9):1790–4.
71 Harewood GC, Holub JL, Lieberman DA. Biopsy specimen acquisition in patients with newly diagnosed peptic ulcer disease as determined from a national endoscopic database. *Gastrointest Endosc* 2004;59(6):664–9.
72 Harewood GC, Olson JS, Mattek NC, et al. Colonic biopsy practice for evaluation of diarrhea in patients with normal endoscopic findings: results from a national endoscopic database. *Gastrointest Endosc* 2005;61(3):371–5.
73 Harewood GC, Mattek NC, Holub JL, et al. Variation in practice of ileal intubation among diverse endoscopy settings: results from a national endoscopic database. *Aliment Pharmacol Ther* 2005;22(6):571–8.
74 Hersh W. Oregon Health Sciences University's 2-for-1 proposition: the fusion of medical informatics and outcomes research. *MD Comput* 1999;16(5):35–7.
75 Holub JL, Silberg DG, Michaels LC, et al. Acid-related upper endoscopy findings in patients with diabetes versus non-diabetic patients. *Dig Dis Sci* 2010;55(10):2853–9.
76 Iles-Shih L, Collins JF, Holub JL, Lieberman DA. Prevalence of significant neoplasia in FOBT-positive patients on warfarin compared with those not on warfarin. *Am J Gastroenterol* 2010;105(9):2030–4.

77 Ko CW, Riffle S, Michaels L, et al. Serious complications within 30 days of screening and surveillance colonoscopy are uncommon. *Clin Gastroenterol Hepatol* 2010;8(2):166–73.
78 Kovalak M, Lake J, Mattek N, et al. Endoscopic screening for varices in cirrhotic patients: data from a national endoscopic database. *Gastrointest Endosc* 2007; 65(1):82–8.
79 Krishnamurthy C, Hilden K, Peterson KA, et al. Endoscopic findings in patients presenting with dysphagia: analysis of a national endoscopy database. *Dysphagia* 2012; doi: 10.1007/s00455-011-9346-0. Epub 2011 Jun 15.
80 Lieberman D. Disease-specific outcomes assessment for gastroesophageal reflux disease. *Gastrointest Endosc Clin N Am* 1999;9(4):657–63, viii.
81 Lieberman D, Fennerty MB, Morris CD, et al. Endoscopic evaluation of patients with dyspepsia: results from the national endoscopic data repository. *Gastroenterology* 2004;127(4):1067–75.
82 Lieberman D, Moravec M, Holub J, et al. Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography. *Gastroenterology* 2008;135(4):1100–5.
83 Lieberman DA, de Garmo PL, Fleischer DE, et al. Patterns of endoscopy use in the United States. *Gastroenterology* 2000;118(3):619–24.
84 Lieberman DA, de Garmo PL, Fleischer DE, et al. Colonic neoplasia in patients with nonspecific GI symptoms. *Gastrointest Endosc* 2000;51(6):647–51.
85 Lieberman DA, Holub J, Eisen G, et al. Prevalence of polyps greater than 9 mm in a consortium of diverse clinical practice settings in the United States. *Clin Gastroenterol Hepatol* 2005;3(8):798–805.
86 Lieberman DA, Holub J, Eisen G, et al. Utilization of colonoscopy in the United States: results from a national consortium. *Gastrointest Endosc* 2005;62(6): 875–83.
87 Lieberman DA, Holub JL, Moravec MD, et al. Prevalence of colon polyps detected by colonoscopy screening in asymptomatic black and white patients. *JAMA* 2008;300(12):1417–22.
88 Logan JR, Klopfer KC. The use of a standardized terminology for comparison of free text and structured data entry. *Proc AMIA Symp* 2000;512–6.
89 Logan JR, McCashland TM, Lieberman DA. Evaluation of reliability in the use of endoscopic terminology. *Stud Health Technol Inform* 2004;107(Pt 1):396–400.
90 Logan JR, Holub JL, Peters D, Brandstater A. Improving quality in cancer screening: the excellence report for colonoscopy. *AMIA Annu Symp Proc* 2010;2010: 462–6.
91 Lowenfels AB, Williams JL, Holub JL, et al. Determinants of polyp size in patients undergoing screening colonoscopy. *BMC Gastroenterol* 2011;11:101.
92 McCashland T, Brand R, Lyden E, de Garmo P. The time and financial impact of training fellows in endoscopy. CORI Research Project. Clinical Outcomes Research Initiative. *Am J Gastroenterol* 2000;95(11): 3129–32.
93 McCashland TM, Brand R, Lyden E, de Garmo P. Gender differences in colorectal polyps and tumors. *Am J Gastroenterol* 2001;96(3):882–6.
94 Olson JS, Lieberman DA, Sonnenberg A. Practice patterns in the management of patients with esophageal strictures and rings. *Gastrointest Endosc* 2007;66(4): 670–5.
95 Rodriguez S, Mattek N, Lieberman D, et al. Barrett's esophagus on repeat endoscopy: should we look more than once? *Am J Gastroenterol* 2008;103(8):1892–7.
96 Rubenstein JH, Mattek N, Eisen G. Age- and sex-specific yield of Barrett's esophagus by endoscopy indication. *Gastrointest Endosc* 2010;71(1):21–7.
97 Saini SD, Eisen G, Mattek N, Schoenfeld P. Utilization of upper endoscopy for surveillance of gastric ulcers in the United States. *Am J Gastroenterol* 2008; 103(8):1920–5.
98 Sharma VK, Nguyen CC, Crowell MD, et al. A national study of cardiopulmonary unplanned events after GI endoscopy. *Gastrointest Endosc* 2007;66(1):27–34.
99 Shumaker DA, de Garmo P, Faigel DO. Potential impact of preoperative EUS on esophageal cancer management and cost. *Gastrointest Endosc* 2002;56(3):391–6.
100 Thakkar K, El-Serag HB, Mattek N, Gilger MA. Complications of pediatric EGD: a 4-year experience in PEDS-CORI. *Gastrointest Endosc* 2007;65(2): 213–21.
101 Varadarajulu S, Eloubeidi MA, Patel RS, et al. The yield and the predictors of esophageal pathology when upper endoscopy is used for the initial evaluation of dysphagia. *Gastrointest Endosc* 2005;61(7):804–8.
102 Beaulieu D, Barkun A, Martel M. A quality audit of colonoscopy reports amongst patients screened for colorectal neoplasia (CRN). *Can J Gastroenterol* 2010;24(Suppl A). [Abstract]
103 Centers for Disease Control and Prevention (CDC). Colorectal cancer test use among persons aged > or = 50 years – United States, 2001. *MMWR Morb Mortal Wkly Rep* 2003;52(10):193–6.
104 Meissner HI, Breen N, Klabunde CN, Vernon SW. Patterns of colorectal cancer screening uptake among men and women in the United States. *Cancer Epidemiol Biomarkers Prev* 2006;15(2):389–94.
105 Centers for Disease Control and Prevention (CDC). Vital signs: colorectal cancer screening among adults

aged 50–75 years – United States, 2008. *MMWR Morb Mortal Wkly Rep* 2010;59:808–12. [Abstract]
106 Ananthakrishnan AN, Schellhase KG, Sparapani RA, et al. Disparities in colon cancer screening in the Medicare population. *Arch Intern Med* 2007;167(3):258–64.
107 Thornton JG, Morris AM, Thornton JD, et al. Racial variation in colorectal polyp and tumor location. *J Natl Med Assoc* 2007;99(7):723–8.
108 Agrawal S, Bhupinderjit A, Bhutani MS, et al. Colorectal cancer in African Americans. *Am J Gastroenterol* 2005;100(3):515–23.
109 Weber TK, Chin HM, Rodriguez-Bigas M, et al. Novel hMLH1 and hMSH2 germline mutations in African Americans with colorectal cancer. *JAMA* 1999;281(24):2316–20.
110 Muller AD, Sonnenberg A. Prevention of colorectal cancer by flexible endoscopy and polypectomy. A case-control study of 32,702 veterans. *Ann Intern Med* 1995;123(12):904–10.
111 Singh H, Turner D, Xue L, et al. Risk of developing colorectal cancer following a negative colonoscopy examination: evidence for a 10-year interval between colonoscopies. *JAMA* 2006;295(20):2366–73.
112 Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. *The National Polyp Study Workgroup*. *New Engl J Med* 1993;329(27):1977–81.
113 Baxter NN, Goldwasser MA, Paszat LF, et al. Association of colonoscopy and death from colorectal cancer. *Ann Intern Med* 2009;150(1):1–8.
114 Brenner H, Chang-Claude J, Seiler CM, et al. Protection from colorectal cancer after colonoscopy: a population-based case-control study. *Ann Intern Med* 2011;154(1):22–30.
115 Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2008;149(9):627–37.
116 Davila RE, Rajan E, Baron TH, et al. ASGE guideline: colorectal cancer screening and surveillance. *Gastrointest Endosc* 2006;63(4):546–57.
117 Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* 2008;58(3):130–60.
118 Rex DK, Johnson DA, Anderson JC, et al. American College of Gastroenterology guidelines for colorectal cancer screening 2009 [corrected]. *Am J Gastroenterol* 2009;104(3):739–50.
119 Burt RW, Barthel JS, Dunn KB, et al. NCCN clinical practice guidelines in oncology. Colorectal cancer screening. *J Natl Compr Canc Netw* 2010;8(1):8–61.
120 Seeff LC, Nadel MR, Klabunde CN, et al. Patterns and predictors of colorectal cancer test use in the adult U.S. population. *Cancer* 2004;100(10):2093–103.
121 Zauber A, Winawer SJ, Lansdrop-Vogelaar I, et al. Effect of initial polypectomy versus surveillance polypectomy on colorectal cancer mortality reduction: micro-simulation modeling of the National Polyp Study. *Am J Gastroenterol* 2007;102. [Abstract]
122 Vogelaar I, van BM, Schrag D, et al. How much can current interventions reduce colorectal cancer mortality in the U.S.? Mortality projections for scenarios of risk-factor modification, screening, and treatment. *Cancer* 2006;107(7):1624–33.
123 Brenner H, Chang-Claude J, Seiler CM, et al. Case-control study supports extension of surveillance interval after colonoscopic polypectomy to at least 5 yr. *Am J Gastroenterol* 2007;102(8):1739–44.
124 Brenner H, Chang-Claude J, Seiler CM, Hoffmeister M. Long-term risk of colorectal cancer after negative colonoscopy. *J Clin Oncol* 2011;29(28):3761–7.
125 Lieberman DA, Weiss DG, Bond JH, et al. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380. *New Engl J Med* 2000;343(3):162–8.
126 Regula J, Rupinski M, Kraszewska E, et al. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. *New Engl J Med* 2006;355(18):1863–72.
127 Schoenfeld P, Cash B, Flood A, et al. Colonoscopic screening of average-risk women for colorectal neoplasia. *New Engl J Med* 2005;352(20):2061–8.
128 Imperiale TF, Wagner DR, Lin CY, et al. Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. *New Engl J Med* 2000;343(3):169–74.
129 Harewood GC, Petersen BT, Ott BJ. Prospective assessment of the impact of feedback on colonoscopy performance. *Aliment Pharmacol Ther* 2006;24(2):313–8.
130 Benin AL, Vitkauskas G, Thornquist E, et al. Validity of using an electronic medical record for assessing quality of care in an outpatient setting. *Med Care* 2005;43(7):691–8.
131 Tang PC, Ralston M, Arrigotti MF, et al. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. *J Am Med Inform Assoc* 2007;14(1):10–5.
132 Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. *New Engl J Med* 2009;360(16):1628–38.

133 DesRoches CM, Campbell EG, Rao SR, et al. Electronic health records in ambulatory care – a national survey of physicians. *New Engl J Med* 2008;359(1):50–60.
134 Jha AK, DesRoches CM, Kralovec PD, Joshi MS. A progress report on electronic health records in U.S. hospitals. *Health Aff (Millwood)* 2010;29(10): 1951–7.
135 Jha AK, Ferris TG, Donelan K, et al. How common are electronic health records in the United States? A summary of the evidence. *Health Aff (Millwood)* 2006;25(6):w496–w507.
136 Lacchetti C, Guyatt G. (2002) Surprising results of randomized controlled trials, in *Users' Guide to the Medical Literature* (eds. G Guyatt, R Drummond), AMA Press, Chicago, IL, pp. 247–65.

Answers to multiple choice questions

1. D
2. E
3. A

19 Epidemiology of colorectal carcinoma

Harminder Singh[1], *Joselito M. Montalban*[2], *& Salaheddin Mahmud*[3]

[1]Departments of Internal Medicine and Community Health Sciences, University of Manitoba; Division of Gastroenterology, University of Manitoba IBD Clinical and Research Centre; Department of Medical Oncology and Haematology, CancerCare Manitoba; Winnipeg, MB, Canada

[2]Departments of Internal Medicine and Community Health Sciences, University of Manitoba, Winnipeg, MB, Canada

[3]Community Health Sciences, University of Manitoba; Epidemiology, CancerCare Manitoba, Winnipeg, MB, Canada

Key points

- Worldwide, colorectal cancer is the third most common cancer, with an increasing number of cases diagnosed in both developed and developing countries.
- Risk factors which markedly increase the risk for developing colorectal cancer include family history of colorectal cancer, certain inherited syndromes, prior personal history of colorectal cancer, and inflammatory bowel disease. However, less than 30 % of individuals developing colorectal cancer have any of these very high risk factors.
- There are several well-established modalities for colorectal cancer screening. A more widespread use of these modalities should lead to reduction in incidence and mortality due to this disease.
- There is an urgency to improve detection and screening for right colonic colorectal cancer, as all of the current colon cancer screening modalities have lower effectiveness in reducing risk of right-sided colorectal cancer.

Epidemiology of colorectal cancer (CRC)

Worldwide, CRC is the third most common cancer, following lung and breast cancers. In 2008, CRC was responsible for 1.2 million out of 12.7 million new cancer cases that occurred that year. Worldwide, it is the fourth most common cause of death from cancer, accounting for 608,000 deaths or 8 % of all cancer deaths in 2008 [1]. CRC continues to be a disease of the developed world where it remains the second most common cause of cancer-related deaths (Figure 19.1). However, the age-standardized CRC incidence and mortality rates have been declining in most developed countries over the last two decades, with the most marked changes occurring in the United States (USA) [2]. On the other hand, the CRC incidence rates have been rising in developing countries [3]. With aging of the population, the actual number of individuals diagnosed with CRC is expected to continue to increase even in developed countries [4].

An interesting change in the colonic site localization of CRC has been reported by several studies – a higher proportion of CRC now occur in the proximal or right part of the colon, than was the case a few decades ago. This has implications for CRC screening,

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

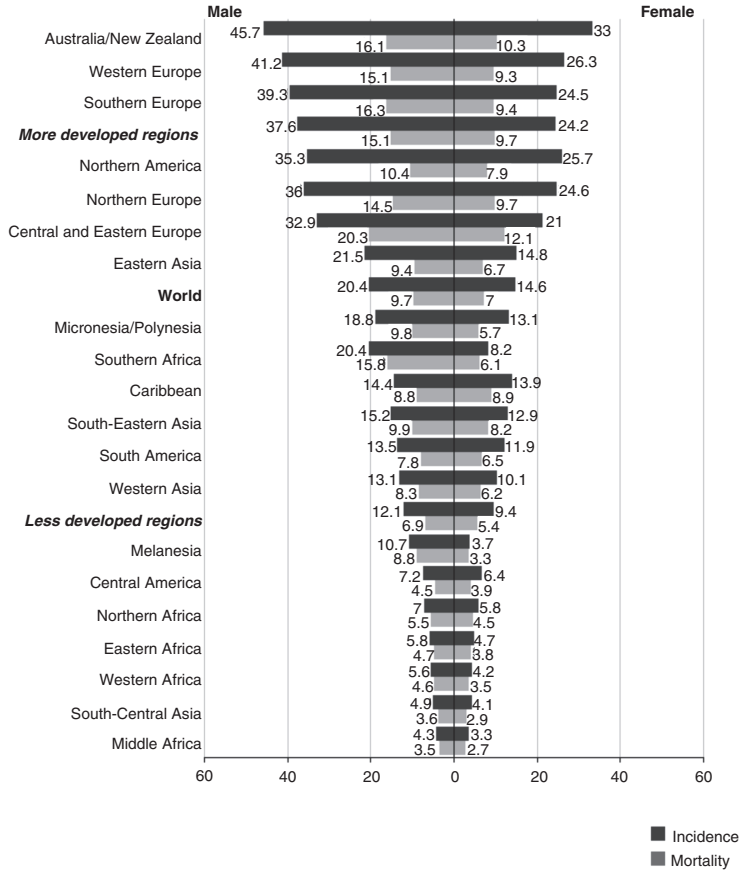


Figure 19.1 Estimated age-standardized incidence and mortality rates for colorectal cancer, 2008. Source: Ferlay et al. 2010 [1]. Reproduced with permission of John Wiley & Sons.

as all modalities of screening for CRC have a lower effectiveness for detecting and preventing right-sided CRC [5]. As a result, currently there is an enormous interest in enhancing the detection and prevention of right-sided CRC.

Etiopathogenesis of CRC

Over 95 % of CRCs are adenocarcinomas, which arise from the normal epithelium lining the colonic mucosa. The development of most colonic adenocarcinomas is thought to involve the intermediate precursor lesions of adenomatous polyps (adenoma-carcinoma sequence). The adenomatous polyps, which are further down this transformation pathway, are known as advanced adenomas or advanced neoplasia (AN) and are typically defined as adenomatous polyps larger than 1 cm in size or ones with villous histology or high-grade dysplasia. Multiple sequential genetic changes are accumulated and drive this transformation. As per the Vogelstein model proposed in 1990, the molecular basis of CRC is a multistep process in which each genetic change leads to a growth advantage in the modified cells.

Recently, an alternate pathway called serrated polyp pathway has been recognized in which the precursor lesions are believed to be serrated polyps [6]. Morphologically, adenomatous and serrated polyps are distinct, with adenomatous polyps characterized by their dysplastic changes and serrated polyps by their architectural alterations. It has been hypothesized that some of the serrated polyps may have more rapid growth and could be responsible for a large proportion of CRCs that are diagnosed soon after colonoscopy (interval CRCs) [6]. On colonoscopy, some of the serrated polyps (sessile serrated polyps (SSPs)) are flatter lesions, often with an overlying

mucus plug that might conceal them so that these lesions may be overlooked by endoscopists. A higher proportion of SSPs occurs in the right side of the colon, and a higher proportion of interval CRCs occurs in the right colon.

In terms of the involved molecular pathways, there are at least three distinct molecular pathways to CRC [6]. The chromosomal instability pathway is driven by mutational changes in oncogenes and tumor suppressor genes, the microsatellite instability pathway by mutations in one of the DNA mismatch repair genes and the epigenetic pathway by hypermethylation-induced silencing of tumor suppressor-like genes and/or DNA mismatch repair genes. The mutations can be inherited or acquired and underlie the inherited CRC syndromes and sporadic CRC cases, respectively. Identification of specific mutations in the cancer cells shed into the fecal specimens is the basis of the fecal DNA tests, an emerging method of CRC screening.

Clinical manifestations

Most symptoms, including bowel habit changes and abdominal pain, have poor sensitivity and specificity for CRC. Although dark red rectal bleeding and the presence of palpable abdominal mass may have the highest specificity, less than 15 % of individuals with CRC present with these symptoms [7]. The lack of sensitive symptoms is one of the rationales for screening asymptomatic individuals for CRC. According to recent guidelines from the British Society of Gastroenterology [8], all individuals (other than menstruating women) presenting with iron deficiency anemia (and no significant overt nongastrointestinal blood loss) should be investigated for gastrointestinal sources of blood loss, including CRC.

Risk factors

Age and gender

The incidence and mortality rates of CRC rise with increasing *age*, with exponential increases starting in the 50s – over 90 % of CRC are diagnosed in individuals over the age of 50 and the median age at diagnosis is in the early 70s [9]. At any particular age in the developed world, *men* have a much higher risk of being diagnosed with CRC compared to women – the age-standardized incidence rate of CRC in women is approximately two-thirds that in men [9]. Men also have a higher prevalence of the high-risk precursor lesion AN. In a meta-analysis [10] of 17 studies involving almost 925,000 participants, men were found to have 83 % greater risk of AN than women, and the risk was higher at all age groups studied (40–70 or older). However, women tend to live longer than men. Hence, the lifetime risk of CRC is very similar in men and women. In Canada, based on current estimates, one in 13 men and one in 16 women are expected to develop CRC in their lifetime, and one in 28 men and one in 32 women are expected to die from it [9]. This emphasizes the need for equal attention for prevention and screening in both genders to reduce the burden of this disease.

Ethnicity

African Americans in the United States have a 20 % higher incidence and a 50 % higher mortality due to CRC [11,12]. They also tend to be diagnosed at an advanced stage of the disease. The differences in rates of CRC screening and healthcare utilization are likely contributing to some of these disparities.

Family history and hereditary syndromes

Family history of CRC is one of the strongest risk factors for the development of CRC and is present in as many as 20–30 % of all CRC cases [13]. Approximately 2–5 % of CRC cases occur in the setting of well-defined inherited syndromes, including Lynch syndrome (also known as hereditary nonpolyposis CRC (HNPCC)) (2–4 % of all CRC), familial adenomatous polyposis (FAP) (~1 % of all CRC), and MUTYH-associated polyposis (MAP) (<1 % of CRC) [13]. The lifetime risk of CRC among individuals with HNPCC is estimated to be 50–80 % and in those with FAP and intact colon is close to 100 %. Both HNPCC and FAP are inherited as autosomal dominant conditions, and are associated with a variety of nonintestinal manifestations. FAP is characterized by development of multiple adenomatous polyps at a young age, and HNPCC by rapid development of CRC. For individuals with HNPCC, screening with colonoscopy is generally recommended to be initiated by 20–25 years

of age and repeated every 1–2 years. For those with FAP, flexible sigmoidoscopy is recommended every 1–2 years starting at age 10–12 years.

Among the individuals with family history of CRC (and no known hereditary syndromes), the risk of CRC starts to increase about a decade earlier than among those without such history [14]. Therefore, it is recommended that screening for CRC should be started a decade earlier among those with a family history of CRC. However, if only second- and third-degree relatives are affected, the risk is very similar to those without a family history of CRC [15]. The risk increases with increasing number of family members affected and the younger the age of those affected with CRC. As an example, the risk has been estimated to be increased twofold among all those with one or more affected first-degree relatives. The specific recommendations for screening among those with family history of CRC vary in different jurisdictions. As an example, in North America, it is recommended that individuals with one first-degree relative affected with CRC at age less than 60 years or two or more first-degree relatives with CRC should be screened with colonoscopy every 5 years starting at age 40 or when 10 years younger than the youngest affected person, whichever is earlier [13].

Personal history of colorectal polyps and CRC

Individuals with prior history of AN or CRC are at a three- to sixfold increased risk of developing additional AN or CRC [16]. Hence, regular surveillance colonoscopy at shortened intervals is recommended for such individuals.

Medical conditions, including inflammatory bowel disease

Inflammatory bowel disease (IBD) is another strong risk factor for development of CRC. Although IBD accounts for only 1–2 % of cases with CRC, individuals with IBD colitis have an approximately sixfold increased risk of CRC [17]. Risk factors for development of CRC among individuals with IBD include the occurrence of pseudopolyps, coexistence of primary sclerosing cholangitis, family history of sporadic CRC, young age at onset of colitis, and extent and duration of colonic disease. There is no increased risk among those with inflammation limited to the rectum. The risk begins to increase approximately 10 years after the onset of the colitis and has been reported to be as high as 30 % after 35 years of pancolitis-colitis involving the entire colon. Regular surveillance colonoscopies at annual or biannual basis are recommended starting at 8–10 years after the onset of colitis.

Although, individuals with coronary artery disease and diabetes have been reported to have an up to twofold increased risk of AN and/or CRC, it has been suggested that this association may be due to concomitant confounding conditions and/or shared risk factors such as smoking, obesity, and physical inactivity [18].

Dietary and lifestyle risk factors

There are several dietary and lifestyle risk factors that have been associated with CRC. These were recently examined in detail in an excellent review by Chan and Giovanucci [19]. Western diet consisting of high intake of red and processed meats, high-fat dairy products, highly refined grains and starches, and sugars has been related to increased risk of CRC. However, the role of specific nutritional elements has not been consistently demonstrated, and it is likely that the dietary pattern as a whole is more important than the individual components. Although there have been concerns that food fortification with folic acid (to prevent neural tube defects) could increase the risk of CRC, several recent studies suggest that folic acid administered at moderate doses (such as by food fortification) does not increase the risk of CRC [20]. It is still unclear whether dietary supplementation with calcium and vitamin D or increased fiber intake could prevent CRC [21].

Lifestyle factors that have been associated with increased risk of CRC include cigarette smoking, alcohol intake, obesity, and physical inactivity. The evidence is most consistent for smoking and physical inactivity. In the United States, it has been estimated that 15–20 % of CRCs could be attributed to smoking. There is an approximately 30- to 40-year time lag between exposure to cigarette smoking and CRC detection. One meta-analysis found a significant effect of dose (38 % increase risk for an increment of 40 cigarettes per day; 51 % increase risk for an increment of 60 pack-years) and duration (20 % increase risk for an increment of 40 cigarettes per day) as well as that of age of initiation (4 % decrease in risk for each 10-year delay in initiation of smoking) [22].

Worldwide, it is estimated that physical inactivity causes about 10–16 % of cases of CRCs [23]. Physically active individuals have a 20–30 % lower risk of adenomatous polyps [24] and CRC [25] than less active individuals. Moderate levels of activity, leisure-time or occupational activities are all associated with decreased risk.

The association of obesity with CRC is stronger among men than among women. Indeed, there are studies that found no association between body mass index and risk of CRC among women [26,27], including a meta-analysis which corrected for publication bias [28].

A pooled analysis of 8 cohort studies had found that more than a moderate amount ($\geq$30 g d^{-1} or 2 drinks per day) of alcohol intake increased the risk of CRC (relative risk (RR) 1.24; 95 % confidence interval (CI) 1.07–1.42) [29]. However, other recent studies have found a decreased risk of CRC with moderate alcohol intake [30].

Primary prevention

Primary prevention of CRC aims to reduce the development of CRC and relies on the modification of the dietary and lifestyle risk factors discussed in the previous section [19]. Since many of these proven and putative risk factors for CRC are also risk factors for other diseases, such as coronary artery disease, modification of these factors would lead to improvement in overall health.

Another potentially attractive strategy for reducing CRC incidence and mortality is the use of pharmaceutical agents to prevent CRC development and/or mortality, that is, chemoprevention. The evidence is most conclusive for use of aspirin and other nonsteroidal anti-inflammatory drugs and for postmenopausal use of hormone replacement therapy; however, because of the potential side effects associated with these agents, they are not recommended for use in primary prevention of CRC. Several other agents have been explored including statins, bisphosphonates, angiotensin-converting enzyme inhibitors, combination of difluoromethylornithine (DFMO), and sulindac.

Secondary prevention: screening

Secondary prevention refers to interventions (e.g. screening) that aim to detect diseases at their early stages so as to improve intervention outcomes. Traditionally, cancer screening aims to reduce the mortality due to specific cancers. For some cancers, such as CRC and cervical cancer, screening can also lead to reduction in the incidence of the cancer through early detection and removal of the precursor lesions. Colorectal cancer can be prevented by the detection and removal of the pre-invasive polyps, and once cancerous, survival is significantly improved when diagnosis is made early, while the lesion is still localized [31]. Screening for both CRC and its adenoma precursors, therefore, is acknowledged as an effective way to reduce mortality due to CRC [32].

There are several tests for CRC screening among average-risk individuals (those without familial/hereditary risk factors or prior personal history of CRC, AN, or IBD). The most commonly used tests can be categorized as fecal specimen-based tests and structural evaluation of the colon. The former, which includes fecal occult blood test (FOBT), examines the stool for evidence of blood or other markers being shed from CRCs and is used mainly for cancer detection. The structural approach is for both cancer detection and prevention and includes optical colonoscopy (OC), flexible sigmoidoscopy (FS), double-contrast barium enema (DCBE), and computed tomography colonography (CTC) [31,32]. Barium enema may not detect up to half of the large polyps and hence is no longer widely used for CRC screening [33].

Many countries have recently instituted organized national or regional screening programs for CRC screening among average-risk individuals [34]. In most countries, less invasive screening methods such as different versions of FOBT are performed first (usually on a biannual basis), followed by colonoscopy in those with positive results on the initial test. On the other hand, in a few countries such as the United States, Poland, and Germany, colonoscopy is often the first-line method used [35]. The United Kingdom is the first and only country so far to announce a national FS screening program.

The most commonly used FOBTs can be divided into hemoccult FOBTs and fecal immunochemical test (FITs). Hemoccult tests rely on the pseudoperoxidase activity of hemoglobin in stool and are also referred to as guaiac FOBTs (gFOBTs). The FITs detect human globin, thereby improving test characteristics.

There are several advantages for FOBT. It is noninvasive, can be performed in the privacy of a patient's own home and has no direct physical side effects. Its disadvantages include the need for repeat testing (annually or biennially) and limited ability to detect adenomas, thereby resulting in limited efficacy in preventing development of CRC. Four large randomized controlled trials (RCTs) evaluated the efficacy of the older version of the guaiac test, Hemoccult II. The results from these tests were pooled in a Cochrane meta-analysis [36], which suggested that biennial screening with FOBT may lead to a 15 % relative reduction in CRC mortality in an intention-to-treat (ITT) analysis, and a 25 % reduction when the analyses were adjusted for attendance for screening. Hemoccult Sensa, a newer version of the guaiac test, has improved sensitivity in detecting CRC. However, it has a higher false-positive rate, and is less well studied compared to the Hemoccult II. The accuracy of the newer FOBTs was the subject of a recent systematic review for the U.S. Preventive Services Task Force [37]. The review concluded that Hemoccult II was less sensitive than FIT for cancer detection and that FIT had sensitivity similar to, or less than, that of Hemoccult Sensa. The specificity of Hemoccult Sensa was reported to be less than that of FIT, which had specificity similar to that of Hemoccult II. The review noted, however, that there are few studies directly comparing different FITs with each other or with regular or high-sensitivity hemoccult tests (Hemoccult Sensa). There is no direct evidence on efficacy of Hemoccult Sensa or FITs on preventing CRC mortality or incidence. However, mathematical modeling studies, using the test characteristics derived from one-time performance of these tests, predict that FITs maybe more cost-effective than endoscopic methods for screening for CRC.

There are four ongoing RCTs of FS for CRC screening. Three European trials [38–40] are evaluating once-in-lifetime FS around the age of 60, and the fourth, a US trial [41] FS once every 5 years. The initial results from these trials have established the efficacy of FS in reducing incidence and mortality due to CRC. For example, the large UK trial, after a median follow-up of 11.2 years, reported a 31 % reduction in CRC mortality in ITT analysis and a 43 % reduction in CRC mortality among those actually having the FS when adjusted for self-selection bias in the FS group. The trial also reported a 23 % reduction in incidence of CRC in ITT analysis and a 33 % CRC incidence reduction in those having the FS. It is remarkable that there was not much attenuation in the beneficial effects of initial FS in the later years of follow-up reported in this trial so far – a result that is consistent with earlier observational studies that found no attenuation in effect after follow-up of as long as 10–16 years after sigmoidoscopy [42,43]. Therefore, more recent guidelines recommend that individuals with a normal FS do not need a repeat screening FS for at least 10 years [33], and indeed the UK will be offering once-in-lifetime FS in its screening program. Of note, in three of these trials, FS was not reported to have any significant effect on right-sided CRC.

The efficacy of OC in reducing the incidence and mortality of CRC has never been proven by RCTs. However, there is considerable indirect evidence on effectiveness of OC in reducing CRC incidence and mortality. For example, there is now RCT-level evidence for efficacy of FS for CRC screening [40,44], and since OC examines more segments of the colon, it is reasonable to assume that OC will be shown in RCTs to be at least as effective as FS, if not more. The results of OC cohort studies do suggest that an FS-only screening strategy would fail to detect 21–65 % of right-sided ANs, which were detected on OC [44]. On the other hand, OC requires greater operator skill and a more intensive bowel cleansing regimen, and therefore more manpower resources, than does FS. OC is also more costly, requires more sedation, and has a higher attendant risk of complications, including bleeding, perforation, and death.

Several recent studies have raised questions on the magnitude of incremental benefit of OC over FS when colonoscopy is performed in the usual clinical practice [45–48]. The limited effectiveness of OC in usual clinical practice for right-sided CRC is likely due to an interplay of the biology of the tumors and technical performance of colonoscopy. Lesions in the right colon are more often endoscopically subtle and therefore may not be detected by some endoscopists, especially those who have a lower rate of reaching the end of the right side of the colon [49]. One of the most disconcerting new evidence in the area of CRC screening is the wide variation in performance of OC by different operators. Irrespective of the initial test, OC is a vital test for CRC screening, and so there is

a renewed focus now to improve the performance of OC in usual practice.

Other less widely used and rapidly emerging techniques for CRC screening include computed tomography colonography (also called virtual colonoscopy) and stool DNA test.

Conclusions

In conclusion, CRC is one of the most common cancers with several well-defined risk factors and modalities for screening and early detection. Additional advances in screening techniques and management of this disease are expected to lead to further reductions in disease burden due to this disease. However, quality assurance and quality enhancement activities are extremely important to avoid unintended consequences of CRC screening, such as missed lesions and complications due to the procedures performed.

Multiple choice questions

1 Which of the following measures has the highest effectiveness in preventing development of right-sided colon cancer?

A Colonoscopy
B Flexible sigmoidoscopy
C Fecal immunochemical test
D Chemoprevention
E Not known

2 The following have increased risk of colorectal cancer, except . . .

A Individuals with inflammatory bowel disease limited to the rectum
B Individuals on a Western diet
C Individuals who are physically active
D A and C
E None of the above

3 Which statement(s) is/are true regarding the epidemiology of colorectal cancer?

A Colorectal cancer is traditionally a disease of the developed world. As such, its incidence rates have been rising over the last decade or two in most of the developed countries
B A higher proportion of colorectal cancers now occur in the distal part of the colon, as compared to the colonic site distribution a few decades ago
C Colorectal cancer is traditionally a disease of the developed world; nevertheless, its incidence rates have been rising in the developing countries
D A and C
E All of the above

4 Which statement(s) is/are true regarding prevention of colorectal cancer?

A Primary prevention refers to activities that aim to detect diseases at their early stages so as to improve intervention outcomes
B Barium enema is no longer widely used for colorectal cancer screening
C Screening is one measure that is in line with primary prevention
D A and C
E None of the above

5 Which statement(s) is/are true regarding the etiopathogenesis of colorectal cancer?

A Mutational changes leading to colorectal cancer are always inherited
B The molecular basis of colorectal cancer is a multistep process
C Sessile serrated polyps are easier to detect than the adenomatous polyps and occur more often in the distal colon, accounting for the higher effectiveness of screening modalities in detecting left-sided colorectal cancer
D A and C
E All of the above

6 Which statement(s) is/are true regarding risk factors for colorectal cancer?

A It is recommended that an individual in North America having two first-degree relatives with colorectal cancer, one of whom was 45 years old at the time of colorectal cancer diagnosis, should undergo colonoscopy every 5 years starting at age 40
B The risk of colorectal cancer is very similar between individuals without a family history of the disease versus those who have only second- and third-degree relatives affected
C The lifetime risk of colorectal cancer is higher in men than in women
D A and C
E All of the above

References

1 Ferlay J, Shin HR, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127(12):2893–917.

2 Vital signs: Colorectal cancer screening, incidence, and mortality – United States, 2002–2010. *MMWR* 2011;60(26):884–9.

3 Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev* 2010;19(8):1893–907.

4 Nowatzki J, Moller B, Demers A. Projection of future cancer incidence and new cancer cases in Manitoba, 2006–2025. *Chronic Diseases in Canada* 2011;31(2):71–8.

5 Haug U, Knudsen AB, Brenner H, Kuntz KM. Is fecal occult blood testing more sensitive for left- versus right-sided colorectal neoplasia? A systematic literature review. *Exp Rev Mol Diagn* 2011;11(6):605–16.

6 Snover DC. Update on the serrated pathway to colorectal carcinoma. *Human Pathol* 2011;42(1):1–10.

7 Ford AC, Veldhuyzen van Zanten SJ, Rodgers CC, et al. Diagnostic utility of alarm features for colorectal cancer: systematic review and meta-analysis. *Gut* 2008;57(11):1545–53.

8 Goddard AF, James MW, McIntyre AS, Scott BB. Guidelines for the management of iron deficiency anaemia. *Gut* 2011;60(10):1309–16.

9 Canadian Cancer Society's Steering Committee (2010) Canadian cancer statistics 2010, Canadian Cancer Society, Toronto.

10 Nguyen SP, Bent S, Chen YH, Terdiman JP. Gender as a risk factor for advanced neoplasia and colorectal cancer: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2009;7(6):676–81, e1–3.

11 Polite BN, Dignam JJ, Olopade OI. Colorectal cancer and race: understanding the differences in outcomes between African Americans and whites. *Med Clin North Am* 2005;89(4):771–93.

12 Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 2010;116(3):544–73.

13 Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* 2010;138(6):2044–58.

14 Brenner H, Hoffmeister M, Haug U. Family history and age at initiation of colorectal cancer screening. *Am J Gastroenterol* 2008;103(9):2326–31.

15 Taylor DP, Burt RW, Williams MS, et al. Population-based family history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology* 2010;138(3):877–85.

16 Atkin WS, Morson BC, Cuzick J. Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. *New Engl J Med* 1992;326(10):658–62.

17 Mattar MC, Lough D, Pishvaian MJ, Charabaty A. Current management of inflammatory bowel disease and colorectal cancer. *Gastrointest Cancer Res* 2011;4(2):53–61.

18 Kahi CJ, Rex DK, Imperiale TF. Screening, surveillance, and primary prevention for colorectal cancer: a review of the recent literature. *Gastroenterology* 2008;135(2):380–99.

19 Chan AT, Giovannucci EL. Primary prevention of colorectal cancer. *Gastroenterology* 2010;138(6):2029–43, e10.

20 Lee JE, Chan AT. Fruit, vegetables, and folate: cultivating the evidence for cancer prevention. *Gastroenterology* 2011;141(1):16–20.

21 Ross AC, Manson JE, Abrams SA, et al. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J Clin Endocrinol Metab* 2011;96(1):53–8.

22 Liang PS, Chen TY, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. *Int J Cancer* 2009;124(10):2406–15.

23 World Health Organization (WHO) (2002) *The World Health Report 2002: Reducing Risks, Promoting Healthy Life*, WHO, Geneva.

24 Wolin KY, Yan Y, Colditz GA. Physical activity and risk of colon adenoma: a meta-analysis. *Br J Cancer* 2011;104(5):882–5.

25 Wolin KY, Yan Y, Colditz GA, Lee IM. Physical activity and colon cancer prevention: a meta-analysis. *Br J Cancer* 2009;100(4):611–6.

26 Bassett JK, Severi G, English DR, et al. Body size, weight change, and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* 2010;19(11):2978–86.

27 Reeves GK, Pirie K, Beral V, et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ* 2007;335(7630):1134.

28 Moghaddam AA, Woodward M, Huxley R. Obesity and risk of colorectal cancer: a meta-analysis of 31 studies with 70,000 events. *Cancer Epidemiol Biomarkers Prev* 2007;16(12):2533–47.

29 Cho E, Smith-Warner SA, Ritz J, et al. Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. *Ann Intern Med* 2004;140(8):603–13.

30 Crockett SD, Long MD, Dellon ES, et al. Inverse relationship between moderate alcohol intake and rectal cancer:

analysis of the North Carolina Colon Cancer Study. *Dis Colon Rectum* 2011;54(7):887–94.
31 Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Cancer* 2008;58(3):130–60.
32 Van Gossum A. Guidelines for colorectal cancer screening – a puzzle of tests and strategies. *Acta Clin Belg* 2010;65(6):433–6.
33 Leddin DJ, Enns R, Hilsden R, et al. Canadian Association of Gastroenterology position statement on screening individuals at average risk for developing colorectal cancer: 2010. *Can J Gastroenterol* 2010;24(12):705–14.
34 Kennedy DA, Stern SJ, Moretti M, et al. Folate intake and the risk of colorectal cancer: a systematic review and meta-analysis. *Cancer Epidemiol* 2011;35(1):2–10.
35 Pizzo E, Pezzoli A, Stockbrugger R, et al. Screenee perception and health-related quality of life in colorectal cancer screening: a review. *Value Health* 2011;14(1):152–9.
36 Hewitson P, Glasziou P, Irwig L, et al. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database Syst Rev* (Online) 2007(1):CD001216.
37 Whitlock EP, Lin JS, Liles E, et al. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* 2008;149(9):638–58.
38 Segnan N, Armaroli P, Bonelli L, et al. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian randomized controlled trial – SCORE. *J Natl Cancer Inst* 2011;103(17):1310–22.
39 Hoff G, Grotmol T, Skovlund E, Bretthauer M. Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. *BMJ* 2009;338:b1846.
40 Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet* 2010;375(9726):1624–33.
41 Weissfeld JL, Schoen RE, Pinsky PF, et al. Flexible sigmoidoscopy in the PLCO cancer screening trial: results from the baseline screening examination of a randomized trial. *J Natl Cancer Inst* 2005;97(13):989–97.
42 Newcomb PA, Storer BE, Morimoto LM, et al. Long-term efficacy of sigmoidoscopy in the reduction of colorectal cancer incidence. *J Natl Cancer Inst* 2003;95(8):622–5.
43 Selby JV, Friedman GD, Quesenberry CP, Jr., Weiss NS. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *New Engl J Med* 1992;326(10):653–7.
44 Canadian Partnership Against Cancer, Expert Panel on Flexible Sigmoidoscopy (2010) Flexible Sigmoidoscopy Watching Brief. 2nd Iteration of Expert Panel Report, Canadian Partnership Against Cancer, Toronto.
45 Baxter NN, Goldwasser MA, Paszat LF, et al. Association of colonoscopy and death from colorectal cancer. *Ann Intern Med* 2009;150(1):1–8.
46 Brenner H, Hoffmeister M, Arndt V, et al. Protection from right- and left-sided colorectal neoplasms after colonoscopy: population-based study. *J Natl Cancer Inst* 2010;102(2):89–95.
47 Lakoff J, Paszat LF, Saskin R, Rabeneck L. Risk of developing proximal versus distal colorectal cancer after a negative colonoscopy: a population-based study. *Clin Gastroenterol Hepatol* 2008;6(10):1117–21; quiz 064.
48 Singh H, Nugent Z, Mahmud SM, et al. Predictors of colorectal cancer after negative colonoscopy: a population-based study. *Am J Gastroenterol* 2010;105(3):663–73; quiz 74.
49 Rex DK. Preventing colorectal cancer and cancer mortality with colonoscopy: what we know and what we don't know. *Endoscopy* 2010;42(4):320–3.

Answers to multiple choice questions

1. A
2. D
3. C
4. B
5. B
6. B

20 Epidemiology of irritable bowel syndrome

Rok Seon Choung[1] *& Yuri A. Saito*[2]

[1]Department of Internal Medicine, Institute of Digestive Diseases and Nutrition, Korea University, Seoul, South Korea

[2]Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

Key points

- About 10 % of the population has IBS at any one time and about 200 people per 100,000 will receive an initial diagnosis of IBS over the course of a year.
- Environmental factors – such as diet, stress, abuse, and infections – have clear links to IBS development or exacerbation, yet the pathophysiology of IBS still remains poorly understood.
- IBS results in significant work absenteeism, decreased productivity, and impaired health-related quality of life, and results in high direct and indirect healthcare costs.

Clinical summary

Irritable bowel syndrome (IBS) is a common and chronic functional gastrointestinal disorder characterized by recurrent abdominal pain or discomfort associated with altered bowel habits, including symptoms of diarrhea, constipation, or both. Typically, the abdominal pain or discomfort is associated with a change in stool consistency (harder or looser) or stool frequency (increased or decreased), and is often relieved by passage of stool. Other symptoms may include abdominal bloating or distension, straining, sensation of incomplete evacuation, or passage of mucus. Subtypes of IBS exist, based on the predominant symptom: constipation-predominant IBS (C-IBS), diarrhea-predominant IBS (D-IBS), and mixed IBS (M-IBS).

The exact pathophysiology of IBS remains unknown, although IBS is considered as a biopsychosocial disorder with disturbances of motor function, heightened visceral sensitivity, and possibly central nervous system disturbances. No diagnostic tests are presently available to diagnose IBS, and symptom- based diagnostic criteria are used to make the diagnosis. Individuals presenting with typical symptoms of IBS may not require additional laboratory, radiologic, or endoscopic evaluation, but those with severe symptoms may warrant additional testing to rule out other disease. Treatment is usually selected based on the predominant symptom. For example, antispasmodics or visceral neuromodulators may be used for those with significant pain; antidiarrheals may be used in those with diarrhea; and laxatives (fiber, osmotic, stimulant) or other prokinetic agents may be used in those with constipation.

Disease definition

The definition of IBS has evolved over time, from a diagnosis of exclusion to the symptom- based diagnostic criteria including Manning, Rome I, Rome II, and Rome III criteria. The Rome I criteria were originally

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

based on clinical studies from Bristol (which defined the Manning criteria) and Germany and were subsequently modified by consensus [1,2]. The Rome II criteria were subsequently also defined by consensus, based on the Rome I criteria and the results of factor analysis studies, which suggested that the previous criteria could be simplified [3]. The most recent version is the Rome III criteria. By Rome III criteria, IBS is defined as "recurrent abdominal pain or discomfort, at least 3 days per month in the last 3 months associated with 2 or more of the following: (i) improvement with defecation, (ii) onset associated with a change in frequency of stool, (iii) onset association with a change in form (appearance) of stool" [4].

Two points should be made about the disease definition in IBS. First, because various diagnostic criteria for IBS have been employed over the last three decades (Manning criteria [5], Rome 1989 [6], Rome 1990[7], Rome I (1992) criteria [8], and Rome II (1999) criteria [9]), the majority of epidemiologic studies are based on the older criteria rather than the most recent Rome III criteria and disease burden estimates may vary due to use of different criteria. Second, although published diagnostic criteria are increasingly recognized and utilized in clinical practice, there are several studies documenting less than optimal knowledge of the diagnostic criteria among gastroenterologists and general practitioners [10–12], as well as poor agreement between diagnostic criteria and physicians [13], suggesting that many providers make the diagnosis based on clinical impression alone, incorporating IBS with other functional gastrointestinal disorders. Which disease definition was utilized in a specific study may impact conclusions [14], thus, any review of epidemiology literature related to IBS must be cognizant of how IBS was defined.

Prevalence and incidence

Many population-based surveys have estimated the prevalence of IBS using the responses of surveys which record bowel symptoms. The prevalence rates in these studies have varied between 3 and 22 per hundred [15–18]. Although this sevenfold difference may represent true differences in populations, it likely reflects differences in the IBS definition. The earlier Manning criteria are more generous and less restrictive than the recent Rome criteria [19–21]. Higher prevalence rates are identified using a threshold of two of six Manning criteria [20]. Lower prevalence rates are identified using more specific criteria, either by increasing the number of Manning criteria necessary to make the diagnosis or utilizing the Rome criteria.

In a meta-analysis, Lovell and Ford [22] reported the pooled prevalence of IBS in the community was 11.2 % (95 % CI 9.85–12.8 %), which was based on data from 80 separate studies. In addition, they draw the World map of IBS prevalence (Figure 20.1). However, this meta-analysis also found significant heterogeneity between studies attributed to different definition, population, response rates, or data collection methods. The major IBS prevalence studies in Western countries are summarized in Table 20.1. In addition, the prevalence of IBS in Asia by the Rome II criteria is shown in Table 20.2 [23–27].

The incidence of IBS is more difficult to estimate. From one population study in the United States, which was based on two surveys sent to a random sample of the community one year apart, the IBS onset rate was 9 %. However, in another study using physician-based clinical diagnosis of IBS in the same population, the incidence rate of clinically diagnosed IBS was much lower, 196 cases per 100,000 person-years [39]. A study from Europe showed a similar annual incidence rate of IBS, about 200 to 300 per 100,000 people [40].

Symptoms may come and go and change over time [41]. Because studies that utilize physician-originated clinical diagnoses will not include people who do not seek care [38], the incidence of IBS from these clinical-based studies likely is an underestimate of the true incidence. Nonetheless, if only half seek care, the observed incidence can be doubled to 400 per 100,000 per year and then multiplied by a 20-year disease duration to get a prevalence of 12 %, which is in keeping with the data. Another population-based study conducted in England and Wales, using first-diagnosis of IBS by a general practitioner, found an estimated incidence of 2.6 per 1000 person-years [42].

Risk factors for disease

Age and gender

Based on prevalence data, IBS appears to be more common in women than men, with up to a 2:1 ratio. The ratio may increase further when outpatient studies

Figure 20.1 Prevalence of IBS according to country. Source: Lovell & Ford 2012 [22], figure 1. Reproduced with permission of American Gastroenterological Association.

are examined; however, it is unclear if the higher ratio in patients may reflect higher healthcare seeking by women than men with comparable symptoms. More interestingly, the female predominance reported in the West has not been reported in some Asian countries [32–35]. Notably, a higher prevalence of IBS in males has been reported in some Asian countries [32,33].

As no real data exists regarding the age-of-onset of IBS or age-specific incidence, it remains unclear whether increasing age is associated with increased or decreased risk of developing IBS. However, the stable prevalence of IBS across various adult age groups [15,16,43] suggests that advancing age is not a major risk factor for IBS.

Geography

IBS is a common disorder around the world, with studies reporting prevalence rates of 6–22 % in Western countries and 1–15 % in Asian countries [23,38,44]. Although IBS has been studied in other continents such as Africa and South America, population-based studies from these regions are lacking. A recent systematic review evaluating geographical and ethnic differences in IBS did not find real differences between countries in the East and countries in the West with respect to overall prevalence rate, but some Asian countries were not included [44].

With regard to IBS bowel habit subtypes, one systemic review [45] reported that population-based studies from the United States using the Manning criteria found similar distributions among constipation-predominant IBS (IBS-C), diarrhea-predominant IBS (IBS-D), and IBS alternating between diarrhea and constipation (IBS-A), while European studies (Rome I, Rome II, or self-reporting) showed either IBS-C or IBS-A as the most prevalent subtypes. For example, in one study approximately 16 % of the IBS patients had IBS-C, 21 % had IBS-D, and 63 % had IBS-A [46]. Whether the agreement between subtyping of IBS patients based on Rome II versus Rome III criteria is good or poor [47,48] is controversial. Very few data by IBS subgroup based on the recent Rome III classification system are available.

Race and ethnicity

Studies evaluating race within the same country suggest that IBS may affect Caucasians more than other ethnic groups. For example, studies of the 1987 National Health Interview Survey (NHIS), the 1976–1980 Second National Health and Nutrition

Table 20.1 Prevalence of irritable bowel syndrome in Western countries

Author, Country	Year	*n*	Case definition	% IBS		
				Overall	Men	Women
Talley [16], USA	1987	835	Manning 2	15.8	15.8	18.2
			Manning 3	12.8	12.1	13.6
Hahn [109], USA	1989	42,392	Manning 2	3	–	–
			Rome I	12	–	–
Drossman [43], USA	1990	5430	Rome I	9.4	7.7	14.5
Saito [14], USA	1992	643	Manning 3	15.7	13.5	17.7
			Rome I	8.4	8.4	8.4
Mearin [19], Spain	2001	2000	Manning	10.3	–	–
			Rome I	12.1	–	–
			Rome II	3.3	1.9	4.6
Brommelaer [110], France	2002	8221	Manning	2.5	1.7	3.1
			Rome I	2.1	1.4	2.8
			Rome II	1.1	0.9	1.3
Thompson [111], Canada	2002	1149	Rome II	12.1	8.7	15.2
Boyce [112], Australia	1997	2910	Manning	13.6	–	–
			Rome I	4.4	4.4	9.1
Jones [17], England	1992	1620	Manning	21.6	18.7	24.3
Agreus [113], Sweden	1988	1290	Rome I	12.5	–	–
Wilson [114] UK	2003	4807	Rome II	8.1	–	–
Hungin [46], Europe (UK, France, Germany, Italy, Holland, Belgium, Spain, Switzerland)	2003	41984	Overall	9.6	7.1	12
			Manning	6.5	–	–
			Rome I	4.2	–	–
			Rome II	2.9	–	–
Kennedy [83], UK	1998	3179	Manning 3	17.2	10.5	22.9
Icks [17], Germany	2002	1281	Patient report	12.5	–	–
Kay [18], Denmark	1994	4581	Symptom criteria	6.6	5.6	7.7
Heaton [115], UK	1992	1896	Manning 3	9.5	5.0	13.0
			Manning 2	21.6	18.7	24.3
Hillila [21], Finland	2004	3650	Manning 2	16.2	13.1	19.2
			Manning 3	9.7	8.3	11.2
			Rome I	5.5	5.1	6.1
			Rome II	5.1	5.1	5.3
Jung [84], USA	2004	2273	Rome III	11	8	14
Olafsdottir [116], Iceland	1996	1336	Manning	31	–	–
	2006	799	Rome III	10	–	–
			Manning	32	–	–
			Rome II	5.0	–	–
			Rome III	13	–	–

Table 20.2 Prevalence of irritable bowel syndrome in Asian countries

Author, Country	Year	*n*	Case definition	% IBS		
				Overall	Men	Women
Gwee [24], Singapore	2000	2276	Manning	11	9.5	12.6
			Rome I	10.4	9.0	11.7
			Rome II	8.6	7.8	9.4
Xiong [25], South China	2002	4178	Manning	11.5	9.7	13.0
			Rome II	5.7	5.0	6.3
Lau [26], Hong Kong	1996	1298	Rome II	3.7	3.6	3.8
Ho [27], Singapore	1990	696	Manning	2.3	–	–
Kwan [28], Hong Kong	2000	1797	Rome II	6.6	–	–
Danivat [29], Thailand	1988	1077	Manning	4.4	–	–
Masud [30], Bangladesh	2000	2426	Rome I	8.5	5.8	10.7
Rajendra [31], Malaysia	2000	949	Rome II	14	–	–
Ghoshal [32], India	2005	7285	Clinical	4.2	4.3	4.0
Han [33], Korea	2004	1066	Rome II	6.6	7.1	6.0
Husain [34], Pakistan	2006	880	Rome II	13.3	13.1	13.4
Lu [35], Taiwan	2001	2865	Rome II	22.1	21.8	22.8
Miwa [36], Japan	2006	10000	Rome III	13.1	10.7	15.5
Sorouri [37], Iran	2006	18180	Rome III	1.1	0.6	1.5

Examination Survey (NHANES II), and the 1985 National Ambulatory Care Medical Care Survey (NAMCS) show that the rate of self-reported spastic colon or mucous colitis was consistently greater in Caucasians than Blacks or African Americans [15]. Data from other ethnic groups were not reported. However, these figures were based on "being told" of these diagnoses, and thus, these figures may reflect lower access to health care rather than true differences in prevalence between race and ethnic groups. Another study comparing prevalence of IBS among US African Americans and Caucasians also found Caucasians were over twofold more likely to report IBS, after adjusting for age, education, and household income [49]. Of note, the study sample was a convenience sample (rather than population-based) raising the question of participation and selection bias affecting the final estimates, but recognizing the paucity of data regarding race, suggests that even after taking into account education level and socioeconomic status, Caucasians may be at higher risk than African Americans for IBS. Another non-population-based study comparing Hispanics to non-Hispanics showed that IBS-type symptoms were less common in Hispanics compared to non-Hispanic Whites, although a significant ethnic difference was not found after controlling for covariates [50]. The authors also reported that Hispanics were less likely to see a physician for their bowel symptoms [51]. In summary, studies suggest IBS is more common among Caucasians than non-Caucasians, but further study is warranted.

Socioeconomic status

Few studies reported the association between socioeconomic status and IBS [19,34,52,53]. The recent meta-analysis study [22], which was conducted as a pooled analysis of four studies [19,34,52,53] reporting the prevalence of IBS according to socioeconomic status, found no significant association of socioeconomic status with IBS status.,

Diet

Although many patients report dietary triggers for their symptoms, various food substances have been reported to be associated with exacerbating IBS symptoms, and some dietary elimination studies show positive symptom benefit, the role of diet and the specific dietary components in causing IBS is perplexing

as there is considerable heterogeneity in response to foods [54]. The only population-based study comparing diet among cases and controls demonstrates little difference in the dietary and nutrient composition among those reporting IBS-like symptoms and those not reporting symptoms, suggesting that food sensitivity rather than dietary excess is associated with IBS [55]. To date, food allergy has not clearly been shown to be a cause of IBS [56]. In summary, there may be select food substances that worsen or trigger symptoms, but in themselves do not cause IBS.

Psychological factors

Psychological and psychiatric comorbidity has been frequently linked with IBS, and several treatments for IBS are used either to directly treat the psychological disorders or as nerve-modulating agents [18,57]. Some have suggested that the high level of comorbidity observed in IBS patients may be a reflection of factors that drive healthcare seeking. However, there is some data arguing that consulters with IBS are not different psychologically from nonconsulters with IBS [58,59], and that neuroticism, psychological morbidity, and abuse history are not predictors of healthcare seeking [60]. Moreover, Choung. et al. [61] showed that even in community-level individuals with IBS (and not clinic-based individuals), somatization is significantly associated with IBS and no IBS community subject was free of psychological distress and somatic symptoms. Thus, psychological factors appear to be closely linked with IBS.

Abuse

Many clinic-based studies have reported a higher prevalence of abuse history in IBS patients compared with controls [62–64], although it should be noted that there are also several clinic-based studies that have not found an association between abuse and IBS [65,66]. One community-based study in Olmsted County affirmed the association between sexual abuse, emotional or verbal abuse, and abuse in childhood and adulthood with IBS [67]. However, a similar study in Penrith, Australia conducted by the same investigator, although finding an association between childhood abuse and IBS, observed that the association disappeared after controlling for age, gender, and psychological factors [68]. This study suggested that abuse may lead to higher neuroticism, and consequently, higher healthcare seeking. Other studies have shown that patients with past abuse demonstrate higher levels of current psychological distress [65], and that the abuse history, although not linked with IBS specifically, may result in an increased number of gastrointestinal and extra-gastrointestinal symptoms, irrespective of the presence of an underlying functional or organic disorder [66]. In summary, although abuse has been linked to IBS, abuse may not lie in the causal pathway to IBS, but this association remains an area of relative controversy.

Infection

Several patient-based or outbreak studies have shown that a subset of individuals with acute gastroenteritis go on to develop persistent IBS [69,70,71–74]. One population-based study utilized a database of clinical diagnoses in the United Kingdom and observed that the cohort with bacteriologically confirmed gastroenteritis were 12-fold more likely to develop IBS within the next year [75]. Another population-based study of patients presenting with bacterial gastroenteritis at a primary care practice in the UK observed that after excluding those with IBS at baseline [76], IBS was 10-fold more common in cases than controls [77]. Another study conducted in Walkerton, Canada following a large outbreak of acute *E. coli* 0157:H7 and *Campylobacter jejuni* gastroenteritis yielded a threefold risk for the development of post-infectious IBS after clinically suspected gastroenteritis [78]. Thus, post-infectious IBS appears to be a real clinical entity. However, it is unlikely that infection is the underlying etiology for all IBS cases, and may represent the major risk factor in only a small subset of patients. Furthermore, psychological characteristics appear to be independent risk factors for the development of post-infectious IBS [69,70], and the role and interaction of inflammatory mediators with IBS remains to be determined.

Family history

Various clinical studies confirm that IBS appears to aggregate in families [79–81]. In the small population-based study [14], it was found that reporting a first-degree relative with abdominal pain or bowel problems was associated with self-report of IBS, with an

estimated odds ratio of 2.3 (95 % CI 1.3–3.9). In contrast, reporting a spouse with pain or bowel problems was not associated with IBS. More recently, a large family study [82], which collected bowel habits directly from IBS cases, controls, and their first-degree relatives to construct pedigrees accurately, showed relatives of a family member with IBS are at two- to threefold higher risk for IBS than control patient relatives. These studies do suggest that a positive family history of IBS remains a relevant risk factor for a diagnosis of IBS; however, whether this is due to genetics or shared environment (including learned illness behavior) remains to be determined.

Overlap with other disorders

Several population and clinical studies [43,83–88] have reported associations with other disorders, specifically other functional gastrointestinal disorders. For example, a community study [83] in the United Kingdom showed that IBS, gastroesophageal reflux disease (GERD) and symptomatic bronchial hyperresponsiveness occurred more frequently together than expected. In subjects with IBS, 47 % also had GERD. Locke et al. [85] showed in a community-based study that 4–9 % of the population had two GI symptom complexes, and 1–4 % of the population had three GI symptom complexes. The mechanism behind these overlapping syndromes is not yet clear.

A subset of IBS patients also experience nongastrointestinal symptoms. IBS patients have two to three times as many nongastrointestinal healthcare visits as control subjects without IBS [43,81]. Nongastrointestinal nonpsychiatric disorders documented to be associated with IBS in a detailed literature review included chronic fatigue syndrome (51 %), chronic pelvic pain (50 %), and temporomandibular joint disorders (64 %) [89]. In referred patients with IBS, psychiatric disorders have also been reported to be very common, leading some to argue that IBS is a part of the psychiatric disease spectrum and not a unique condition [90,91]. Whitehead et al. [89] performed a study comparing the comorbidities between 3153 patients with IBS and age- and gender-matched controls in a health maintenance organization. They argued that the elevated incidence of nongastrointestinal disorders might occur in a subset because patients with IBS are hyper-vigilant and consult much more readily for problems than those without IBS. There has been a recent movement to overhaul the classification of somatoform disorder which may incorporate IBS [92].

Natural history

Reviews of studies evaluating the natural history of IBS demonstrate that it is indeed a chronic disorder in clinic-based patients [93]. With long-term follow-up, 20–50 % of patients have unchanged symptoms, 2–18 % of patients have worsening symptoms, and in the balance, symptoms improve. For example, in a large 1-year prospective, observational study of 400 primary care and gastroenterology clinic patients in Spain, half of the patients and half of the physicians considered their symptoms to have improved, although objective review of diary data showed that the improvement was small and that the major predictor of improvement was severe baseline symptoms [94]. However, population-based studies that include patients as well as nonconsulters, show considerable fluctuation of IBS and non-IBS symptoms. For example, a random sample survey in Sweden in 1988, 1989, and then 1995 showed that among those with IBS at baseline, 55 % continued to report IBS at both follow-up surveys [95]; 3 % were symptom-free at year 1, and 13 % were symptom-free at year 7, thus implying that among a small subset, there is perhaps complete resolution of symptoms. Fifteen percent and 8 % had changed from IBS symptoms to dyspepsia symptoms at years 1 and 7, respectively, suggesting that other GI symptoms may develop or predominate in the natural history of IBS. Further, a recent long-term follow-up study from Olmsted County evaluated the transitions amongst FGIDs over 12 years [41]. Halder et al. [41] showed the substantial transition among the categories, with about one third of subjects with IBS developing another functional gastrointestinal disorder.

The diagnosis of IBS appears also to be durable, with only an estimated 2–5 % of IBS patients being given an initial misdiagnosis that is subsequently changed [93].

Disability and quality of life

A number of studies have been conducted to quantitate the disability that results from IBS. A recent

systematic review of the available literature found that the average number of days off work per year because of IBS was between 8.5 and 21.6 [96]. Patients also report being late for work or leaving work early, and having to make other work–life adjustments including working shorter hours, refraining from applying for promotions or a new job [97], and/or selecting work based on settings for reasons such as restroom access (including working from home or being self-employed). IBS also impacts the personal and social lives of affected individuals resulting in avoidance or reduction of activities, inhibited personal relationships, interference with sex life, and embarrassment at using public toilets [97–99].

Not surprisingly, health-related quality of life (HRQoL) is lower in patients with IBS compared to the general population. A number of studies have evaluated HRQoL in patients with IBS, many of which were evaluated and summarized in a recent, well-conducted systematic review [100]. This review found:
1 HRQoL is lower in patients with moderate to severe IBS compared with healthy controls;
2 patients with IBS have impaired HRQoL comparable to diseases such as moderate to severe GERD, end-stage renal disease, peptic ulcer disease, inflammatory bowel disease, and liver disease;
3 patients with a response to therapy have a correlative improvement in HRQoL;
4 the subtype of IBS does not affect the degree of impact of IBS on HRQoL;
5 degree of impairment of HRQoL is directly related to severity of bowel symptoms.

Healthcare utilization and costs

In 2002, the American Gastroenterological Association (AGA) published findings of their study to determine the burden of selected gastrointestinal diseases [101]. Using publicly available and proprietary databases to assess inpatient hospital stays, physician office visits, emergency room visits, and hospital outpatient visits, the study found that IBS was second only to GERD as the most prevalent chronic gastrointestinal disorder. In a separate study using the National Ambulatory Medical Care Survey (NAMCS), compared to non-GI disease, IBS-related outpatient physician visits occurred at the same rate as for asthma and 2.6 times the rate of visits for migraine headaches [102]. In a recent database analysis by Everhart et al. [103] on the burden of digestive diseases in the United States, IBS and chronic constipation are the most commonly diagnosed functional intestinal disorders. Thus, visits directly related to IBS care appear to be extremely common in the United States.

Besides visits directly related to IBS, patients with IBS utilize more healthcare resources overall. Studies of managed care administrative databases [104–106], administrative claims data from a national Fortune 100 manufacturer collecting information on medical, pharmaceutical and disability claims for employees, spouses, and retirees [107], and Medicaid administrative databases [108] have demonstrated that overall healthcare utilization was greater in patients with IBS compared to controls without the syndrome.

Estimates for the direct and indirect costs attributed to IBS have been evaluated in many settings. The AGA figures estimated that the direct costs from inpatient and outpatient visits and prescription medications for IBS exceeded $1.6 billion in 1998, or $1.7 billion in year 2000 dollars. The costs arose from 3.65 million physician visits, 500,000 hospital inpatient stays, 150,000 hospital outpatient visits, and 87,000 emergency room visits. Estimated indirect costs, based exclusively on lost work days due to consumption of health care, was estimated at US$205 million, but using different methodology applying wage figures to age, work loss was estimated at US$19.2 billion in 1998, or US$20.2 billion in year 2000 dollars.

Areas for further study

Until the pathophysiology of IBS is better understood, there remain many lines of investigation to pursue further study. Several gaps in our understanding of the epidemiology of IBS remain:
- The accuracy of symptom-based diagnostic criteria, such as the Rome criteria.
- The determination of whether IBS is one disorder, or an etiologically heterogeneous collection of multiple disorders.
- The identification of environmental and genetic risk factors that lead to the clustering of IBS in families, including the role of learned illness behavior in IBS.
- The determination of the long-term natural history of IBS including better description of its onset (e.g. incidence, age-of-onset), its evolution from childhood

through adulthood, and its long-term consequences (mortality, morbidity).

Conclusions

IBS is a common disorder that exists in individuals of all ages and various ethnic and cultural backgrounds. Because not everyone needs to seek care, population-based studies are needed to truly understand the epidemiology of IBS. Moreover, as it is one of the most prevalent gastrointestinal disorders, results in disability, decreased productivity, and absenteeism in working-age individuals, and costs the healthcare system considerable dollars, a better understanding of the pathophysiology is needed. Several environmental risk factors – such as diet and stress – have been well studied, but clearly are not the sole determinants of disease development and exacerbation. Further epidemiologic studies are warranted to identify the environmental, psychosocial, and genetic risk factors for IBS occurrence and prognosis so that better diagnostic tests and treatments may be developed.

Multiple choice questions

1 You read an epidemiologic paper in which the following result is given: "30 per 1000 report ever having received a diagnosis of spastic colon or mucus colitis." What does this result represent?

A Incidence
B Prevalence
C Cumulative incidence
D Attributable risk

2 Which of the following is *NOT* a benefit of performing population-based research?

A Avoids selection bias since not everyone seeks care
B Avoids severity bias, some clinics see sicker patients
C Provides more precise information than available in clinic-based studies
D Population-based research using noninvasive techniques can be cheaper

3 Which risk factor has *NOT* been identified for functional gastrointestinal disorders thus far?

A Cigarette smoking
B Prior enteric infection
C Family history
D Mood disturbance

4 How common is IBS? The prevalence of IBS is . . .

A One in ten
B One in a hundred
C One in a thousand
D One in ten thousand
E One in a hundred thousand

References

1 Manning AP, et al. Towards positive diagnosis of the irritable bowel. *BMJ* 1978; 2(6138):653–4.
2 Thompson WG, et al. Functional bowel disorders and functional abdominal pain. *Gut* 1999;45 (Suppl 2):II43–7.
3 Kruis W, et al. A diagnostic score for the irritable bowel syndrome. Its value in the exclusion of organic disease. *Gastroenterology* 1984;87(1):1–7.
4 Longstreth GF, et al. Functional bowel disorders. *Gastroenterology* 2006;130(5):1480–91.
5 Manning AP, et al. Towards positive diagnosis of the irritable bowel. *BMJ* 1978;2(6138):653–4.
6 Thompson WG, et al. Irritable bowel syndrome: guidelines for the diagnosis. *Gastroenterol Int* 1989;2:92–5.
7 Drossman DA, et al. Identification of sub-groups of functional gastrointestinal disorders. *Gastroenterol Int* 1990;3(4):159–72.
8 Thompson WG, et al. Functional bowel disease and functional abdominal pain. *Gastroenterol Int* 1992;5(2):75–91.
9 Thompson WG, et al. Functional bowel disorders and functional abdominal pain. *Gut* 1999;45(Suppl 2):II43–7.
10 Lea R, et al. Diagnostic criteria for irritable bowel syndrome: utility and applicability in clinical practice. *Digestion* 2004;70(4):210–3.
11 Charapata C, Mertz H. Physician knowledge of Rome symptom criteria for irritable bowel syndrome is poor among non-gastroenterologists. *Neurogastroenterol Motil* 2006;18(3):211–6.
12 Longstreth GF, Burchette RJ. Family practitioners' attitudes and knowledge about irritable bowel syndrome: effect of a trial of physician education. *Fam Pract* 2003;20(6):670–4.
13 Vandvik PO, Aabakken L, Farup PG. Diagnosing irritable bowel syndrome: poor agreement between general practitioners and the Rome II criteria. *Scand J Gastroenterol* 2004;39(5):448–53.
14 Saito YA, et al. A comparison of the Rome and Manning criteria for case identification in epidemiological

investigations of irritable bowel syndrome. *Am J Gastroenterol* 2000;95(10):2816–24.
15 Sandler RS. Epidemiology of irritable bowel syndrome in the United States. *Gastroenterology* 1990;99(2):409–15.
16 Talley NJ, et al. Epidemiology of colonic symptoms and the irritable bowel syndrome. *Gastroenterology* 1991;101(4):927–34.
17 Jones R, Lydeard S. Irritable bowel syndrome in the general population. *BMJ* 1992;304(6819):87–90.
18 Kay L, Jorgensen T, Jensen KH. The epidemiology of irritable bowel syndrome in a random population: prevalence, incidence, natural history and risk factors. *J Intern Med* 1994;236(1):23–30.
19 Mearin F, et al. Irritable bowel syndrome prevalence varies enormously depending on the employed diagnostic criteria: comparison of Rome II versus previous criteria in a general population. *Scand J Gastroenterol* 2001;36(11):1155–61.
20 Saito YA, et al. The effect of new diagnostic criteria for irritable bowel syndrome on community prevalence estimates. *Neurogastroenterol Motil* 2003;15(6):687–94.
21 Hillila MT, Farkkila MA. Prevalence of irritable bowel syndrome according to different diagnostic criteria in a non-selected adult population. *Aliment Pharmacol Ther* 2004;20(3):339–45.
22 Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol* 2012; doi: 10.1016/j.cgh.2012.02.029. Epub 2012 Mar 15.
23 Choung RS, Locke GR, 3rd. Epidemiology of IBS. *Gastroenterol Clin North Am* 2011;40(1):1–10.
24 Gwee KA, et al. The prevalence, symptom characteristics, and impact of irritable bowel syndrome in an Asian urban community. *Am J Gastroenterol* 2004;99(5):924–31.
25 Xiong LS, et al. A population-based epidemiologic study of irritable bowel syndrome in South China: stratified randomized study by cluster sampling. *Aliment Pharmacol Ther* 2004;19(11):1217–24.
26 Lau EM, et al. Epidemiology of irritable bowel syndrome in Chinese. *Dig Dis Sci* 2002;47(11):2621–4.
27 Ho KY, Kang JY, Seow A. Prevalence of gastrointestinal symptoms in a multiracial Asian population, with particular reference to reflux-type symptoms. *Am J Gastroenterol* 1998;93(10):1816–22.
28 Kwan AC, et al. Prevalence of irritable bowel syndrome in Hong Kong. *J Gastroenterol Hepatol* 2002;17(11):1180–6.
29 Danivat D, Tankeyoon M, Sriratanaban A. Prevalence of irritable bowel syndrome in a non-Western population. *Br Med J (Clin Res Ed)* 1988;296(6638):1710.
30 Masud MA, Hasan M, Khan AK. Irritable bowel syndrome in a rural community in Bangladesh: prevalence, symptoms pattern, and health care seeking behavior. *Am J Gastroenterol* 2001;96(5):1547–52.
31 Rajendra S, Alahuddin S. Prevalence of irritable bowel syndrome in a multi-ethnic Asian population. *Aliment Pharmacol Ther* 2004;19(6):704–6.
32 Ghoshal UC, et al. Epidemiological and clinical profile of irritable bowel syndrome in India: report of the Indian Society of Gastroenterology Task Force. *Indian J Gastroenterol* 2008;27(1):22–8.
33 Han SH, et al. Prevalence of irritable bowel syndrome in Korea: population-based survey using the Rome II criteria. *J Gastroenterol Hepatol* 2006;21(11):1687–92.
34 Husain N, et al. A population-based study of irritable bowel syndrome in a non-Western population. *Neurogastroenterol Motil* 2008;20(9):1022–9.
35 Lu CL, et al. Current patterns of irritable bowel syndrome in Taiwan: the Rome II questionnaire on a Chinese population. *Aliment Pharmacol Ther* 2003;18(11–12):1159–69.
36 Miwa H. Prevalence of irritable bowel syndrome in Japan: Internet survey using Rome III criteria. *Patient Prefer Adherence* 2008;2:143–7.
37 Sorouri M, et al. Functional bowel disorders in Iranian population using Rome III criteria. *Saudi J Gastroenterol* 2010;16(3):154–60.
38 Cremonini F, Talley NJ. Irritable bowel syndrome: epidemiology, natural history, health care seeking and emerging risk factors. *Gastroenterol Clin North Am* 2005;34(2):189–204.
39 Locke GR, 3rd, et al. Incidence of a clinical diagnosis of the irritable bowel syndrome in a United States population. *Aliment Pharmacol Ther* 2004;19(9):1025–31.
40 Ruigomez A, et al. One-year follow-up of newly diagnosed irritable bowel syndrome patients. *Aliment Pharmacol Ther* 1999;13(8):1097–102.
41 Halder SL, et al. Natural history of functional gastrointestinal disorders: a 12-year longitudinal population-based study. *Gastroenterology* 2007;133(3):799–807.
42 García Rodríguez LA, et al. Detection of colorectal tumor and inflammatory bowel disease during follow-up of patients with initial diagnosis of irritable bowel syndrome. *Scand J Gastroenterol* 2000;35(3):306–11.
43 Drossman DA, et al. U.S. householder survey of functional gastrointestinal disorders. Prevalence, sociodemography, and health impact. *Dig Dis Sci* 1993;38(9):1569–80.
44 Kang JY. Systematic review: the influence of geography and ethnicity in irritable bowel syndrome. *Aliment Pharmacol Ther* 2005;21(6):663–76.
45 Guilera M, Balboa A, Mearin F. Bowel habit subtypes and temporal patterns in irritable bowel syndrome:

systematic review. *Am J Gastroenterol* 2005; 100(5):1174–84.
46 Hungin AP, et al. The prevalence, patterns and impact of irritable bowel syndrome: an international survey of 40,000 subjects. *Aliment Pharmacol Ther* 2003;17(5): 643–50.
47 Dorn SD, et al. Irritable bowel syndrome subtypes defined by Rome II and Rome III criteria are similar. *J Clin Gastroenterol* 2009;43(3):214–20.
48 Ersryd A, et al. Subtyping the irritable bowel syndrome by predominant bowel habit: Rome II versus Rome III. *Aliment Pharmacol Ther* 2007;26(6):953–61.
49 Wigington WC, Johnson WD, Minocha A. Epidemiology of irritable bowel syndrome among African Americans as compared with whites: a population-based study. *Clin Gastroenterol Hepatol* 2005;3(7): 647–53.
50 Zuckerman MJ, et al. Comparison of bowel patterns in Hispanics and non-Hispanic whites. *Dig Dis Sci* 1995;40(8):1763–9.
51 Zuckerman MJ, et al. Health-care-seeking behaviors related to bowel complaints. Hispanics versus non-Hispanic whites. *Dig Dis Sci* 1996;41(1):77–82.
52 Lee S, et al. Irritable bowel syndrome is strongly associated with generalized anxiety disorder: a community study. *Aliment Pharmacol Ther* 2009;30(6):643–51.
53 Li FX, et al. Irritable bowel syndrome and health-related quality of life: a population-based study in Calgary, Alberta. *Can J Gastroenterol* 2003;17(4):259–63.
54 Lea R, Whorwell PJ. The role of food intolerance in irritable bowel syndrome. *Gastroenterol Clin North Am* 2005;34(2):247–55.
55 Saito YA, et al. Diet and functional gastrointestinal disorders: a population-based case-control study. *Am J Gastroenterol* 2003;98(9):S275.
56 Bischoff S, Crowe SE. Gastrointestinal food allergy: new insights into pathophysiology and clinical perspectives. *Gastroenterology* 2005;128(4):1089–113.
57 Palsson OS, Drossman DA. Psychiatric and psychological dysfunction in irritable bowel syndrome and the role of psychological treatments. *Gastroenterol Clin North Am* 2005;34(2):281–303.
58 Weinryb RM, et al. Psychological factors in irritable bowel syndrome: a population-based study of patients, non-patients and controls. *Scand J Gastroenterol* 2003;38(5):503–10.
59 Kanazawa M, et al. Patients and nonconsulters with irritable bowel syndrome reporting a parental history of bowel problems have more impaired psychological distress. *Dig Dis Sci* 2004;49(6):1046–53.
60 Talley NJ, Boyce PM, Jones M. Predictors of health care seeking for irritable bowel syndrome: a population based study. *Gut* 1997;41(3):394–8.
61 Choung RS, et al. Psychosocial distress and somatic symptoms in community subjects with irritable bowel syndrome: a psychological component is the rule. *Am J Gastroenterol* 2009;104(7):1772–9.
62 Walker EA, et al. Histories of sexual victimization in patients with irritable bowel syndrome or inflammatory bowel disease. *Am J Psychiatry* 1993;150(10): 1502–6.
63 Delvaux M, Denis P, Allemand H. Sexual abuse is more frequently reported by IBS patients than by patients with organic digestive diseases or controls. Results of a multicentre inquiry. French Club of Digestive Motility. *Eur J Gastroenterol Hepatol* 1997;9(4):345–52.
64 Talley NJ, Fett SL, Zinsmeister AR. Self-reported abuse and gastrointestinal disease in outpatients: association with irritable bowel-type symptoms. *Am J Gastroenterol* 1995;90(3):366–71.
65 Hobbis IC, Turpin G, Read NW. A re-examination of the relationship between abuse experience and functional bowel disorders. *Scand J Gastroenterol* 2002;37(4):423–30.
66 Baccini F, et al. Prevalence of sexual and physical abuse and its relationship with symptom manifestations in patients with chronic organic and functional gastrointestinal disorders. *Dig Liver Dis* 2003;35(4):256–61.
67 Talley NJ, et al. Gastrointestinal tract symptoms and self-reported abuse: a population-based study. *Gastroenterology* 1994;107(4):1040–9.
68 Talley NJ, Boyce PM, Jones M. Is the association between irritable bowel syndrome and abuse explained by neuroticism? A population based study. *Gut* 1998;42(1):47–53.
69 Gwee KA, et al. Psychometric scores and persistence of irritable bowel after infectious diarrhoea. *Lancet* 1996;347(8995):150–3.
70 Gwee KA, et al. The role of psychological and biological factors in postinfective gut dysfunction. *Gut* 1999;44(3):400–6.
71 Neal KR, Hebden J, Spiller R. Prevalence of gastrointestinal symptoms six months after bacterial gastroenteritis and risk factors for development of the irritable bowel syndrome: postal survey of patients. *BMJ* 1997;314(7083):779–82.
72 Urfer E, et al. Outbreak of Salmonella braenderup gastroenteritis due to contaminated meat pies: clinical and molecular epidemiology. *Clin Microbiol Infect* 2000;6(10):536–42.
73 Thornley JP, et al. Relationship of Campylobacter toxigenicity in vitro to the development of postinfectious irritable bowel syndrome. *J Infect Dis* 2001;184(5):606–9.
74 Cumberland P, et al. The infectious intestinal disease study of England: a prospective evaluation of symptoms

and health care use after an acute episode. *Epidemiol Infect* 2003;130(3):453–60.
75 Rodríguez LA, Ruigómez A. Increased risk of irritable bowel syndrome after bacterial gastroenteritis: cohort study. *BMJ* 1999;318(7183):565–6.
76 Parry SD, et al. Is irritable bowel syndrome more common in patients presenting with bacterial gastroenteritis? A community-based, case-control study. *Am J Gastroenterol* 2003;98(2):327–31.
77 Parry SD, et al. Does bacterial gastroenteritis predispose people to functional gastrointestinal disorders? A prospective, community-based, case-control study. *Am J Gastroenterol* 2003;98(9):1970–5.
78 Marshall JK, et al. Incidence and epidemiology of irritable bowel syndrome after a large waterborne outbreak of bacterial dysentery. *Gastroenterology* 2006;131(2):445–50; quiz 660.
79 Whorwell PJ, et al. Non-colonic features of irritable bowel syndrome. *Gut* 1986;27(1):37–40.
80 Bellentani S, et al. A simple score for the identification of patients at high risk of organic diseases of the colon in the family doctor consulting room. The Local IBS Study Group. *Fam Pract* 1990;7(4):307–12.
81 Levy RL, et al. Intergenerational transmission of gastrointestinal illness behavior. *Am J Gastroenterol* 2000;95(2):451–6.
82 Saito YA, et al. Familial aggregation of irritable bowel syndrome: a family case-control study. *Am J Gastroenterol* 2010;105(4):833–41.
83 Kennedy TM, et al. Irritable bowel syndrome, gastro-oesophageal reflux, and bronchial hyper-responsiveness in the general population. *Gut* 1998;43(6):770–4.
84 Jung HK, et al. Overlap of gastro-oesophageal reflux disease and irritable bowel syndrome: prevalence and risk factors in the general population. *Aliment Pharmacol Ther* 2007;26(3):453–61.
85 Locke GR, 3rd, et al. Overlap of gastrointestinal symptom complexes in a US community. *Neurogastroenterol Motil* 2005;17(1):29–34.
86 Koloski NA, Talley NJ, Boyce PM. Epidemiology and health care seeking in the functional GI disorders: a population-based study. *Am J Gastroenterol* 2002;97(9):2290–9.
87 Corsetti M, et al. Impact of coexisting irritable bowel syndrome on symptoms and pathophysiological mechanisms in functional dyspepsia. *Am J Gastroenterol* 2004;99(6):1152–9.
88 Talley NJ, et al. Gastrointestinal symptoms and subjects cluster into distinct upper and lower groupings in the community: a four nations study. *Am J Gastroenterol* 2000;95(6):1439–47.
89 Whitehead WE, et al. Comorbidity in irritable bowel syndrome. *Am J Gastroenterol* 2007;102(12):2767–76.
90 Whitehead WE, Palsson O, Jones KR. Systematic review of the comorbidity of irritable bowel syndrome with other disorders: what are the causes and implications? *Gastroenterology* 2002;122(4):1140–56.
91 Henningsen P, Zipfel S, Herzog W. Management of functional somatic syndromes. *Lancet* 2007;369(9565):946–55.
92 Kroenke K. Physical symptom disorder: a simpler diagnostic category for somatization-spectrum conditions. *J Psychosom Res* 2006;60(4):335–9.
93 El-Serag HB, Pilgrim P, Schoenfeld P. Systemic review: Natural history of irritable bowel syndrome. *Aliment Pharmacol Ther* 2004;19(8):861–70.
94 Mearin F, et al. Predictive factors of irritable bowel syndrome improvement: 1-year prospective evaluation in 400 patients. *Aliment Pharmacol Ther* 2006;23(6):815–26.
95 Agreus L, et al. Natural history of gastroesophageal reflux disease and functional abdominal disorders: a population-based study. *Am J Gastroenterol* 2001;96(10):2905–14.
96 Maxion-Bergemann S, et al. Costs of irritable bowel syndrome in the UK and US. *Pharmacoeconomics* 2006;24(1):21–37.
97 Silk DB. Impact of irritable bowel syndrome on personal relationships and working practices. *Eur J Gastroenterol Hepatol* 2001;13(11):1327–32.
98 Dapoigny M, et al. Irritable bowel syndrome in France: a common, debilitating and costly disorder. *Eur J Gastroenterol Hepatol* 2004;16(10):995–1001.
99 Hungin AP, et al. Irritable bowel syndrome in the United States: prevalence, symptom patterns and impact. *Aliment Pharmacol Ther* 2005;21(11):1365–75.
100 El-Serag HB, Olden K, Bjorkman D. Health-related quality of life among persons with irritable bowel syndrome: a systematic review. *Aliment Pharmacol Ther* 2002;16(6):1171–85.
101 Sandler RS, et al. The burden of selected digestive diseases in the United States. *Gastroenterology* 2002;122(5):1500–11.
102 Kozma CM, et al. A comparison of office-based physician visits for irritable bowel syndrome and for migraine and asthma. *Manag Care Interface* 2002;15(9):40–3, 49.
103 Everhart JE, Ruhl CE. Burden of digestive diseases in the United States. Part II: lower gastrointestinal diseases. *Gastroenterology* 2009;136(3):741–54.
104 Levy RL, et al. Costs of care for irritable bowel syndrome patients in a health maintenance organization. *Am J Gastroenterol* 2001;96(11):3122–9.
105 Patel RP, et al. The economic impact of irritable bowel syndrome in a managed care setting. *J Clin Gastroenterol* 2002;35(1):14–20.

106 Longstreth GF, et al. Irritable bowel syndrome, health care use, and costs: a U.S. managed care perspective. *Am J Gastroenterol* 2003;98(3):600–7.
107 Leong SA, et al. The economic consequences of irritable bowel syndrome: a US employer perspective. *Arch Intern Med* 2003;163(8):929–35.
108 Martin BC, et al. Utilization patterns and net direct medical cost to Medicaid of irritable bowel syndrome. *Curr Med Res Opin* 2003;19(8):771–80.
109 Hahn BA, Saunders WB, Maier WC. Differences between individuals with self-reported irritable bowel syndrome (IBS) and IBS-like symptoms. *Dig Dis Sci* 1997;42(12):2585–90.
110 Bommelaer G, et al. Prevalence of irritable bowel syndrome (IBS) and variability of diagnostic criteria. *Gastroenterol Clin Biol* 2004;28(6–7 Pt 1):554–61.
111 Thompson WG, et al. Functional gastrointestinal disorders in Canada: first population-based survey using Rome II criteria with suggestions for improving the questionnaire. *Dig Dis Sci* 2002;47(1):225–35.
112 Boyce PM, Koloski NA, Talley NJ. Irritable bowel syndrome according to varying diagnostic criteria: are the new Rome II criteria unnecessarily restrictive for research and practice? *Am J Gastroenterol* 2000; 95(11):3176–83.
113 Agreus L, et al. Irritable bowel syndrome and dyspepsia in the general population: overlap and lack of stability over time. *Gastroenterology* 1995;109(3): 671–80.
114 Wilson S, et al. Prevalence of irritable bowel syndrome: a community survey. *Br J Gen Pract* 2004;54(504):495–502.
115 Heaton KW, et al. Symptoms of irritable bowel syndrome in a British urban community: consulters and nonconsulters. *Gastroenterology* 1992;102(6): 1962–7.
116 Olafsdottir LB, et al. Stability of the irritable bowel syndrome and subgroups as measured by three diagnostic criteria – a 10-year follow-up study. *Aliment Pharmacol Ther* 2010;32(5):670–80.

21 Epidemiology of constipation

Brain E. Lacy & John M. Levenick
Section of Gastroenterology and Hepatology, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA

Key points

- Constipation is a highly prevalent disorder affecting ~15 % of the US population.
- The incidence of chronic constipation is 17.4 % in adults, 6.8/1000 person-years in children less than 5 years, and 3.9/1000 person-years from age 5–21.
- Risk factors include diet, gender, age, socioeconomic status, and medications.
- Primary causes include colonic inertia, normal transit constipation, pelvic floor dysfunction, and IBS-C. Secondary causes include medications, neurologic or metabolic disorders, and obstruction.

Introduction

Chronic constipation is a highly prevalent, heterogeneous disorder that significantly affects patients' lives. Estimates on the prevalence of constipation vary based on how the disorder is defined; a recent review estimated the overall prevalence of constipation in the United States to be approximately 15 % [1]. Recent studies have demonstrated that chronic constipation reduces a patient's quality of life and imposes a significant economic burden on the healthcare system [2,3]. In this chapter we will review the epidemiology of constipation in adults, with an emphasis on the impact of this disorder on patients and society, risk factors, incidence, prevalence, and natural history. Information on the pediatric population will be reviewed, where available.

Constipation defined

The definition of constipation has evolved over the last decade. Once simply defined by the limiting criteria of infrequent bowel movements, the Rome III criteria (see Table 21.1) now defines constipation using a number of different symptoms, including stool frequency, straining, feelings of incomplete evacuation, the need for digital manipulation, and rectal pressure or pain [4]. The use of the Rome criteria for research protocols and in clinical practice represents an important step forward in the field of chronic constipation. Having a precise definition of constipation is critical in order to accurately diagnose the problem, appropriately assess the natural history, and follow the response to treatment. In addition, the use of a broader definition of constipation should improve communication between patients and physicians, since patients and physicians differ dramatically in their definition of constipation. Physicians tend to use objective measures to define constipation and this invariably involves measuring stool frequency [5]. Several large population studies have shown that

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

most people have anywhere from 3 bowel movements per day to 3 bowel movements per week [6,7]. As such, many physicians define constipation as less than 3 bowel movements per week. Patients, however, frequently define constipation using terms other than stool frequency. For example, one study showed that a patient's definition of constipation agreed with a physician's definition only 50 % of the time and most often focused on symptoms rather than stool frequency [8]. In fact, patients use stool frequency as a measure of constipation only 32 % of the time [9]. Patients are likely to report that they are constipated if they have straining at stool (52 %), have hard stools (44 %), have the urge to pass stool but cannot (34 %), or have abdominal discomfort (20 %). Analysis of the National Health Interview Survey data from 1999 found that, in 10,875 subjects older than age 60, straining and hard bowel movements were most strongly associated with self-reported constipation [10]. Another study found that in adults (age >18) with self-reported constipation, the most bothersome symptoms were straining and passing hard stools [11]. The broader definition provided by the Rome III criteria should enable physicians to communicate more effectively with patients and identify and treat a larger number of symptomatic patients.

Table 21.1 Rome III definition of chronic constipation.

- Symptom onset at least 6 months prior to diagnosis
- Presence of symptoms for the last 3 months (see below)
- Insufficient criteria for IBS
- Loose stools are rarely present without the use of laxatives
- Symptoms include two or more of the following during at least 25 % of defecations:
  - Straining
  - Lumpy or hard stools
  - Sensation of incomplete evacuation
  - Sensation of anorectal obstruction or blockade
  - Manual maneuvers to facilitate evacuation
  - Less than 3 bowel movements per week

Source: Modified from Longstreth 2006 [4]

Incidence

The incidence of constipation has not been well studied. Chitkara and colleagues found the incidence of constipation in children <5 years old to be 6.8/1000 person-years, with a decreased incidence of 3.9/1000 person-years from age 5 to less than 21 in Rochester County, Minnesota [12]. The incidence of constipation in adults over a 12-year period in the same county was 17.4 % [13]. After three months of residency in a nursing home, after removing all with a preceding diagnosis of constipation, the incidence of constipation was 7 % [14]. Over an 8-year period in the US military, postinfectious functional constipation was found at a rate of 127/100,000 patient-years; seven times higher in females compared to males (480 vs. 67/100,000 patient-years [15] (Table 21.2).

Prevalence

An accurate measurement of constipation prevalence rates is problematic because self-reporting of constipation symptoms is very subjective, influenced by societal customs, does not correlate well with stool frequency, and does not allow accurate discrimination between constipation subtypes [16–20]. In addition, researchers have used a variety of definitions and questionnaires to assess prevalence rates. Despite these limitations, a number of studies have attempted to measure prevalence rates in both children and adults. The prevalence of constipation in children ranges from 0.7 % to 30 %. The largest study reported to date ($n = 9660$) involved Italian primary care pediatricians who evaluated patients over a 3-month period and found a prevalence rate of 0.7 % [20]. Constipation was defined using Rome II criteria as defecation frequency <3 per week. In contrast to this low prevalence, most other studies have shown higher prevalence rates of constipation (see Table 21.3). A prospective study of 1932 children (ages 2–14 yrs) in Greece identified a prevalence rate of 6 % using a definition of constipation of <3 bowel movements per week or hard stools with painful defecation [21]. A prospective study of 8341 children in southeast Sweden followed over 2.5 years found a prevalence rate of constipation of 6.5 % [22]. A stepwise backward regression model identified low maternal education, female gender, lack of older siblings, and living in a community with more than 3000 people as risk factors for constipation. A study of Turkish children (ages 5–9), which used the North American Society of Gastroenterology and Nutrition (NASPGHAN) definition of constipation (a delay or difficulty in defecation present for 2 weeks or more),

Table 21.2 Constipation: calculated incidence

Population	Incidence	Sample size	Country	Reference
Admission to nursing home		21,012	USA	Robsun 2000
	7 % incidence after 3 months			
General population over a 12-year period				
Children		5299	USA	Chitkara 2007
<5	6.8/1000 person-years			
5–20	3.1/1000 person-years			
Military personnel, post-infectious			USA	Porter 2011
Total	127/100,000 person-years			
Female	480/100,000 person-years			
Male	67/100,000 person-years			

reported a prevalence rate of 12.4 % [23]. A prevalence rate of 17.3 % was reported from a study involving 378 young children (17–19 months) when constipation was identified by parental self-report and the child underwent treatment by a pediatrician [24]. A prospective study of 150 Italian pediatricians monitoring 2879 infants (birth to 6 months; 49 % female) found that constipation was present in 17.6 % [25]. Constipation was defined a priori as one bowel movement every 3 days or more. Breast-fed infants were less likely to have symptoms of constipation than bottle-fed infants ($P = 0.007$). A study of 5282 school children in Japan (range 7–12; 51 % boys) found a prevalence rate of 18.5 % when constipation was defined as <3 bowel movements per week [26]. Some of these studies [26,27] demonstrated an increased prevalence of constipation in girls, while others did not show any gender-related differences [28].

In adults, prevalence rates range from 1.9 % to 40.1 % [17,29–52; and see Table 21.4). The mean prevalence rate is approximately 14 %. The largest study published reported on questionnaires distributed by the American Cancer Society to 890,394 US adults during 1959 and 1960. The self-reported prevalence rate was 27.1 % [33]. A telephone survey study of 10,018 US adults aged 18 and older using self-reported symptoms of constipation over a 3-month period reported a prevalence rate of 14.7 % (16 % of women and 12 % of men; [31]). A questionnaire study of 835 adults in Olmsted County, MN (ages 30–64) identified an identical prevalence rate of 14.7 % (95 % CI 11.9–17.4; [35]). A telephone survey study of 1149 Canadian adults (age 18 or older) found that 27.2 % reported symptoms of constipation over the prior 3 months, while 16.7 % and 14.9 % met Rome I and Rome II criteria for constipation, respectively [43]. A large postal survey study of Australian women (41,724) identified self-reported prevalence rates of constipation of 14.1 % in young women (ages 18–23), 26.6 % in middle-aged women (ages 45–50), and

Table 21.3 Constipation: prevalence rates in children

Author	Study date	Location	Sample size	Age range (yrs)	Prevalence
Miele	1999	Italy	9660	0–12	0.7 %
Roma	1999	Greece	1932	2–14	6.0 %
Ludvigsson	1997	Sweden	8341	1–2.5	6.5 %
Uguralp	2003	Turkey	1377	5–9	12.4 %
Blum	2004	USA	378	1.5	17.3 %
Iacono	1999	Italy	2879	0–0.5	17.6 %
Kajiwara	2002	Japan	5282	7–12	18.5 %
Ip	2003	Hong Kong	561	3–5	29.6 %

Table 21.4 Constipation: prevalence rates in adults

Author	Study date	Location	Sample size	Definition	Prevalence
Hammond	1964	USA	890,394	Self-report	27.1 %
Everhart	1989	USA	11,024	Self-report	15.8 %
Sandler	1990	USA	15,014	Self-report	12.8 %
Talley	1992	USA	835	Self-report (BDQ*)	14.7 %
Talley	1993	USA	690	Rome I	19.2 %
Drossman	1993	USA	5430	Rome I	3.6 %
Johanson	1994	USA	NR**	Self-report	1.9 %
Harari	1996	USA	43,375	Self-report	3.4 %
Talley	1996	USA	1375	Self-report	40.1 %
Frexinos	1998	France	6000	Self-report	35 %
Stewart	1999	US	10,018	Rome II	14.7 %
Chiarelli	2000	Australia	41,724	Self-report	14.1–27 % (all women)
Pare	2001	Canada	1149	Rome I	16.7
				Rome II	14.9 %
				Self-report	27.2 %
Bytzer	2001	Australia	8555	Self-report	6.3–10.3 %***
Walter	2002	Sweden	1610	Self-report	19.8 % (in women)
Haug	2002	Norway	62,651	Self-report	20.2 %
Cheng	2003	Hong Kong	3282	Rome II	14.3 %
Talley	2004	New Zealand	924	Self-report	19.9 %
Garrigues	2004	Spain	349	Self-report	19.2 %
				Rome I	14 %
Siproudhis	2006	France	7196	Self-report	22.4 %
Jun	2006	Korea	1029	Self-report	16.5 %
				Rome II	9.2 %
Murakami	2006	Japan	1705	Self-report	26 % (all women)
Howell	2006	Australia	1673	Rome II	30.7 %
Murakami	2007	Japan	3835	Rome I	26.2 % (all women)

*Bowel Disease Questionnaire.

**NR, not reported.

***Based on socioeconomic class, with increased prevalence rates in patients with lower socioeconomic status.

27 % in older women (ages 70–75; [42]). In contrast to these studies, Johanson used data from the National Health Interview Survey, which involved face-to-face interviews, and identified a prevalence rate of 1.9 % [38]. The direct interview, and absence of Rome criteria, may have contributed to the low prevalence rate. Other population-based studies reported the prevalence of constipation to range from 14 % to 29 % in France, Spain, and Sweden [30,31,48].

Prevalence rates in the elderly are generally higher. The prevalence of constipation in elderly adults residing in the community ranged from 11.6 % in Asian men and women [53], to 20.3 % in the community elderly in New Zealand [54], to as high as 39.5 % in community-residing US adults [15] and 45 % in homebound elderly US adults (mean age = 79; [55]). Using a validated self-report questionnaire, Talley and colleagues identified a prevalence rate of 40.1 % (95 % CI 38.9–44.4) in a survey of 1375 adults, aged 65 and older [40]. As noted above, Chiarelli and colleagues noted a prevalence rate of 27 % in Australian women aged 70–75 [42].

A few studies have focused on prevalence rates of constipation in special populations. Active duty US

Marines and Navy servicemen had constipation rates of 7.2 % while at home, 10.4 % while on ship, and 34.1 % while deployed in the field [56]. Pregnant women experienced constipation between 16–26 % of the time during and immediately after pregnancy [57]. Fifty percent of patients undergoing thoracic surgery experienced symptoms of constipation, with nearly all restoring normal bowel function by postoperative day 17 [58]. Finally, a single study of stroke victims reported a 55.2 % rate of constipation after a first stroke at four weeks of follow-up [59].

Risk factors

A number of risk factors associated with constipation have been identified (see Table 21.5). Some risk factors, such as opioids and socioeconomic status, have been implicated in worsening constipation symptoms, while others are considered protective, such as fiber intake, tobacco and alcohol use. Opioid use increases the risk of developing constipation with an odds ratio (OR) 1.6–5.26 times above nonusers [60,61]. Similarly, nonopioid analgesics worsen constipation. Acetaminophen use of any kind has an OR of 1.92, nonsteroidal anti-inflammatory agents (NSAIDs) of 1.69 [62], and daily aspirin of 1.38 [63] compared to nonusers. Many other pharmaceutical agents including diuretics, iron and calcium supplementation, antidepressants, and antispasmodics have also been implicated in worsening constipation symptoms [60]. Female gender increases prevalence rates of constipation in most studies, although the impact varies widely with studies showing no influence on constipation symptoms (OR 1.0 [50]) to a significant impact on constipation symptoms (OR 2.9 [30]). The majority of studies had OR ranging from 1.62 to 2.3 [13,31,60]. Non-Whites are affected by constipation more frequently than Whites by approximately 50 % [32]. The impact of BMI on constipation is unclear. Chang et al. [62] showed that normal and overweight people (BMI 24.2–30) have less constipation than those with BMIs <24.2 or >30. Dukas and colleagues [63] described a significant difference in constipation in those with a BMI >29 compared to <21 (OR 0.48). Conversely, a population study in Iran did not show any difference in constipation based upon BMI [64].

Several factors appear to be protective against constipation. Tobacco use is protective with ORs of about 0.9 for current or former users [62,63], while constipation symptoms increase in the weeks following cessation [65]. Alcohol is protective against constipation, more so in daily consumers, as is coffee [62,63]. Fiber intake appears to lessen constipation as well. In a subset of females from the Nurse's Health Study, Dukas et al. showed that those who consumed >20 g of fiber per day had less constipation compared to those who ate <7 g day^{-1} (OR 0.66 [63]). Moderate fiber intake was associated with improved odds of having constipation versus either low or high fiber diets in Spain [30], while in young Japanese women the use of fiber did not change rates of constipation [66]. Being married, graduating college, and higher income levels lessen constipation [62]. Johanson showed that living in the southern United States compared to the northern areas carried a correlation coefficient of 0.291 to -0.441 in about 11 million American discharge summaries [67]. In the same study, living in metropolitan areas decreased constipation rates by a coefficient of -0.661. There is little good data on religion as a risk factor for constipation.

Natural history

The natural history of chronic constipation has been measured in several studies. In adults, one study, which was of fairly short duration, found that 89 % of patients with chronic constipation were still symptomatic when surveyed 12–20 months after the initial diagnosis [35]. A larger study of 1365 adults performed over a 12-year follow-up period found that symptoms of functional constipation resolved in 77.8 % of patients [68]. The latter study may be overly optimistic, however, since some patients were likely asymptomatic due to the use of medications. In addition, many of these patients (up to 40 %) transitioned from one functional gastrointestinal disorder (i.e. chronic constipation) to another (i.e. IBS-C or functional dyspepsia). Thus, although their symptoms of constipation resolved, other symptoms of functional gastrointestinal disorders developed. In a longitudinal study of childhood constipation ($n = 418$; median age = 8; median follow-up = 5 years), symptoms of constipation were still present in 30 % of those children after puberty, despite the use of medications [69]. A study of 47 children (60 % boys) with severe constipation in their first year of life found that 69 %

Table 21.5 Risk factors for constipation

Risk factor	Odds ratio	Sample size	Country	Reference
Sex (Female: Male)	2.9	349	Spain	Garrigues2004
	2.3	1610	Sweden	Walter S 2002
	2 (age <50)	5507	USA	Choung 2004
	1.0	1176	USA	Chang 2007
	1.62	20,795	UK	Talley 2003
Age				
>50 : <50	0.94	1176	USA	Chang 2007
35–39 : >60	0.41	62,036	USA	Dukas 2003
<40 rate (OR*)	5.9 % (1.0)	43,375	USA	Harari 1996
60–69	3.8 % (0.64)			
>80	6.3 % (1.07)			
BMI				
<24.2	1.0	1176	USA	Chang 2007
24.2–26.7	0.66			
26.7–30	0.92			
>30	1.09			
<25, 25–30, >30	1.0	18,180	Iran	Pourhoseingholi 2009
>29 : <21	0.48	62,036	USA	Dukas 2003
Coffee drinker	0.94	1176	USA	Chang 2007
Tobacco	0.89 (ever)	62,036	USA	Dukas 2003
	0.90 (current)	1176	USA	Chang 2007
	0.88 (past)			
Opioid use	5.26	10,094	USA	Pappagallo 2007
	1.6	20,795	UK	Talley 2007
Analgesics (nonopioid)				
Acetaminophen	1.92 (any)	1176	USA	Chang 2007
NSAIDS	1.69			
Aspirin	1.38	62,036	USA	Dukas 2003
Alcohol	0.89 (ever)	1176	USA	Chang 2007
	0.66 (>30 g d^{-1})	62,036	USA	Dukas 2003
Fiber	0.64 (>20 g vs. <7 g d^{-1})	62,036	USA	Dukas 2003
	0.38 (moderate)	349	Spain	Garrigues 2004
	1.05 (large)			
	1.0	3835	Japan	Murakami 2007
Exercise				
Daily:<weekly	0.56	62,036	USA	Dukas 2003
None	1.0	349	Spain	Garrigues 2004
Sometimes	0.43			
Habitual	0.31			
Geography/Income				
North : South	0.291: −0.441	~11,000,000	USA	Johanson 1998
Metropolitan	−0.688			
Public aid	0.355			
Pregnancy	1.3	20,795	UK	Talley 2003
Marital status				
Married	1.0	1176	USA	Chang 2007
Single	1.44			
Divorced	1.13			

Table 21.5 (*Continued*)

Risk factor	Odds ratio	Sample size	Country	Reference
Education		1176	USA	Chang 2007
HS, some college	1.0			
Not HS graduate	0.8			
College graduate	0.7			
Medications (nonanalgesics)		20,795	UK	Talley 2003
Diuretics	1.7			
Antidepressants	1.9			
Antispasmodics	3.3			
Iron	1.48			
Calcium	2.49			
Anticonvulsants	2.8			
Antihistamines	1.9			

had symptom resolution at the end of 6 months [70]. A longer follow-up study found that half of constipated children remain symptomatic at 5 years [71].

The impact of constipation on patients and society

Quality of life

Health-related quality of life measures are multidimensional constructs designed to capture the patient's subjective evaluation of how their medical condition affects their physical, psychological, social functioning and well-being. Quality of life (QOL) measures include generic questionnaires (i.e. SF-36, SF-12, SCL-90-R, Peds QOL, Psychological General Well-Being (PGWB), Short Portable Mental Status Questionnaire (SPMSQ)) or disease-specific questionnaires (i.e. PAC-SYM (Patient Assessment of Constipation – Symptoms), PAC-QOL (Patient Assessment of Constipation – Quality of Life), Defecation Disorder List (DDL)). Quality of life measures are important because they may help predict healthcare seeking behavior and response to therapy.

There is a limited amount of information regarding quality of life and constipation in pediatric patients. A survey of 80 constipated children (ages 5–18) using the Peds QOL questionnaire found that QOL scores were lower in children with constipation than in healthy controls and those with inflammatory bowel disease or gastroesophageal reflux disease [72]. Another study using the Peds QOL found that Australian children ($n = 51$; ages 8–18) with slow transit constipation had significantly lower QOL scores than did healthy children [73]. Lastly, a study conducted in Brazil evaluated 57 children with functional constipation (Rome II) using the CHQ-PF50 questionnaire, a 50-item survey instrument completed by the patient's parents [74]. Scores were lower in children with constipation compared to healthy controls for both physical and psychological domains.

More data is available regarding QOL in constipated adults. Using the SPMSQ, Whitehead and colleagues found that psychological distress was greater, and mental status poorer, in a study of US constipated older adults (age >65; $n = 209$ [29]). A US survey study of older adults living in the community (age >65; $n = 126$) found that SF-36 scores were lower in patients with constipation for mental health, general health perception, physical functioning, and bodily pain than those without constipation [75]. In a prospective study to validate the PAC-SYM questionnaire, adult patients (ages 18–71; $n = 216$) with constipation were found to have lower QOL scores than healthy controls using both the SF-36 and PGWB [76]. A multinational prospective survey (SF-36) of 1435 adults with constipation (Rome III criteria) found that QOL was markedly reduced compared to an identical number of nonconstipated controls. Constipated women had lower QOL scores than did constipated

men [77]. An analysis of the National Health and Wellness Survey (NHWS) found that, using the SF-12 questionnaire, patients with constipation ($n = 1430$) had lower physical and mental component scores compared to matched controls ($n = 1430$) without constipation [78]. These results confirm earlier studies involving Canadian adults (Rome II criteria), and employees of a VA healthcare system (Rome I criteria), demonstrating that constipation significantly worsens QOL compared to nonconstipated individuals [2,79]. Although data is limited, psychological distress and QOL appears to be lower in patients with normal transit constipation compared to slow transit constipation [11,79,80]. Lastly, appropriate therapeutic interventions may lead to an improvement in QOL. A small study of 16 patients (mean age = 54.5) with constipation due to puborectalis dyssynergia found that, using the PAC-QOL, scores improved along with symptoms after treatment with a guided physical therapy program [81]. Patients with slow transit constipation ($n = 59$; mean follow-up = 11 years) who underwent ileorectal anastomosis reported an improvement in symptoms and overall QOL compared to before surgery, although a direct comparison of pre- and postsurgical QOL scores using identical questionnaires was not performed [82]. Medical therapy for constipation symptoms, using prucalopride, tegaserod, lubiprostone, and linaclotide all led to an improvement in QOL scores in a number of studies [83–87].

Economics

A number of studies conducted over the past decade have shown that functional gastrointestinal disorders, such as chronic constipation, are associated with increased healthcare costs. The significant economic costs associated with treating chronic constipation arise due to direct costs associated with evaluation and treatment, as well as indirect costs, such as missing school or work (absenteeism) or not being as productive at school or work as usual (presenteeism). In the adult population, constipation was the primary diagnosis or main reason to seek care for 5.7 million patient visits in 2001, while the primary symptom of constipation led to 6.3 million patient visits in the United States in 2004 [88,89]. An analysis of the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medical Care Survey (NHAMCS) in the USA showed that ambulatory care visits for constipation increased from 4 million during 1993–1996 to 8 million during 2001–2004 [90]. Martin analyzed three different US national health surveys (2001 NAMCS, 2001 NHAMCS, and the 2001 National Hospital Discharge Survey) and estimated an annual cost of US$235 million directly attributable to the primary diagnosis of constipation for calendar year 2001 [88]. This figure did not include costs of over-the-counter or prescription medications. Although the vast majority of the visits took place in the outpatient arena, 55 % of the costs were incurred due to inpatient care. The authors determined that, overall, 2 % of the US population had an ambulatory visit for constipation; women and patients residing in metropolitan areas had higher rates of ambulatory care utilization. A survey study of over 100,000 adult patients enrolled in the California Medicaid program (mean age = 48.5; 65 % women) found that the total direct costs for constipation over a 15-month period was US$246 per patient [91]. The majority of these costs were for gastrointestinal procedures and laboratory tests. A study from a large West Coast HMO (>525,000 members) found that, in a survey of 1352 patients (mean age = 52.4; 73.4 % women), annual healthcare costs for patients with chronic constipation were US$7522 – nearly 50 % higher than for patients with IBS (US$5049), and nearly equal to those with abdominal pain (US$7646 [92]). Healthcare costs were not related to sex or race, were lower for college graduates compared to those with less education, and increased with advancing age. In a longitudinal, case-control study of women with constipation ($n = 168$; mean age = 33.2), direct medical costs were double those of controls over a 15-year period (US$63, 591 vs. $24,529; [93]). The authors noted that women with a diagnosis of constipation used more of all types of services, ranging from outpatient clinics to emergency department visits.

Although data from the pediatric population is limited, a representative household survey in the United States found that 1.1 % of children reported symptoms of constipation during a 2-year period [94]. This same study found that healthcare costs for children with constipation were three times higher than children without constipation, translating into additional healthcare costs of US$3.9 billion annually. A longitudinal study of children ($n = 250$; mean age at diagnosis = 11.6 years) with constipation found that, during

an average 13-year follow-up period, both inpatient (US$9,994) and outpatient costs (US$13,927) were four times higher compared to matched controls [95]. Depression, anxiety, and a history of otitis media were all associated with higher medical costs in the patients with constipation.

Etiology

Mechanistically, constipation can be categorized into primary or secondary causes (see Table 21.6). Primary causes of constipation include IBS with constipation, evacuation disorders, colonic inertia, and normal transit constipation [5]. Secondary causes of constipation are more numerous and can be grouped into several broad categories including medications, neurologic disorders, metabolic disorders, and mechanical obstruction. Unfortunately, individual symptoms are neither sensitive nor specific enough to enable accurate identification of the underlying pathophysiologic process causing the constipation [80,96]. A brief overview of the abnormal physiologic processes involved in the development of constipation is provided in the next section.

Table 21.6 Mechanistic classification of chronic constipation

Primary causes
- Normal transit
- Slow transit (colonic inertia)
- Irritable bowel syndrome with constipation
- Evacuation disorders (i.e. pelvic floor dyssynergia, descending perineum syndrome, intussusception)

Secondary causes
- Mechanical obstruction (masses, strictures)
- Neurologic disorders (Parkinson's, multiple sclerosis)
- Metabolic disorders (elevated serum calcium, diabetes, hypothyroidism)
- Medications (narcotics, high-dose tricyclic antidepressants)
- Anorectal disorders (prolapse, descending perineum syndrome)
- Psychogenic (anorexia, severe depression)
- Dietary/lifestyle (low fiber intake; ignoring call to stool)
- Iatrogenic (prior surgery)

Abnormal colonic and anorectal physiology

Constipation is a symptom, rather than a disease unto itself. A number of different conditions may cause constipation, and these secondary causes can be classified as structural, mechanical, metabolic, or medication-related (see Table 21.6). In the colon, a number of different pathophysiologic processes may lead to constipation [97–106]. Primary constipation is typically categorized as normal transit constipation, slow transit constipation, obstructed defecation (pelvic floor dysfunction), or irritable bowel syndrome.

Normal transit constipation can be a difficult concept to understand in the evaluation of a patient with constipation. These patients complain of constipation (infrequent stools, bloating, fullness, abdominal pressure), although there is no evidence of a mechanical obstruction, and colonic motility and pelvic floor function are both normal. This is considered a functional gastrointestinal disorder.

Slow transit constipation typically develops because of a neuropathic process [102,103]. In slow transit constipation, the number of HAPCs may be reduced in the postprandial period, and thus colonic transit and the number of mass movements is reduced. Alternatively, the number of HAPCs may be normal, although they are uncoordinated. Abnormalities in rectal function can also occur. In this setting, abnormally strong rectal contractions can impede the flow of colonic contents distally. Constipation may also develop because of injury to the pacemaker cells in the colon – the interstitial cells of Cajal, or abnormalities in sensory processing [99]. In the latter situation, injury to rectal sensory afferents prevents the initiation of a normal rectal reflex. Finally, uncommon disorders of smooth muscle (scleroderma, amyloidosis, hollow visceral myopathy) may lead to a myopathic process and loss of contraction within the colon.

Defecatory disorders encompass a variety of abnormalities in the pelvic floor [104,105]. A large rectocele or sigmoidocele, rectal prolapse or intussusception may all impede the normal evacuation of stool. Less commonly, descending perineum syndrome (descent of greater than 3 cm) can develop, or the rectum may have diminished contractile function. Pelvic floor dyssynergia, a condition where the internal anal sphincter fails to relax properly or the external anal sphincter inappropriately contracts during attempted defecation, is primarily a disorder of women.

Conclusion

Chronic constipation is a highly prevalent disorder found worldwide. Risk factors for the development of constipation are many, although diet, gender, age, and lower socioeconomic status are some of the most important. Constipation affects both young and old alike, and is more likely to affect women than men. As a symptom rather than a specific disorder, constipation represents a number of different pathophysiologic processes. Symptoms, unfortunately, are nonspecific and cannot be used to accurately predict either the pathophysiology of the disorder or the response to treatment. Similar to other functional gastrointestinal disorders, symptoms are chronic in nature for many patients. The chronicity associated with constipation directly contributes to the negative impact of this disorder on quality of life and healthcare economics. The success of future treatments should be judged, in part, on improving quality of life and reducing the economic impact of this common disorder.

Multiple choice questions

1 Which definition best fits chronic constipation as defined by the Rome III criteria?

A Infrequent bowel movements for more than 3 months

B Straining at stool and feelings of incomplete evacuation for more than 6 months

C Active symptoms of constipation during the last 3 months (straining, incomplete evacuation, need for manual maneuvers) present at least 25 % of defecations, with onset 6 months ago

D The presence of lower abdominal pain or discomfort with symptoms of constipation

E Symptoms of constipation present at least 3 days per week, and not meeting criteria for IBS

2 The evaluation and treatment of chronic constipation is important because:

A Symptoms of chronic constipation for >15 years increase the likelihood of colon cancer

B Symptoms markedly reduce patients' quality of life and increase healthcare costs

C The natural history of chronic constipation is that nearly all children and adults remain symptomatic

D Symptoms can be easily confused with IBS-C, which is more difficult to treat

E Symptoms are nonspecific and frequently hide serious underlying organic disorders

3 Risk factors for the development of constipation include:

A Opioids, nonnarcotic analgesics, low socioeconomic class, and high fiber intake

B Opioids, high socioeconomic class, BMI >33, and low fiber intake

C Opioids, low socioeconomic class, tobacco use, alcohol use, and BMI >35

D Opioids, nonnarcotic analgesics, low socioeconomic class, and low fiber intake

E Opioids, low socioeconomic class, low fiber intake, religion, and alcohol use

4 Analysis of chronic constipation prevalence rates in adults has determined a mean prevalence rate of approximately:

A 1.9 %

B 40.1 %

C 27.1 %

D 14 %

E 3.4 %

5 Mechanistically, chronic constipation can be categorized into primary and secondary causes. Primary causes of constipation include:

A Opioids

B Normal transit, slow transit, irritable bowel syndrome with constipation, and evacuation disorders (i.e. pelvic floor dyssynergia, descending perineum syndrome, intussusception)

C Slow transit, IBS with constipation, and evacuation disorders

D Slow transit constipation and IBS with constipation

E Opioids, nonnarcotic analgesics, antibiotics, and iron supplements

References

1 Higgins PDR, Johanson JF. Epidemiology of chronic constipation in North America: a systematic review. *Am J Gastroenterol* 2004;99:750.

2 Irvine EJ, et al. Health-related quality of life in functional GI disorders: Focus on constipation and resource utilization. *Am J Gastroenterol* 2002;97: 1986.

3 Sonnenberg A, Koch TR. Physician visits in the United States for constipation: 1958 to 1986. *Dig Dis Sci* 1989;34:606.
4 Longstreth GF, et al. Functional bowel disorders. *Gastroenterology* 2006;130:1480.
5 Lembo A, Camilleri M. Chronic constipation. *New Engl J Med* 2003;349:1360.
6 Connell AM, et al. Variation of bowel habit in two population samples. *BMJ* 1965;5470:1095.
7 Heaton KW, et al. Defecation frequency and timing, and stool form in the general population: a prospective study. *Gut* 1992;33:818.
8 Herz MJ, et al. Constipation: a different entity for patients and doctors. *Fam Pract* 1996;13:156.
9 Sandler RS, Drossman DA. Bowel habits in young adults not seeking health care. *Dig Dis Sci* 1987;32:841.
10 Harari D, et al. How do older persons define constipation? Implications for therapeutic management. *J Gen Intern Med* 1997;12:63.
11 Glia A, Lindberg G. Quality of life in patients with different types of functional constipation. *Scan J Gastroenterol* 1997;32:1083.
12 Chitkara DK, et al. Medical presentation of constipation from childhood to early adulthood: a population-based cohort study. *Clin Gastroenterol Hep* 2007;5:1059.
13 Choung RS, et al. Cumulative incidence of chronic constipation: a population-based study 1998–2003. *Aliment Pharmacol Ther* 2007;26:1521.
14 Robson KM, et al. Development of constipation in nursing home residents. *Dis Colon Rectum* 2000;43: 940.
15 Porter CK, et al. The incidence and gastrointestinal risk of functional gastrointestinal disorders in a healthy US adult population. *Am J Gastroenterol* 2001;106:130.
16 Talley NJ, et al. Functional constipation and outlet delay: a population-based study. *Gastroenterology* 1993;105:781.
17 Sandler RS, et al. Demographic and dietary determinants of constipation in the US population. *Am J Public Health* 1990;80:185.
18 Hale WE, et al. Symptom prevalence in the elderly. An evaluation of age, sex, disease, and medication use. *Am J Geriatr Soc* 1986;34:333.
19 Harari D, et al. How do older persons define constipation? Implications for therapeutic management. *J Gen Intern Med* 1997;12:63.
20 Miele E, et al. Functional gastrointestinal disorders in children: an Italian prospective survey. *Pediatrics* 2004;114:73.
21 Roma-Giannikou E, et al. Epidemiology of chronic constipation in Greek children. *Hell J Gastroenterol* 1999;12:58.
22 Ludvigsson JF, et al. Epidemiological study of constipation and other gastrointestinal symptoms in 8000 children. *Acta Pediatrica* 2006;95:573.
23 Uguralp S, et al. Frequency of enuresis, constipation and enuresis association with constipation in a group of school children aged 5–9 years in Malatya, Turkey. *Turk J Med Sci* 2003;33:315.
24 Blum NJ, et al. Why is toilet training occurring at older ages? A study of factors associated with later training. *J Pediatr* 2004;145:107.
25 Iacono G, et al. Gastrointestinal symptoms in infancy: a population-based prospective study. *Dig Liver Dis* 2005;37:432.
26 Kajiwara M, et al. The micturition habits and prevalence of daytime urinary incontinence in Japanese primary school children. *J Urol* 2004;171:403.
27 Ip KS, et al. A community-based study of the prevalence of constipation in young children and the role of dietary fibre. *Hong Kong Med J* 2005;11:431.
28 van Ginkel R, et al. Childhood constipation: longitudinal follow-up beyond puberty. *Gastroenterology* 2003;125:357.
29 Whitehead WE, et al. Constipation in the elderly living at home: definition, prevalence, and relationship to lifestyle and health status. *J Am Geriatric Soc* 1989;37:423.
30 Garrigues V, et al. Prevalence of constipation: Agreement among several criteria and evaluation of diagnosis accuracy of qualifying symptoms and self-reported definition in a population-based survey in Spain. *Am J Epidemiol* 2004;5:520.
31 Walter S, et al. A population-based study on bowel habits in a Swedish community: prevalence of faecal incontinence and constipation. *Scand J Gastroenterol* 2002;8:911.
32 Stewart WF, et al. Epidemiology of constipation (EPOC) study in the United States: relation of clinical subtypes to sociodemographic features. *Am J Gastroenterol* 1999;12:3530.
33 Hammond E. Some preliminary findings on physical complaints from a prospective study of 1,064,004 men and women. *Am J Pub Health* 1964;54:130.
34 Everhart JE, et al. A longitudinal survey of self-reported bowel habits in the United States. *Dig Dis Sci* 1989;34:1153–62.
35 Talley NJ, et al. Onset and disappearance of gastrointestinal symptoms and functional gastrointestinal disorders. *Am J Epidemiol* 1992;136:165.
36 Talley NJ, et al. Functional constipation and outlet delay: a population-based study. *Gastroenterology* 1993;105:781.
37 Drossman DA, et al. U.S. householder survey of functional gastrointestinal disorders. Prevalence,

sociodemography, and health impact. *Dig Dis Sci* 1993;38:1569.
38 Johanson JF. (1994) Constipation, in *Digestives Diseases in the United States: Epidemiology and Impact*, Vol. 94-1447 (ed. JE Everhart), US Department of Health and Human Services, Public Health Service, National Institutes of Health, NIDDK, Washington, DC, p. 567.
39 Harari D, et al. Bowel habit in relation to age and gender. Findings from the National Health Interview Survey and clinical implications. *Arch Intern Med* 1996;156:315.
40 Talley NJ, et al. Constipation in an elderly community: a study of prevalence and potential risk factors. *Am J Gastroenterol* 1996;91:19.
41 Frexinos J, et al. Descriptive study of digestive functional symptoms in the French general population. *Gastroenterol Clin Biol* 1998;22:785.
42 Chiarelli P, et al. Constipation in Australian women: prevalence and associated factors. *Int Urogynecol* 2000;11:71.
43 Pare P, et al. An epidemiological survey of constipation in Canada: definitions, rates, demographics, and predictors of health care seeking. *Am J Gastroenterol* 2001;96:3130.
44 Bytzer P, et al. Low socioeconomic class is a risk factor for upper and lower gastrointestinal symptoms: a population-based study in 15,000 Australian adults. *Gut* 2001;49:66.
45 Haug TT, et al. Are anxiety and depression related to gastrointestinal symptoms in the general population? *Scand J Gastroenterol* 2002;37:294.
46 Cheng C, et al. Coping strategies, illness perception, anxiety and depression of patients with idiopathic constipation: a population-based study. *Aliment Pharmacol Ther* 2003;18:19.
47 Talley NJ, et al. Obesity and chronic gastrointestinal tract symptoms in young adults: a birth cohort study. *Am J Gastroenterol* 2004;99:1807.
48 Siporoudihis L, et al. Defecation disorders: a French population survey. *Dis Colon Rectum* 2006;49: 219.
49 Jun DW, et al. A population-based study on bowel habits in a Korean community: prevalence of functional constipation and self-reported constipation. *Dig Dis Sci* 2006;51:1471.
50 Murakami K, et al. Dietary intake in relation to self-reported constipation among Japanese women aged 18–20 years. *Eur J Clin Nutr* 2006;60:650.
51 Howell SC, et al. Low social class is linked to upper gastrointestinal symptoms in an Australian sample of urban adults. *Scand J Gastroenterol* 2006;41: 657.
52 Murakami K, et al. Food intake and functional constipation: a cross-sectional study of 3835 Japanese women aged 18–20 years. *J Nutr Sci Vitaminol* 2007;53:30.
53 Wong ML, et al. Sociodemographic and lifestyle factors associated with constipation in an elderly Asian community. *Am J Gastroenterol* 1999;94:1283.
54 Campbell AJ, et al. Factors associated with constipation in a community based sample of people aged 70 years and over. *J Epidemiol Community Health* 1993;47:23.
55 Wolfsen CR, et al. Constipation in the daily lives of frail elderly people. *Arch Fam Med* 1993;2:853.
56 Sweeney WB, et al. The constipated serviceman: prevalence among deployed U.S. troops. *Military Medicine* 1993;8:546.
57 Bradley CS, Kennedy CM, Turcea AM, et al. Constipation in pregnancy: prevalence, symptoms, and risk factors. *Obstet Gynecol* 2007;6:1351.
58 Rasmussen LS, Pedersen PU. Constipation and defecation pattern the first 30 days after thoracic surgery. *Scand J Caring Sci* 2010;2:244.
59 Su Y, Zhang X, Zeng J, et al. New-onset constipation at acute stage after first stroke: Incidence, risk factors, and impact on the stroke outcome. *Stroke* 2009;40:1304.
60 Talley NJ, et al. Risk factors for chronic constipation based on a general practice sample. *Am J Gastroenterol* 2003;5:1107.
61 Pappagallo M. Incidence, prevalence, and management of opioid bowel dysfunction. *Am J Surgery* 2001;November(Suppl):11S.
62 Chang JY, et al. Risk factors for chronic constipation and a possible role of analgesics. *Neurogastroenterol Motil* 2007;19:905.
63 Dukas L, et al. Association between physical activity, fiber intake, and other lifestyle variables and constipation in a study of women. *Am J Gastroenterol* 2003;8:1790.
64 Pourhoseingholi MA, et al. Obesity and functional constipation. A community-based study in Iran. *J Gastrointest Liver Dis* 2009;2:151
65 Hajeke P, et al. Stopping smoking can cause constipation. *Addiction* 2003;98:1563.
66 Murakami K, et al. Association between dietary fiber, water, and magnesium intake and functional constipation among young Japanese women. *European J Clin Nutr* 2007;61:616.
67 Johanson JF. Geographic distribution of constipation in the United States. *Am J Gastroenterol* 1998;2:188.
68 Halder SLS, et al. Natural history of functional gastrointestinal disorders: a 12-year longitudinal population-based study. *Gastroenterology* 2007;133:799.
69 van Ginkel R, et al. Childhood constipation: longitudinal follow-up beyond puberty. *Gastroenterology* 2003:125:357.

70 Van den Berg MM, et al. Functional constipation in infants: a follow-up study. *J Ped* 2005;147:700.
71 Staiano A, et al. Long-term follow-up of children with chronic idiopathic constipation. *Dig Dis Sci* 1994;39:561–4.
72 Youssef NN, et al. Chronic childhood constipation is associated with impaired quality of life: a case controlled study. *J Ped Gastro Nutr* 2005;41:56.
73 Clarke MCC, et al. Quality of life in children with slow transit constipation. *J Pediatr Surg* 2008;43:320.
74 Faleiros FT, Machado NC. Assessment of health-related quality of life in children with functional defecation disorders. *J Pediatr (Rio J)* 2006;82:421.
75 O'Keefe EA, et al. A bowel symptom questionnaire for the elderly. *J Gerontology* 1992;47:M116.
76 Frank L, et al. Psychometric validation of a constipation symptom assessment questionnaire. *Scan J Gastroenterol* 1999;34:870.
77 Wald A, et al. The burden of constipation on quality of life: results of a multinational survey. *Aliment Pharmacol Ther* 2007;26:227.
78 Sun SX, et al. Impact of chronic constipation on health-related quality of life, work productivity, and healthcare resource use: an analysis of the National Health and Wellness Survey. *Dig Dis Sci* 2011; ePub 2011 March 6.
79 Wald A, et al. Psychological and physiological characteristics of patients with severe idiopathic constipation. *Gastroenterology* 1989;97:932.
80 Grotz RL, et al. Discriminant value of psychological distress, symptom profiles, and segmental dysfunction in outpatients with severe idiopathic constipation. *Gut* 1994;35:798.
81 Lewicky-Gaupp C, et al. Successful physical therapy for constipation related to puborectalis dyssynergia improves symptom severity and quality of life. *Dis Col Rectum* 2008;51:1686.
82 Hassan I, et al. Ileorectal anastomosis for slow transit constipation: Long-term functional and quality of life results. *J Gastrointest Surg* 2006;10:1330.
83 Quigley EMM, et al. Clinical trial: the efficacy, impact on quality of life, and safety and tolerability of prucalopride in severe chronic constipation: a 12week, randomized, double-blind, placebo-controlled study. *Aliment Pharmacol Ther* 2006;29:315.
84 Dubois D, et al. Psychometric performance and clinical meaningfulness of the Patient Assessment of Constipation – Quality of Life questionnaire in prucalopride (RESOLOR) trials for constipation. *Neurogastroenterol Motil* 2010;22:e54.
85 Chan AO, et al. Efficacy of tegaserod for functional constipation in Chinese subjects: a randomized, double-blind controlled trial in a single centre. *Aliment Pharmacol Ther* 2007;25:463.
86 Johanson JF, et al. Multicenter, 4-week, double-blind, randomized, placebo-controlled trial of lubiprostone, a locally acting type-2 chloride channel activator, in patients with chronic constipation. *Am J Gastroenterol* 2007;103:170.
87 Lembo A, et al. Efficacy of linaclotide for patients with chronic constipation. *Gastroenterology* 2010;138:886.
88 Martin BC, et al. Direct medical costs of constipation in the United States. *Managed Care Interface* 2006;19:43.
89 Everhart JE, et al. Burden of digestive diseases in the United States. Part II: lower gastrointestinal diseases. *Gastroenterology* 2009;136:741.
90 Shah ND, et al. Ambulatory care for constipation in the United States, 1993–2004. *Am J Gastroenterol* 2008;103:1746.
91 Singh G, et al. Use of health care resources and cost of care for adults with constipation. *Clin Gastroenterol Hepatol* 2007;5:1053–8.
92 Nyrop KA, et al. Costs of health care for irritable bowel syndrome, chronic constipation, functional diarrhoea and functional abdominal pain. *Aliment Pharmacol Ther* 2007;26:237.
93 Choung RS, et al. Longitudinal direct medical costs associated with constipation in women. *Aliment Pharmacol Ther* 2011;33:251.
94 Liem O, et al. Health utilization and cost impact of childhood constipation in the United States. *J Pediatr* 2009;154:258–62.
95 Choung RS, et al. Direct medical costs of constipation from childhood to early adulthood: a population-based birth cohort study. *J Pediatr Gastroenterol Nutr* 2011;52:47.
96 Knowles CH, et al. Linear discriminant analysis of symptoms in patients with chronic constipation. *Dis Colon Rectum* 2000;43:1419.
97 Tzavella K, et al. Decreased substance P levels in rectal biopsies from patients with slow transit constipation. *Eur J Gastroenterol Hepatol* 1996;8:1207.
98 Cortesini C, et al. Nitric oxide synthase and VIP distribution in enteric nervous system in idiopathic chronic constipation. *Dig Dis Sci* 1995;40:2450.
99 He CL, et al. Decreased interstitial cell of Cajal volume in patients with slow-transit constipation. *Gastroenterology* 2000;118:14.
100 Mertz H, et al. Physiology of refractory chronic constipation. *Am J Gastroenterol* 1999;94:609.
101 Ashraf W, et al. An examination of the reliability of reported stool frequency in the diagnosis of idiopathic constipation. *Am J Gastroenterol* 1996;91:26.

102 Preston DM, Lennard-Jones JE. Severe chronic constipation of young women: "idiopathic slow transit constipation." *Gut* 1986;27:41.
103 Nyam DC, et al. Long-term results of surgery for chronic constipation. *Dis Colon Rectum* 1997;40:273.
104 Kuijpers HC, Bleijenberg G. The spastic pelvic floor syndrome: a cause of constipation. *Dis Colon Rectum* 1985;28:669.
105 Preston DM, Lennard-Jones JE. Anismus in chronic constipation. *Dig Dis Sci* 1985;28:413–18.
106 Johanson JF, et al. Association of constipation with neurologic diseases. *Dig Dis Sci* 1992;37:179.

Answers to multiple choice questions

1. C
2. B
3. D
4. D
5. B

22 Epidemiology of diverticular disease

Robin Spiller & David Humes
Nottingham Digestive Disease Biomedical Research Unit, University of Nottingham, Queen's Medical Centre, Nottingham, UK

Key points

- Diverticulosis is a common gastrointestinal disorder whose prevalence increases with age.
- The majority of patients with diverticulosis will remain asymptomatic.
- The disease burden associated with diverticulosis and its complications are increasing.
- The mortality associated with the spectrum of complicated diverticular disease is high.

Clinical summary

Diverticulosis and its complications represent a significant burden on healthcare utilization [1]. The underlying structural abnormality, diverticulosis is common in Western populations, predominately affecting the left side of the colon [2,3] which is in contrast to that seen in Oriental populations where a predominantly right-sided distribution is seen [4–6]. It is important to be clear in defining the terminology used to identify the various clinical manifestations of diverticulosis and these are clarified in Table 22.1 [7]. The majority of patients are asymptomatic [8]. The main disease burden in primary care is those who present with diverticular disease, common symptoms being a change in bowel habit and abdominal pain. The principle burden of disease in secondary care is from the complications associated with diverticulosis, which include inflammation, perforation, stricture, fistula, bleeding, and abscess formation. Diverticulosis becomes more common with increasing age and complications are more likely in those with significant comorbidity so an increase in disease burden is expected with the aging population. Diagnosis typically follows screening for colorectal cancer in asymptomatic subjects or investigation for underlying symptoms such as change in bowel habit, abdominal pain, or rectal bleeding and is made by either endoscopic (colonoscopy/flexible sigmoidoscopy) or radiological (computed tomography) means. Current treatments for symptomatic improvements in patients with diverticular disease are limited [7]. In those with complicated diverticular disease treatment often involves surgical intervention [7].

Diverticulosis/diverticular disease

Incidence and prevalence

The underlying incidence and prevalence of diverticulosis is difficult to determine as the condition is asymptomatic and requires either radiological or endoscopic investigation to be confirmed. The prevalence of diverticulosis has been reported in autopsy, endoscopy, and radiological studies [8]. A small sample of 264 patients who were representative of the general population of asymptomatic patients from General Practices in Oxford (United Kingdom) underwent a barium

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 22.1 Disease definitions

Term	Definition
Colonic diverticulosis	The presence of one or more mucosal outpouchings (diverticulum) through the large bowel wall
Colonic diverticular disease	Colonic diverticulosis associated with attributable symptoms such as change in bowel habit or pain
Acute diverticulitis	Acute inflammation associated with a diverticulum
Complicated diverticular disease	Development of a complication associated with the diverticulum of perforation, bleeding, stricture, fistula, or abscess formation

Source: Reproduced from Humes et al. 2011 [7].

follow-through study which identified 33 % (88/264) as having diverticulosis over the age of 45 years [9]. The prevalence increased with age with 25 % (36/144) under the age of 60 years and 43 % (52/120) in those greater than 60 years [9]. A smaller study in Oxford of 109 subjects with no gastrointestinal symptoms who underwent a barium follow-through study reported a prevalence of 7.6 % (5/66) in those under the age of 60 years, 34.9 % (15/43) in those over the age of 60 years [10]. There is therefore an increase in the prevalence of diverticulosis associated with aging. An autopsy study of patients in Belfast reported a prevalence of 37 % (111/300) with a female predominance (42 % vs. 33 %) and an increase in prevalence with increasing age [11]. An autopsy study from Norway of 280 patients reported similar results with a female predominance and an increase in frequency of diverticulosis with increasing age, which was 46 % for those aged >65 years [3]. Autopsy, radiological and endoscopic studies have reported widely on the anatomical location of diverticulosis and there appears to be a substantial difference in location with regard to geographical location. Studies reporting anatomical location from Singapore and Japan have confirmed predominantly right-sided disease with studies from Europe and North America confirming mainly left-sided disease with the most common location being the sigmoid colon [3,4,6]. A study of 12,335 barium enemas in Edinburgh used census data to report on the incidence of diverticular disease. The study reported an incidence of 1.55 per 1000 population [12]. A prospective cohort study of patients from the European Prospective Investigation into Cancer and Nutrition (EPIC) Oxford cohort using record linkage to Hospital Episodes Statistics (HES) data (see Chapter 9 for explanation of data source) reported an incidence of diverticular disease which included all patients with a hospital diagnosis of diverticular disease of 148 per 100,000 person-years of follow-up [13].

Those studies reporting the prevalence from endoscopic and radiological series have inherent bias as the majority of these patients will have reported symptoms to initiate investigation of the colon. The two small studies in Oxford attempted to account for this by selecting asymptomatic individuals; however, only a small sample of the population was assessed in each study and the nature by which the samples were obtained was not clear. The results from autopsy studies vary widely with reporting bias resulting from whether the autopsy was performed to detail the occurrence of diverticulosis or for other reasons. The study from Edinburgh only included symptomatic cases and will have been biased by the selection of patients for testing. The record linkage study by Crowe et al. gives an overall estimate of incidence for diverticulosis and its complications. The incidence quoted is high compared to the incidence of, for example, perforated diverticular disease (4 per 100,000 person-years) [13,14]. It is likely the true incidence and prevalence of diverticulosis will remain unknown, though the recent adoption of colorectal cancer screening in asymptomatic subjects may give less biased data for those in the screened age group in the future.

The occurrence of diverticular disease is difficult to determine. Some authors have questioned whether diverticulosis contributes to colonic symptoms resulting in diverticular disease [15]. A follow-up study of 261 patients diagnosed with barium enema proven diverticulosis demonstrated that 36 % (94/261) experienced short-lived pain on a median of five days per month with a median duration of 3 hours [16], though when taking only those with nonpainful indications for investigation (iron deficiency/family history of cancer) the proportion with chronic pain fell to 20 %. These patients did not meet the strict Rome II criteria for irritable bowel syndrome mainly because of lack of relief of pain by defecation. It is likely that a percentage of patients do develop symptoms secondary

to their diverticulosis [17]. It is difficult to determine the prevalence and incidence of diverticular disease. A study of hospital admissions using HES data coded for diverticular disease using International Classification of Disease (ICD) codes reported hospital admission rates of 28.0 per 100,000 population with a significant increase in admissions for both men (16 %) and women (12 %) over a 10-year period (1989/1990 to 1999/2000) [18]. Admission rates were lowest in those under the age of 35 years (2.0 per 100,000) and greatest in those over 85 years (250.9 and 314.2 per 100,000 for men and women, respectively). A further study of HES data reported all admissions with diverticular disease and reported an increase in admissions from 0.56 to 1.20 per 1000 population per year (1996/2006) with an increasing number of admissions with increasing age [19]. These studies were performed in secondary care only and it is likely the majority of disease burden associated with diverticular disease occurs in primary care. There are as yet no estimates of disease burden for primary care where the majority of these patients will be treated. ICD-10 coding identifies all cases with diverticular disease but neither study was able to give specific admission rates for diverticular disease and included admissions with acute diverticulitis and complicated disease. Both studies reported an increase in the number of admissions with diverticular disease with the biggest change identified in those having day-case procedures [19]. This increase in day-case procedures likely reflects increased use of endoscopy and therefore an increased ascertainment of the diagnosis of diverticulosis. However, the data does suggest an increase in both elective and emergency admissions, the latter suggesting an increase in disease burden.

Mortality

Mortality rates using HES data for England have reported a 30-day mortality rate associated with an admission for diverticular disease of 5.1 % and a 1-year mortality rate of 14.5 % between 2000 and 2005. The 30-day and 1-year mortality associated with surgery for diverticular disease was 10.1 % and 15.5 % respectively, with those undergoing emergency surgery having a 30-day and 1-year mortality of 15.9 % and 22.8 %, respectively [19]. The independent predictors of mortality included increasing age, comorbidity as indicated by the Charlson index [20], surgery, and an emergency admission [19]. Age-standardized mortality rates from 1979 to 1999 remained constant but were significantly higher in females than males at all time points [18]. In the United States mortality rates from the National Center for Health Statistics for 1998 reported a mortality rate of 2.5 per 100,000 population with a female excess (1.74 vs. 0.76 per 100,000 population) [21]. All these studies reported mortality rates coded for using ICD codes which encompass all cases of diverticular disease including complications and diverticulosis. It is likely therefore that they represent an overestimate of the mortality associated with diverticulosis and diverticular disease as the majority of the mortality associated with diverticulosis is from its complications, as confirmed by the excess mortality associated with surgery for the condition and emergency admissions [19].

Risk factors

Painter and Burkitt first hypothesized that refinement of dietary carbohydrates resulted in a deficiency of vegetable fiber from the diet which resulted in the development of diverticulosis [22]. To support their hypothesis they suggested that the increasing incidence of diverticulosis in Western countries was seen alongside a reduction of total dietary fiber and supported this by the low prevalence of diverticulosis from autopsy studies in Africa where fiber intake was assumed to be higher. Subsequently two small case-control studies reported that low fiber intake was associated with diverticular disease [23,24]. A small cross-sectional study of diverticulosis in vegetarian and nonvegetarians confirmed that diverticulosis was more common among nonvegetarians than vegetarians (33 % vs. 12 %) [9]. A prospective cohort study demonstrated that dietary fiber intake was inversely associated with risk of diverticular disease (relative risk (RR) 0.58; 95 % CI 0.41–0.83 for men in the highest compared with the lowest quintile of dietary fibre) [25]. These findings have again been confirmed recently in a large cohort study which reported a 41 % lower risk of acquiring a diagnosis of diverticular disease when comparing the highest quintile of dietary fiber intake to the lowest (RR 0.59; 95 % CI 0.46–0.78) [13]. The initial paper by Painter and Burkitt provided little direct evidence of an association between diverticulosis and dietary fiber and

subsequent studies have been small, liable to have residual confounding and not include a baseline assessment of the colon to establish the presence of diverticulosis [26]. This has resulted in calls for prospective studies and a wider examination of dietary factors in the development of diverticulosis and diverticular disease as their etiology is likely multifactorial [26,27].

Consumption of red meat has been reported to increase the risk of development of diverticular disease in two cohort studies with meat eaters having a 1.4 % increased risk of hospitalization than non-meat eaters [13,25]. The role of alcohol, caffeine, and smoking in the development of diverticular disease was studied in a prospective cohort study and no relationship between the development of symptoms and their use was found [28]. A subsequent study in Sweden of a cohort of 35,809 women from the Swedish Mammography Cohort reported an increased risk in both smokers and ex-smokers associated with symptomatic diverticular disease; however, the association did not reach statistical significance for current smoking (RR 1.23; 95 % CI 0.99–1.52) [29]. Vigorous physical activity has been reported to reduce the risk of development of diverticular disease when compared to those not undertaking physical activity (RR 0.60; 95 % CI 0.41–0.87) [30]. An increased body mass index (BMI) in men has been reported to increase the likelihood of diverticular disease [31]. A Swedish cohort study of 7494 men who were followed up by record linkage found that those with a BMI of 30 kg m^{-2} or greater had a fourfold increase in the risk of hospitalization [32]. The evidence suggests that lifestyle and dietary factors are important in both the development of diverticulosis and subsequent symptomatic disease. There is a clear need for larger well-designed studies to try and answer these important questions as on a population level this represents the best opportunity for intervention to reduce the disease burden associated with this condition [33].

Occurrence of other diseases

The patterns of occurrence of disease are similar for both colorectal cancer and diverticulosis. Both tend to occur later in life and may present with similar disturbances of bowel habit. There have been some reports that have suggested an increased risk of colonic polyps and colorectal cancer associated with a diagnosis of diverticulosis or acute diverticulitis [34–37]. Two endoscopic studies have not found an increased risk of colorectal cancer or polyps in patients with diverticular disease [38,39]. Three of these studies reporting an increased risk used the same cohort, 7159 patients from Central Sweden, using standardized incidence ratios and then a case-control design. The finding of an increased risk of colorectal cancer was greatest in the first two years following diagnosis [34]. A recent case-control study of all patients with colon cancer from the Swedish Cancer Registry reported an increased risk of having a colon cancer diagnosed in the first 12 months following an initial diagnosis of diverticular disease but no subsequent increase in risk or mortality [40]. The risk was greatest in the first 6 months following a diagnosis of diverticular disease and the authors stated that a screening effect could not be excluded as the explanation of the increased occurrence. The authors of this study advise that all patients with a new diagnosis of suspected diverticular disease should have investigation of the colon to ensure they do not have a concurrent cancer which is in line with current guidance [41].

Acute diverticulitis

Occurrence

Studies reporting the incidence of acute diverticulitis have concentrated on hospitalized cases only. A study of data from the US Nationwide Inpatient Sample (NIS) reported an increase in incidence of hospital admissions coded for acute diverticulitis from 0.59 to 0.71 per 1000 population over the period 1998–2005 [42]. This represented a 26 % increase in age-adjusted admissions from 1998 to 2005. The incidence increased with increasing age but the greatest increase in incidence came in those aged 18–44 years (82 %) and 45–74 years (36 %) [42]. There was no increase in incidence among different ethnic groups over this period. A second analysis of the same dataset, which did not exclude colorectal cancer patients, revealed a similar increase in hospital admissions with an increase from 61.8 to 75.5 per 100,000 population. The authors also noted a female predominance to admissions except in the lowest age group. A geographical difference in incidence rates was noted with the Northeast having the greatest

percentage increase in admissions and the West having both the lowest rates of hospitalization and smallest percentage increase in admissions [43]. A further study of the NIS from 2002 to 2007 reported a 9.5 % year on year rise in hospitalization for acute diverticulitis with 85 % of these emergency admissions being treated medically [44]. There were few changes in the ethnicity of the patients admitted over the study period but a substantial proportion of cases (25 %) did not have this data available. An analysis of discharge data from the Office of Statewide Health Planning and Development in California from 1995–2006 reported a 2.1 % estimated annual percentage change for admissions with acute diverticulitis confirming the observations in the NIS data [45]. The authors also reported the largest increase in admissions in those aged 20–34 years and 35–49 years although they only contributed to a small number of the overall cases.

These studies all fail to represent the true disease burden associated with acute diverticulitis as many cases will be treated in the community [7,46]. Possible explanations for this increase in acute diverticulitis could be systematic biases in the data used due to systematic coding errors. However, despite differences in coding from different institutions it is likely that trends at a population level would represent an underestimate of disease. There could have been systematic changes in the diagnosis of acute diverticulitis over time resulting in an increase in the diagnosis of more mild cases of the disease being treated in hospital. This should have resulted in changes in diagnosis at all age levels and not the differential patterns of age-related increase that have been described. The increase in hospital admissions noted in these studies from the NIS for patients with acute diverticulitis mirror increases in admissions reported from UK HES data [18,19]. The studies from the UK were unable to report directly on rates of acute diverticulitis as ICD-10 fails to code it individually. If these data are to be believed then further studies are required to try and determine the underlying factors responsible for this increase in disease occurrence and thus allow the development of potential preventative strategies.

Rates of recurrence of acute diverticulitis were first reported by Parks from his study of 521 patients admitted with diverticular disease [47]. A 26 % (78/297) recurrence rate was reported for those treated medically on their first admission with 46 % (36/78) occurring in the first year after the initial attack. A retrospective study of 672 patients admitted with acute diverticulitis reported a 5-year recurrence of 36 % (95 % CI 31.4–40.6 %) with those with a long disease segment and an abscess having the greatest risk of recurrence [48]. A Swedish study of 234 patients admitted with a first episode of acute diverticulitis reported an overall readmission rate of 20.8 % (46/221) following medical treatment [49]. A further study of 502 patients from New Zealand reported a recurrence rate of 18.8 % (60/320) following conservative management for acute diverticulitis with no increase in mortality related to recurrent episodes [50]. A recurrence rate of 42 % (234/55) of patients treated nonsurgically for acute diverticulitis was reported from a study from Finland [51]. In only 3.9–5 % of cases did complicated disease arise following the initial attack [48,50]. These studies are all small and have varying diagnostic criteria for acute diverticulitis and its recurrence. A study of 314 patients treated nonoperatively for acute diverticulitis from California reported a recurrence rate of 13.3 % with age >50 years associated with a lower recurrence rate and those with higher comorbidity having the greatest recurrence rate [52]. The rates of recurrence vary considerably and occur within a short time period following the index admission. These studies are all limited to episodes that are documented in secondary care and mandate hospital admission and thus may underestimate the true rate of recurrence as mild attacks may be treated in primary care. Whilst these cases will not represent a significant burden in terms of mortality, they may well increase the burden placed on healthcare resources from recurrent episodes.

Mortality

A reduction in the mortality associated with a hospital admission with acute diverticulitis has been reported with a decrease from 1.6 % in 1998 to 1.0 % in 2004–2005 and a reduction in mortality following surgery for acute diverticulitis from 5.7 % to 4.3 % across the same period [42]. A study on NIS data from 2002–2007 reported a 55 % relative reduction in inhospital mortality from 4.5 % to 2.5 % [44]. A further study of the NIS data from 1999–2003 reported a 2.9 % (1330/45476) mortality following surgery for acute

diverticulitis [53]. These studies used only data coding inhospital mortality and one study failed to define the mortality rate used [53]. The inhospital mortality associated with an admission for acute diverticulitis is relatively low with a significant increase if emergency surgery is required. There are no population-based estimates of the excess mortality compared to the general population associated with acute diverticulitis, and no estimates including community-treated cases.

Risk factors

An increased BMI of greater than 30 kg m^{-2} resulted in a relative risk of 1.78 (95 % CI 1.08–2.94) compared to normal for diverticulitis in a cohort study of male American healthcare professionals [31]. There was also an independent association for waist circumference and waist-to-hip ratio. This relationship has been suggested in other smaller studies [54] and the mechanism responsible for this are unknown but may include the pro-inflammatory nature of adipose tissue which secretes cytokines which may promote the inflammatory response. Physical activity was reported to decrease the risk of developing diverticulitis in the same cohort by 25 % (RR 0.75; 95 % CI 0.58–0.95) when comparing those in the highest quintile of physical activity to those in the lowest [55]. The consumption of nuts, popcorn, and corn had been contraindicated for patients with known diverticular disease [56]. A recent report from the Health Professionals Follow-up Study found no relationship between consumption of corn, nuts, or popcorn with the development of diverticulitis [57]. The diagnoses of diverticulitis in these cohort studies included patients with complicated disease as well as simple acute diverticulitis. An increased risk of acute diverticulitis has been reported from a Danish cohort of inpatient alcoholic patients who were followed up with record linkage from 1977 to 1993 [58]. The study reported a two- and 2.9-fold increase in relative risk for men and women respectively following diagnosis of alcoholism. The authors speculated that the association could relate to the immunosuppressive effects of alcohol consumption resulting in an increased likelihood of acute diverticulitis. However, this association could also be due to an ascertainment bias due to increased investigation in alcoholics requiring hospital admission compared to the general population.

Complicated diverticular disease

Occurrence

There are few published studies detailing the occurrence of complicated diverticular disease. A national audit of 196 cases of complicated diverticular disease reported perforation as the most common complication (32 %) and fistula (14 %) as the least common, excluding cases of a mass in the left iliac fossa [59]. In a 5-year audit of cases admitted to a single center, bleeding (39 %) was the most common complication with fistula the least common (1 %) [60]. A further study from the Mayo Clinic of patients requiring surgery for complicated diverticular disease identified perforation (18 %) as the most common cause for surgery with bleeding being the least common (5 %) [61]. Clearly reporting bias in these small samples will alter the true occurrence of the disease but even from these limited studies perforated diverticular disease appears to be a common and severe presentation of complicated diverticular disease. The majority of studies therefore have focused on the incidence and prevalence of perforated diverticular disease. A study of 133 patients admitted to a hospital in northern Finland reported an increase in the prevalence of perforated diverticular disease from 2.4 per 100,000 in 1986 to 3.8 per 100,000 in 2000 [62]. A small study of 58 cases of perforated diverticular disease from Norwich (UK) reported an incidence of perforated diverticular disease of 4.0 per 100,000 per year [63]. A subsequent follow-up study from the same group identified 202 patients with perforated diverticular disease and reported an incidence of 3.5 per 100, 000 per year [64]. They reported a standardized female-to-male ratio of 1.3 (95 % CI 1.1–1.5) [64]. These studies were small and based on identification of cases from hospital records retrospectively. A subsequent population-based study using primary care data from the UK General Practice research database (GPRD) (see Chapter 9 for explanation of data source) reported an incidence of 2.66 per 100,000 person-years [65]. The incidence was higher in females (3.1 per 100,000 person-years) than males (2.1 per 100,000 person-years) and increased with increasing

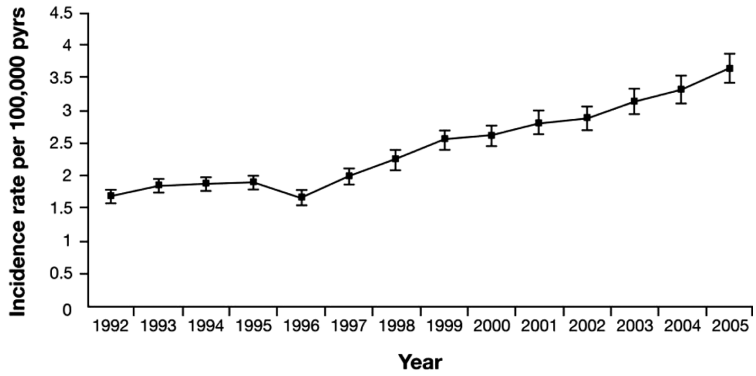


Figure 22.1 Increasing incidence of perforated diverticular disease. Source: Humes et al. 2008 [65]. Reproduced with permission of Elsevier.

age with the highest incidence in those aged 75–84 years (12.19 per 100,000 person-years). There was a twofold increase in the incidence of perforation over the time period of the study from 1990 to 2005 (Figure 22.1) [65]. This study identified cases from primary care and as such it is likely to underestimate the true incidence associated with the condition. The only population-based study of the incidence of hospital admission for lower gastrointestinal bleeding was a small study of 219 patients which reported an incidence of 22.0 per 100,000 population, with 41.6 % of the cases due to diverticular disease, an increasing incidence with increasing age and a male predominance [66].

Mortality

The mortality of complicated diverticular disease has been reported in cases series, national audits, and from population data. A study of all admissions to a single center over a year of complicated diverticular disease reported an inhospital mortality following surgery for perforated diverticular disease of 20 % (10/51) [60]. A National Audit in the United Kingdom reported a mortality of 11.3 % (34/300) overall with the greatest mortality, 35 % (22/63), reported for those with peritonitis secondary to perforation [67]. These studies were small, retrospective and likely contained reporting bias. Morris et al. reported an overall inhospital mortality of 24.3 % (49/202) from perforated diverticular disease following admission to five hospitals in East Anglia [64]. Mortality for complicated diverticular disease was reported as higher in Black race and uninsured patients when adjusting for age, sex, teaching status of hospital, Charlson score, and colostomy (OR 1.45; 95 % CI 1.08–1.94) using data from the NIS [53]. Rates of mortality following admission with lower gastrointestinal bleeding have reported rates of between 0 % and 25 % [68]. The mortality associated with diverticular bleeding has been estimated at 1.4–8.8 % [69–74]. The mortality associated with bleeding in these series has been largely due to other factors such as surgery or comorbidity. A population-based study using the UK GPRD with 953 incident cases of perforated diverticular disease reported a twofold increase (hazard ratio (HR) 2.2; 95 % CI 1.95–2.50) excess mortality compared to the general population with a sixfold excess in the first year following diagnosis (HR 5.63; 95 % CI 4.68–6.77). The majority of this excess mortality occurred in the first 3 months following diagnosis with a 3-month mortality of 13.7 % [14]. The absolute rates of mortality (263.1 per 1000 in Charlson group 2> vs. 176.6 per 1000 in Charlson group 0) were highest in those with most comorbidity but the relative risk of death was greatest in those with no recorded comorbidity (HR 8.27 vs. 3.25 Charlson group 0 compared to Charlson group >2) independent of age and sex. A 2.4-fold increase (HR 2.41; 95 % CI 1.86–3.11) in 1-year mortality was reported following a diagnosis of diverticular stricture compared to the general population. A 2.6-fold increase (HR 2.61; 95 % CI 1.47–4.62) was reported for those with a diverticular fistula [75]. The 1-year mortality following a diagnosis of fistula was 8.6 % and for stricture 10.3 %. These studies were large and population-based and accounted for other confounding factors such as comorbidity using the Charlson index [20]. They were, however, unable to report on operative mortality as they were limited to primary care data for the majority of the cohort and some

cases from secondary care may not have been captured. There is clearly a substantial inhospital mortality associated with a diagnosis of complicated diverticular disease. This represents a significant increase in mortality compared to the general population and this excess appears to be limited to the first year following diagnosis except in the case of fistula where there is a continued increase in mortality compared to the general population (HR 1.41; 95 % CI 1.04–1.91). As expected those with the greatest comorbidity had the highest absolute rates of death highlighting the important effect of comorbid illness on mortality. As more linked population data becomes available we may be able to better understand other factors that are important in predicting mortality following a diagnosis of complicated diverticular disease.

Risk factors

Lifestyle factors (smoking/obesity/physical activity)

Smokers have an excess risk of development of complicated diverticular disease. Turunen et al. analyzed a retrospective series of patients undergoing elective surgery for complicated diverticular disease and reported an increased rate of diverticular stricture amongst smokers and an increased risk of subsequent recurrent acute diverticulitis [76]. A further small retrospective study of 80 patients compared rates of smoking among those with a diagnosis of complicated diverticular disease and those with diverticulosis [77]. The study reported an odds ratio of 2.9 (95 % CI 1.1–7.3) associated with current smoking and a diagnosis of complicated diverticular disease. A subsequent study in Sweden of a cohort of 35,809 women from the Swedish Mammography Cohort reported a near twofold increase (RR 1.89; 95 % CI 1.15–3.10) in the risk of diverticular perforation or abscess formation for smokers [29]. In a population-based study of patients with diverticular perforation/abscess, fistula, and stricture identified from linked primary and secondary care data in the United Kingdom, patients were more likely to be smokers than the general population (41.5 % vs. 29.7 %, X^2 $P < 0.001$) [75]. It does appear that smoking represents a risk factor for the development of complicated diverticular disease and those patients with known disease should be advised to stop smoking.

Several lines of evidence point to an adverse effect of obesity on the development of complications of diverticular disease. A small retrospective study of just 61 patients reported a significantly lower BMI in their control population compared to those with complications of diverticular disease [78]. A much larger cohort study of 47,228 male American healthcare professionals reported a relative risk of 3.19 (95 % CI 1.45–7.00) for diverticular bleeding in those with a BMI >30 versus those with a BMI of <21 kg m^{-2} [31]. A population-based study of UK patients with complicated diverticular disease including stricture, fistula, and perforation/abscess reported a greater proportion of patients with a BMI greater than 30 kg m^{-2} compared to the general population (16.9 % vs. 12.1 %) [75]. Physical activity has been shown to confer a 46 % (RR 0.54; 95 % CI 0.38–0.77) risk reduction in the risk of diverticular bleeding when comparing those in the highest quintile for physical activity to those in the lowest [55]. As already detailed there are biologically plausible explanations for the association between complicated diverticular disease and obesity. Weight reduction with an associated increase in physical activity in those with known diverticular disease may be of benefit in preventing future complications as well as conferring other health benefits.

Prescription medications

Drugs such as opiate analgesics, corticosteroids, and nonsteroidal anti-inflammatory drugs (NSAIDS) have been suggested as potential risk factors for the development of perforated diverticular disease [79–82] with aspirin cited as a potential risk factor for both diverticulitis and diverticular bleeding [83], while calcium antagonists have been reported to have a potentially beneficial effect [84]. There are biologically plausible explanations for why these associations may occur [82,83]. The largest study detailing the excess risk associated with these medications and perforation found a twofold (OR 2.16; 95 % CI 1.55–3.01) increase in risk associated with current use of opiate analgesics and a near threefold (95 % CI 1.63–4.61) increase risk in those currently using corticosteroids, but found no association with current use of an NSAID which was in contrast to prior studies [82]. The authors suggested a possible reason for this was failure to capture over-the-counter medication use or previous studies being affected by recall bias. The largest study detailing the excess risk in bleeding

reported an increased risk of bleeding associated with both regular use of aspirin (HR 1.74; 95 % CI 1.21–2.39) and NSAIDs (HR 1.74; 95 % CI 1.15–2.64) [83]. While these associations have been reported, the absolute risk is very small with, for example, 5 bleeds per year per 100,000 subjects on aspirin compared to 3.6 per 100,000 in those not on aspirin, therefore advising patients with known disease to avoid them should be balanced against their possible benefit, which may outweigh the small risk.

Acute diverticulitis

Prior episodes of acute diverticulitis have long been thought to predispose to the development of complicated diverticular disease and therefore guidance stated that prophylactic resection following two episodes of acute diverticulitis was indicated [85]. Subsequent studies have found, however, that recurrence of acute diverticulitis is low [52], the majority of patients presenting with complicated diverticular disease have no prior history of acute diverticulitis [75,86] and the outcomes for those with two prior episodes of acute diverticulitis are no worse than those with fewer episodes [87]. Therefore it has been suggested that the indications for elective surgery should be for those patients with complications of diverticular disease and the number of prior episodes of acute diverticulitis should not be the only determining factor, with patients being assessed on an individual basis [88–90]. This stance has been reflected in the current guidelines from the American Society of Colon and Rectal Surgeons [41].

Disability and quality of life

There are no data on the quality of life of patients with diverticular disease and its associated complications. A cohort study of patients with a barium enema proven diagnosis of diverticulosis who self-reported short-lived abdominal pain showed that those with a raised score on the Hospital Anxiety and Depression Scale [91] were twice as likely to report short-lived recurrent abdominal pain [92]. A further study of somatization and healthcare use in patients with irritable bowel syndrome, healthy volunteers, and diverticular disease found that those reporting short-lived abdominal pain consistent with diverticular disease had elevated scores on the Patient Health Questionnaire 12 [93,94]. These patients were also more likely to visit their general practitioner. Those who reported prolonged abdominal pain with a fever and requiring antibiotic use had higher levels of somatization and were likely to visit their general practitioner. There is evidence therefore that with escalating symptoms there is an increase in both healthcare utilization and associated psychological morbidity. These studies, however, were small and relied on self-reported questionnaires. There is a clear need for larger studies to address the deficiencies in knowledge regarding the disability and quality of life changes that result from a diagnosis of diverticular disease or one of its complications.

Healthcare costs

There is evidence of an increasing burden of disease associated with diverticular disease in terms of hospital admissions and increasing incidence of complications [14,18,19]. There is little published data on healthcare-related cost associated with diverticular disease from Europe. Sandler et al. detailed direct and indirect healthcare costs associated with 17 selected gastrointestinal diseases using data from publicly available and proprietary national databases [21]. They estimated the total direct and indirect costs associated with diverticular disease to be US$2.5 billion. This comprised mainly direct costs at US$2.4 billion, which represented the fifth largest expenditure on gastrointestinal disease for the year 1998. They were unable to report quality of life based assessments of costs. Given the increasing incidence of complications and rates of hospitalization associated with an aging population the costs associated with diverticular disease can only be anticipated to rise [1].

Multiple choice questions

1 Which one of the following statements regarding diverticulosis is true:

A Diverticulosis becomes less common with age

B The incidence of diverticulosis is 148 per 100,000 person years

C Diverticulosis is more common in the left colon in those of Asian descent

D Up to a third of patients with diverticulosis may develop symptoms

E Diverticulosis is more common in those with a high-fiber diet

2 Which one of the following statements is true regarding acute diverticulitis?

A A BMI of greater than 30 kg m^{-2} is associated with an increased risk of developing acute diverticulitis

B The mortality associated with hospital admission for acute diverticulitis is increasing

C There is no association between eating nuts, popcorn, or corn and an increased risk of acute diverticulitis

D Hospital admissions with acute diverticulitis are decreasing

E There are no associated geographical differences in the occurrence of acute diverticulitis

3 Which one of the following statements is true regarding complicated diverticular disease?

A Prior episodes of acute diverticulitis are associated with an increased risk of perforated diverticular disease

B Physical activity is associated with an increased risk of development of complicated diverticular disease

C The incidence of perforated diverticular disease is stable

D The mortality associated with complicated diverticular disease is low

E Corticosteroids are associated with an increase risk of perforated diverticular disease

References

1 Hall KE, Proctor DD, Fisher L, Rose S. American Gastroenterological Association Future Trends Committee Report: Effects of Aging of the Population on Gastroenterology Practice, Education, and Research. *Gastroenterology* 2005;129(4):1305–38.

2 Parks TG. Natural history of diverticular disease of the colon. *Clin Gastroenterol* 1975;4(1):53–69.

3 Eide TJ, Stalsberg H. Diverticular disease of the large intestine in Northern Norway. *Gut* 1979;20(7):609–15.

4 Lee YS. Diverticular disease of the large bowel in Singapore. An autopsy survey. *Dis Colon Rectum* 1986;29(5):330–5.

5 Munakata A, Nakaji S, Takami H, et al. Epidemiological evaluation of colonic diverticulosis and dietary fiber in Japan. *Tohoku J Exp Med* 1993;171(2):145–51.

6 Nakaji S, Danjo K, Munakata A, et al. Comparison of etiology of right-sided diverticula in Japan with that of left-sided diverticula in the West. *Int J Colorectal Dis* 2002;17(6):365–73.

7 Humes D, Smith JK, Spiller RC. Colonic diverticular disease. *Clin Evid (Online)* 2011; doi:pii:0405.

8 Jun S, Stollman N. Epidemiology of diverticular disease. *Best Prac Res Clin Gastroenterol* 2002;16(4):529–42.

9 Gear JSS, Fursdon P, Nolan DJ, et al. Symptomless diverticular disease and intake of dietary fibre. *Lancet* 1979;313(8115):511–14.

10 Manousos ON, Truelove SC, Lumsden K. Prevalence of colonic diverticulosis in general population of Oxford area. *BMJ* 1967;3(5568):762–3.

11 Parks TG. Post-mortem studies on the colon with special reference to diverticular disease. *Proc R Soc Med* 1968;61(9):932–4.

12 Eastwood MA, Sanderson J, Pocock SJ, Mitchell WD. Variation in the incidence of diverticular disease within the city of Edinburgh. *Gut* 1977;18(7):571–4.

13 Crowe FL, Appleby PN, Allen NE, Key TJ. Diet and risk of diverticular disease in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Oxford cohort: a prospective study of British Vegetarians and non-vegetarians. *BMJ* 2011.

14 Humes DJ, Solaymani-Dodaran M, Fleming KM, et al. A Population-based study of perforated diverticular disease incidence and associated mortality. *Gastroenterology* 2009;136(4):1198–205.

15 Kang JY, Firwana B, Green AE, et al. Uncomplicated diverticular disease is not a common cause of colonic symptoms. *Aliment Pharmacol Ther* 2011;33(4):487–94.

16 Simpson J, Neal KR, Scholefield JH, Spiller RC. Patterns of pain in diverticular disease and the influence of acute diverticulitis. *Eur J Gastroenterol Hepatol* 2003;15(9):1–6.

17 Simpson J, Scholefield JH, Spiller RC. Origin of symptoms in diverticular disease. *Br J Surg* 2003;90(8):899–908.

18 Kang JY, Hoare J, Tinto A, et al. Diverticular disease of the colon on the rise: a study of hospital admissions in England between 1989/1990 and 1999/2000. *Aliment Pharmacol Ther* 2003;17(9):1189–95.

19 Jeyarajah S, Faiz O, Bottle A, et al. Diverticular disease hospital admissions are increasing, with poor outcomes in the elderly and emergency admissions. *Aliment Pharmacol Ther* 2009;30(11–12):1171–82.

20 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in

longitudinal studies: development and validation. *J Chronic Dis* 1987;40(5):373–83.
21 Sandler RS, Everhart JE, Donowitz M, et al. The burden of selected digestive diseases in the United States. *Gastroenterology* 2002;122(5):1500–11.
22 Painter NS, Burkitt DP. Diverticular disease of the colon: a deficiency disease of Western civilization. *BMJ* 1971;2(5759):450–4.
23 Brodribb AJ, Humphreys DM. Diverticular disease: three studies. Part I: Relation to other disorders and fibre intake. *BMJ* 1976;1(6007):424–5.
24 Manousos O, Day NE, Tzonou A, et al. Diet and other factors in the aetiology of diverticulosis: an epidemiological study in Greece. *Gut* 1985;26(6):544–9.
25 Aldoori WH, Giovannucci EL, Rimm EB, et al. A prospective study of diet and the risk of symptomatic diverticular disease in men. *Am J Clin Nutr* 1994;60(5):757–64.
26 Commane DM, Arasaradnam RP, Mills S, et al. Diet, ageing and genetic factors in the pathogenesis of diverticular disease. *World J Gastroenterol* 2009;15(20):2479–88.
27 Hart AR, Kennedy HJ, Day NE. Beyond Burkitt – is diverticular disease more than just cereal fibre deficiency? *Postgrad Med J* 2000;76(895):257–58.
28 Aldoori WH, Giovannucci EL, Rimm EB, et al. A prospective study of alcohol, smoking, caffeine, and the risk of symptomatic diverticular disease in men. *Ann Epidemiol* 1995;5(3):221–28.
29 Hjern F, Wolk A, Håkansson N. Smoking and the risk of diverticular disease in women. *Br J Surg* 2011;98(7):997–1002.
30 Aldoori WH, Giovannucci EL, Rimm EB, et al. Prospective study of physical activity and the risk of symptomatic diverticular disease in men. *Gut* 1995;36(2):276–82.
31 Strate L, Liu Y, Aldoori W, et al. Obesity increases the risks of diverticulitis and diverticular bleeding. *Gastroenterology* 2009;136(1):115–22.e1.
32 Rosemar A, Angerås U, Rosengren A. Body mass index and diverticular disease: a 28-year follow-up study in men. *Dis Colon Rectum* 2008;51(4):450–55.
33 Humes DJ, West J. Diet and risk of diverticular disease. *BMJ* 2011;343.
34 Stefansson T, Ekbom A, Sparen P, Pahlman L. Increased risk of left-sided colon cancer in patients with diverticular disease. *Gut* 1993;34(4):499–502.
35 Stefánsson T, Ekbom A, Sparèn P, Påhlman L. Association between sigmoid diverticulitis and left-sided colon cancer: a nested, population-based, case control study. *Scand J Gastroenterol* 2004;39(8):743–47.
36 Lee K-M, Paik C-N, Chung WC, et al. Clinical significance of colonic diverticulosis associated with bowel symptoms and colon polyp. *J Korean Med Sci* 2010;25(9):1323–29.
37 Stefansson T, Ekbom A, Sparen P, Pahlman L. Cancers among patients diagnosed as having diverticular disease of the colon. *Eur J Surg* 1995;161(10):755–60.
38 Lam TJ, Meurs-Szojda MM, Gundlach L, et al. There is no increased risk for colorectal cancer and adenomas in patients with diverticulitis: a retrospective longitudinal study. *Colorectal Dis* 2010;12(11):1122–26.
39 Meurs-Szojda M, Droste J, Kuik D, et al. Diverticulosis and diverticulitis form no risk for polyps and colorectal neoplasia in 4,241 colonoscopies. *Int J Colorectal Dis* 2008;23(10):979–84.
40 Granlund J, Svensson T, Granath F, et al. Diverticular disease and the risk of colon cancer – a population-based case-control study. *Aliment Pharmacol Ther* 2011;34:675–81.
41 Rafferty J, Shellito P, Hyman N, Buie W. Practice parameters for sigmoid diverticulitis. *Dis Colon Rectum* 2006;49(7):939–44.
42 Etzioni DA, Mack TM, Beart RWJ, Kaiser AM. Diverticulitis in the United States: 1998–2005: Changing patterns of disease and treatment. *Ann Surg* 2009;249(2):210–17.
43 Nguyen GC, Sam J, Anand N. Epidemiological trends and geographic variation in hospital admissions for diverticulitis in the United States. *World J Gastroenterol* 2011;17(12):1600–5.
44 Masoomi H, Buchberg BS, Magno C, et al. Trends in diverticulitis management in the United States from 2002 to 2007. *Arch Surg* 2011;146(4):400–06.
45 Etzioni DA, Cannom RR, Ault GT, et al. Diverticulitis in California from 1995 to 2006: Increased rates of treatment for younger patients. *The American Surgeon* 2009;75:981–85.
46 Jacobs DO. Diverticulitis. *New Engl J Med* 2007;357(20):2057–66.
47 Parks TG. Natural history of diverticular disease of the colon. A review of 521 cases. *BMJ* 1969;4(684):639–42.
48 Hall JF, Roberts PL, Ricciardi R, et al. Long-term follow-up after an initial episode of diverticulitis: What are the predictors of recurrence? *Dis Colon Rectum* 2011;doi:10.1007/DCR.0b013e3182028576.
49 Hjern F, Josephson T, Altman D, et al. Outcome of younger patients with acute diverticulitis. *Br J Surgery* 2008;95(6):758–64.
50 Eglinton T, Nguyen T, Raniga S, et al. Patterns of recurrence in patients with acute diverticulitis. *Br J Surg* 2010;97(6):952–57.
51 Makela JT, Kiviniemi HO, Laitinen ST. Spectrum of disease and outcome among patients with acute diverticulitis. *Dig Surg* 2010;27(3):190–6. Epub 2010 Jun 22.
52 Broderick-Villa G, Burchette RJ, Collins JC, et al. Hospitalization for acute diverticulitis does not mandate routine elective colectomy. *Arch Surg* 2005;140(6):576–81; discussion 81–3.

53 Lidor AO, Gearhart SL, Wu AW, Chang DC. Effect of race and insurance status on presentation, treatment, and mortality in patients undergoing surgery for diverticulitis. *Arch Surg* 2008;143(12):1160–65.
54 Schauer PR, Ramos R, Ghiatas AA, Sirinek KR. Virulent diverticular disease in young obese men. *Am J Surg* 1992;164(5):443–6; discussion 46–8.
55 Strate LL, Liu YL, Aldoori WH, Giovannucci EL. Physical activity decreases diverticular complications. *Am J Gastroenterol* 2009;104(5):1221–30.
56 Schechter S, Mulvey J, Eisenstat TE. Management of uncomplicated acute diverticulitis: results of a survey. *Dis Colon Rectum* 1999;42(4):470–5; discussion 75–6.
57 Strate LL, Liu YL, Syngal S, et al. Nut, corn, and popcorn consumption and the incidence of diverticular disease. *JAMA* 2008;300(8):907–14.
58 Tønnesen H, Engholm G, Møller H. Association between alcoholism and diverticulitis. *Br J Surg* 1999;86(8):1067–68.
59 Tudor R, Farmakis N, Keighley M. National audit of complicated diverticular disease: analysis of index cases. *Br J Surg* 1994;81(5):730–2.
60 Elliott T, Yego S, Irvin T. Five-year audit of the acute complications of diverticular disease. *Br J Surg* 1997;84(4):535–9.
61 McConnell EJ, Tessier DJ, Wolff BG. Population-based incidence of complicated diverticular disease of the sigmoid colon based on gender and age. *Dis Colon Rectum* 2003;46(8):1110–4.
62 Makela J, Kiviniemi H, Laitinen S. Prevalence of perforated sigmoid diverticulitis is increasing. *Dis Colon Rectum* 2002;45(7):955–61.
63 Hart AR, Kennedy HJ, Stebbings WS, Day NE. How frequently do large bowel diverticula perforate? An incidence and cross-sectional study. *Eur J Gastroenterol Hepatol* 2000;12(6):661–5.
64 Morris CR, Harvey IM, Stebbings WSL, Hart AR. Incidence of perforated diverticulitis and risk factors for death in a UK population. *Br J Surg* 2008;95(7):876–81.
65 Humes DJ, Solaymani-Dodaran M, Simpson J, et al. Incidence of perforated diverticular disease and its associated mortality: a UK general population-based study. *Gastroenterology* 2008;134(4).
66 Longstreth GF. Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: a population-based study. *Am J Gastroenterol* 1997;92(3):419–24.
67 Tudor RG, Farmakis N, Keighley MR. National audit of complicated diverticular disease: analysis of index cases. *Br J Surg* 1994;81(5):730–2.
68 Strate LL. Lower GI Bleeding: Epidemiology and diagnosis. *Gastroenterol Clin North Am* 2005;34(4):643–64.
69 Thomas S, Wong RCK, Das A. Economic burden of acute diverticular haemorrhage in the U.S.: A nationwide estimate. *Gastroenterology* 2004;126(4 Suppl 2):A606–7.
70 McGuire HH. Bleeding colonic diverticula. A reappraisal of natural history and management. *Ann Surg* 1994;220(5):653–6.
71 McGuire HH, Haynes BW. Massive hemorrhage for diverticulosis of the colon: guidelines for therapy based on bleeding patterns observed in fifty cases. *Ann Surg* 1972;175(6):847–55.
72 Poncet G, Heluwaert F, Voirin D, et al. Natural history of acute colonic diverticular bleeding: a prospective study in 133 consecutive patients. *Aliment Pharmacol Ther* 2010;32(3):466–71.
73 Schmulewitz N, Fisher DA, Rockey DC. Early colonoscopy for acute lower GI bleeding predicts shorter hospital stay: a retrospective study of experience in a single center. *Gastrointestinal Endoscopy* 2003;58(6):841–46.
74 Chen C-Y, Wu C-C, Jao S-W, et al. Colonic diverticular bleeding with comorbid diseases may need elective colectomy. *J Gastrointest Surg* 2009;13(3):516–20.
75 Humes DJ, West J. The role of acute diverticulitis in the development of complicated colonic diverticular disease and one year mortality following diagnosis in the UK: population based cohort study. *Gut* 2011; May 6. [Epub ahead of print.]
76 Turunen P, Wikstrom H, Carpelan-Holmstrom M, et al. Smoking increases the incidence of complicated diverticular disease of the sigmoid colon. *Scand J Surg* 2010;99(1):14–7.
77 Papagrigoriadis S, Macey L, Bourantas N, Rennie JA. Smoking may be associated with complications in diverticular disease. *Br J Surg* 1999;86(7):923–26.
78 Dobbins C, Defontgalland D, Duthie G, Wattchow D. The relationship of obesity to the complications of diverticular disease. *Colorectal Dis* 2006;8(1):37–40.
79 Morris CR, Harvey IM, Stebbings WSL, et al. Anti-inflammatory drugs, analgesics and the risk of perforated colonic diverticular disease. *Br J Surg* 2003;90(10):1267–72.
80 Wilson RG, Smith AN, Macintyre IM. Non-steroidal anti-inflammatory drugs and complicated diverticular disease: a case-control study. *Br J Surg* 1991;78(9):1148.
81 Campbell K, Steele RJ. Non-steroidal anti-inflammatory drugs and complicated diverticular disease: a case-control study. *Br J Surg* 1991;78(2):190–1.
82 Humes DJ, Fleming KM, Spiller RC, West J. Concurrent medication use and the risk of perforated colonic diverticular disease: a population-based case control study. *Gut* 2011;60(2):219–24.
83 Strate LL, Liu YL, Huang ES, et al. Use of aspirin or nonsteroidal anti-inflammatory drugs increases risk for

diverticulitis and diverticular bleeding. *Gastroenterology* 2011;140(5):1427–33.
84 Morris CR, Harvey IM, Stebbings WSL, et al. Do calcium channel blockers and antimuscarinics protect against perforated colonic diverticular disease? A case control study. *Gut* 2003;52(12):1734–37.
85 Wong W, Wexner S, Lowry A, et al. Practice parameters for the treatment of sigmoid diverticulitis – supporting documentation. The Standards Task Force. The American Society of Colon and Rectal Surgeons. *Dis Colon Rectum* 2000;43(3):290–7.
86 Chapman J, Davies M, Wolff B, et al. Complicated diverticulitis: is it time to rethink the rules? *Ann Surg* 2005;242(4):576–81; discussion 81–3.
87 Chapman JR, Dozois EJ, Wolff BG, et al. Diverticulitis: a progressive disease? Do multiple recurrences predict less favorable outcomes? *Ann Surg* 2006;243(6):876–30; discussion 80–3.
88 Salem L, Veenstra DL, Sullivan SD, Flum DR. The timing of elective colectomy in diverticulitis: A decision analysis. *J Am Coll Surg* 2004;199(6):904–12.
89 Janes S, Meagher A, Frizelle FA. Elective surgery after acute diverticulitis. *Br J Surg* 2005;92(2):133–42.
90 Klarenbeek BR, Samuels M, van der Wal MA, et al. Indications for elective sigmoid resection in diverticular disease. *Ann Surg* 2010; doi:10.1097/SLA.0b013e3181d447d.
91 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67(6):361–70.
92 Humes DJ, Simpson J, Neal KR, et al. Psychological and colonic factors in painful diverticulosis. *Br J Surg* 2008;95(2):195–98.
93 Spiller RC, Humes DJ, Campbell E, et al. The Patient Health Questionnaire 12 Somatic Symptom scale as a predictor of symptom severity and consulting behaviour in patients with irritable bowel syndrome and symptomatic diverticular disease. *Aliment Pharmacol Ther* 2010;32(6):811–20.
94 Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med* 2002;64(2):258–66.

Answers to multiple choice questions

1. D
2. A
3. E

23 Epidemiology of infectious diarrhea

Christina M. Surawicz[1] *& Crenguta Stepan*[2]
[1]University of Washington School of Medicine, Seattle, WA, USA
[2]University of Washington, Valley Medical Center, Seattle, WA, USA

Key points

- Infectious diarrhea is a major health problem in both the developing and developed world.
- Each year 1.5 million children die worldwide as a result of diarrhea, much of which is infectious. Diarrheal diseases mainly affect children under two years old.
- Each year, 1 in 6 (or 48 million) Americans gets a food-borne illness.
- Prevention through sanitation, education, vaccination, and oral rehydration are high priorities.

Introduction

From sporadic cases to familial outbreaks or epidemics, diarrheal disease is a common public health problem worldwide and is one of the leading causes of morbidity and mortality among infants and children in developing countries. Data from the World Health Organization from 2009 indicate that there are two billion cases of diarrhea every year and that 1.5 million children die every year as a result of diarrheal disease [1]. In the United States (USA), there are 76 million cases of diarrhea per year, with 325,000 hospitalizations and 5000 deaths, according to data published in 1999 [2]. Many cases of infectious diarrhea are food-borne. The Produce Safety Project report, "Health-related costs from food-borne illness in the United States" showed that these illnesses cost the USA $152 billion per year in health-related costs, with an average cost of $1850 per illness [3]. Acute diarrhea is defined as three or more loose or watery stools per day lasting for up to 14 days or less; persistent diarrhea lasts for 2–4 weeks, and chronic diarrhea lasts longer than a month. Chronic diarrhea does not commonly have an infectious etiology. The most common pathogens causing acute diarrhea are shown in Table 23.1. Most acute infectious diarrhea is due to viruses such as noroviruses (formerly called the Norwalk agent), rotaviruses, astroviruses, and enteric adenoviruses. Typical symptoms of viral gastroenteritis are watery stools, crampy abdominal pain, nausea, vomiting, myalgia, fatigue, dehydration, and low-grade fever. Bacteria such as *Salmonella*, *Shigella*, *Campylobacter*, *Yersinia*, *Aeromonas*, and *Plesiomona* can cause watery diarrhea, but they are invasive bacteria and can cause bloody diarrhea, as well as crampy abdominal pain, fever and signs of systemic illness when severe. Protozoa that affect the small bowel such as *Giardia*, *Cryptosporidium*, and *Cyclospora* typically cause watery or malabsorptive diarrhea. Any enteric infection can have a prolonged course in immunosuppressed patients. Most diarrhea is self-limited, so diagnostic tests and treatment are not necessary. In some cases (e.g. extremes of age, severe diarrhea, bloody diarrhea, fever, or immunocompromise) stool culture may be indicated as the

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 23.1 Common causes of infectious diarrhea

Etiology: pathogen	Frequency by stool culture (in USA)
Norovirus	8500 cases per 100,000 population
Rotavirus	1100 cases per 100,000 population
Giardia	750 cases per 100,000 population
Salmonella	14.7 cases per 100,000 population
Campylobacter	12.9 cases per 100,000 population
Shigella	5.1 cases per 100,000 population
E. coli 0157:H7 (Shiga toxin *E. coli*)	0.9 cases per 100,000 population

Source: Adapted from Mead et al. 1999 [2].

likelihood of a pathogen being detected is higher than in other cases of mild diarrhea. Complications of enteric infection can include hemolytic uremic syndrome (HUS), usually associated with infection with *Shigella* or Shiga toxin-producing *E. coli* (STEC), Guillain–Barré syndrome (associated with *Campylobacter* infection) and a post-infectious arthritis syndrome which can be associated with *Campylobacter*, *Salmonella*, *Shigella*, or *Yersinia*. Other complications include post-infectious irritable bowel syndrome, and persistent or chronic diarrhea.

Incidence

Past data estimate that 211–270 million episodes of diarrhea occur annually in the USA, with about 0.72 episodes per adult-year [3]. Of these, 76 million cases of food-borne disease occur each year, 82 % with an unknown etiology, and result in 3000 and 5000 deaths per year [4]. The incidence of various pathogens varies with reports to the Centers for Disease Control, reports to health departments, and how often stools are tested. In 2004, the incidence of diarrheal pathogens calculated using the laboratory-confirmed infections is shown in Table 23.1 [5]. For some common pathogens there are no routine diagnostic tests.

In children, the incidence of parent-defined diarrhea is 2.2 episodes per person-year. Rotavirus is the leading cause of hospitalization for diarrhea among children worldwide. The median detection rate for rotavirus among children hospitalized with diarrhea ranges from 29 % to 45 %, with a peak during the winter time [6]. A retrospective study of 402 children hospitalized for acute diarrhea documented that rotavirus was the most common cause (77 %), followed by *Salmonella* (23 %) [7]. In the USA, following the introduction of rotavirus vaccination in 2006, there has been a significant decrease in all-cause diarrhea-related hospitalizations as well as documented decreases in rotavirus diarrhea cases in children [8]. Reductions have ranged from 29 % to 50 % in various studies [9,10]. Persistent diarrhea accounts for 8 % of diarrheal illnesses during childhood and can be associated with *Cryptosporidium*, *Giardia*, enteric adenoviruses, and enterotoxigenic *E. coli*.

Developing countries

In developing countries, there is a median of 3.2 episodes of diarrhea per child-year in children under 5 years of age [11]. The etiology is mostly viral but the incidence of bacterial diarrhea is greater than in developed countries. An interesting difference has been noticed in the epidemiology of *Campylobacter*. In developing countries, *Campylobacter* infections are endemic among young children, asymptomatic infections are common and outbreaks are rare. In developed countries, asymptomatic *Campylobacter* infections are unusual and outbreaks are common [12]. However, worldwide, *Campylobacter* remains one of the most common bacterial causes of diarrhea.

Risk factors

For any case of diarrhea, the epidemiologic history may reveal important clues about the etiology. Specific pathogens can be associated with specific risk factors, as shown in Table 23.2. Epidemiologic surveys have revealed several factors that may influence the risk and incidence of infectious diarrhea. For some of these factors, data remain very limited.

Age

Infants, toddlers and young adults (15–35 years old) are prone to develop traveler's diarrhea, which may be related to hygiene and dietary habits. Younger age is also a risk factor for rotavirus diarrhea and rotavirus-related hospitalization; its incidence decreases with

Table 23.2 Common risk factors for specific pathogens causing infectious diarrhea

Pathogen	Risk factors
E. coli 0157: H7	Undercooked beef, unpasteurized milk or apple cider, visits to animal farms, petting zoos
Shigella	Contaminated water and vegetables, day-care centers, custodial institutions
Campylobacter	Undercooked poultry, contaminated milk, tuna salad, eggs
Salmonella	Raw eggs, undercooked poultry and turkey, unrefrigerated dressing, reptiles as a pet, family members with *Salmonella*
Noncholera Vibrio	Raw/undercooked seafood
Giardia lamblia	Contaminated water, recreational exposure in lakes, rivers or swimming pools, day-care centers
Cryptosporidium	International travel, contact with cattle, freshwater swimming, petting zoos, water parks
Rotavirus	Winter outbreaks, in children under 2 years
Norovirus	Winter outbreaks in adults, raw oysters consumption, cruise ships
Yersinia	Pig intestine
Aeromonas	Brackish water, fresh water

age, presumably due to developing immunity. Children under 5 years also have the highest rate of HUS.

Gender

There is no significant difference in diarrhea incidence rate according to gender; women and men have the same risk of developing infectious diarrhea or traveler's diarrhea. Women may have a higher incidence rate for HUS [24], although the incidence rates of *E. coli* O157:H7 showed no differences by gender. Women have a higher risk of developing post-infectious IBS.

Ethnicity

Because data on ethnicity are incomplete, conclusions cannot be made about differences in the epidemiology of infectious diarrhea. However, Caucasians have a higher incidence of diarrhea-related HUS and ethnicity-specific hospitalization rates for rotavirus diarrhea [13].

Geography and socioeconomic status

The country of origin and host country are important determinants for traveler's diarrhea. Coming from developed countries and traveling to developing countries is associated with highest attack rates. Although the pathogens causing diarrhea are the same worldwide, the incidence of specific pathogens varies.

Seasonality

Rotavirus and norovirus infections occur more frequently during winter months. Bacterial diarrhea is most common during the summer possibly because of traveling, swimming, and food choices. The higher rate of HUS during summer and fall reflects the exposure to *E. coli* 0157:H7 infections.

Genetics

Among blood types, certain ABO phenotypes have been reported to be associated with susceptibility to some enteric pathogens. Type O phenotype presents a greater susceptibility to *Vibrio cholerae* [14] and possibly also to norovirus [15] infection. Post-infectious arthritis as a complication of *S. flexneri* infection occurs especially in persons with the genetic predisposition human leukocyte antigen (HLA) B27. There is also evidence of genetic factors associated with EAEC diarrhea and increased level of fecal interleukin 8 (IL-8) [16]. Other studies have documented a genetic predisposition to post-infectious IBS, identifying genes in the toll-like receptor 9, IL-6, and cadherin 1 regions [17].

Environmental

Behavioral risk factors associated with diarrhea include dietary habits (consumption of "risky" food or beverages, unpasteurized milk), or environmental

exposure (animal exposure, poor hygiene); these are important determinants for traveler's diarrhea. Trends that lead to increase in food-borne infection include a global market for foods, changing eating habits, changes in diet, farming practices, climate change, and an older population. Major issues in the food chain need to be addressed to decrease food-borne infections [18]. Globalization of the US food supply is an additional factor as over 15 % of the food consumed in the USA is imported. In the USA, the Food-borne Diseases Active Surveillance Network (FoodNet) reporting system showed that the leading causes of laboratory-confirmed food-borne infections were *Salmonella*, *Campylobacter*, *Shigella*, *Cryptosporidium*, and Shiga toxin *E. coli* (STEC) 0157:H7 [19,20].

The Geosentinel Surveillance Network comprises 42 specialized travel medicine sites on all six continents. Data collected in this database provides interesting epidemiologic information, as the spectrum of patients includes self-referred outpatients as well as hospitalized patients. An analysis of data on infectious gastrointestinal diseases (IGD) was performed between 1996 and 2005 [21]. IGD was more associated with female gender, younger age, and traveling for tourism. Interestingly, of 2902 pathogens isolated, 65 % were parasitic (commonly *Giardia*, *E. histolytica*, and *Strongyloides*), 31 % were bacterial (commonly *Campylobacter*, *Shigella*, and *Salmonella*), and 3 % were viral. It should be noted, however, that these studies did not include testing data for Enterotoxigenic *E. coli* (ETEC) or common enteric viruses.

An increase in *Salmonella* infection in the 1980s has been epidemiologically related to shell eggs and poultry. The organism has adapted to survive in hen oviduct and ovary, and infection in eggs has been linked to the use of antimicrobials in food production. Moreover, there is a marked increase in epidemics including Salmonella and STEC related to fresh produce (fruit, vegetables, meat), soil, water, sewage, human handling, infected meat [18] as well as parasitism from free-living parasites from the environment (eggs, cysts).

Epidemiologic studies show that among travelers to different areas of the world, the risk of traveler's diarrhea is about 7 % in developed countries and 20–50 % in the developing world, with a total of 15–20 million cases annually [22]. For travel to high-risk areas, such as Africa (excluding South Africa), Asia, South and Central America, Mexico, and India, rates can vary from 30 % to 50 %. Intermediate-risk areas are the countries around the Mediterranean Sea (including Israel), the Caribbean countries, and South Africa with attack rates of 10–20 %. Low risk areas are southern Europe and North America where traveler's diarrhea rates are less than 8 % [23]. Enterotoxigenic *E. coli* is the main cause. A recent systematic review of 51 publications that included data on the etiology of traveler's diarrhea (from 1973 to 2008) showed that ETEC and Enteroaggregative *E. coli* were the most common pathogens. ETEC accounted for 30–33 % of traveler's diarrhea cases in Latin America, Africa, and South Asia [23]. Enteroaggerative *E. coli* was second: 24 % in Latin America; rates were lower in Africa and South Asia. *Campylobacter* infection was more common in Asia than Latin America and Africa. *Vibrio* and non-cholera Vibrio, and *Giardia* and *E. histolytica* were also more common in Asia. *Shigella* cases were more common in Africa, and *Salmonella* cases were more common causes of traveler's diarrhea cases in Asia. In many (40–50 %) of traveler's diarrhea cases, no pathogen is identified. In a study of enteric infections in British Columbia, international travel accounted for 40 %, and rates were highest in 30–39-year olds [24]. Asia and Mexico were common destinations. In most cases, diarrhea is mild with less than 6 bowel movement per day, and resolves in 1–2 days. Severe cases can occur, and require treatment with antibiotics. The natural history is that 90 % of cases resolve within 1 week, and 98 % resolve within 1 month.

Tropical sprue can affect travelers in Asia, some Caribbean islands, and parts of South America. It usually presents as persistent diarrhea. It is possible that the disease is either initiated or sustained by a still-undefined infection because it responds to antibiotics.

Medication, nosocomial infection and comorbid conditions

Acid suppression medication increases the risk of enteric infections and increasing evidence suggests that proton pump inhibitors, in particular, are associated with an increased risk of *Clostridium difficile* infection (CDI) [25–28]. Diarrhea is a common side effect of most antibiotics; antibiotic-associated diarrhea (AAD) occurs in approximately 20 % of patients who take antibiotics. It is estimated that 10–15 % of individuals who develop AAD will have CDI. Since 2000, the incidence of CDI in the USA, Canada, Europe,

and other countries has been increasing, especially in patients aged 65 years or older. Diarrhea due to *C. difficile* is more common in those who take clindamycin, the second- and third-generation cephalosporins, and more recently quinolones. Many outbreaks have been associated with a hypervirulent strain (NAP1/BI/027) that has a gene deletion that allows for increased toxin production *in vitro*. It also has quinolone resistance, and widespread use of quinolones over the past decade may have allowed it to emerge. In the USA, between 2000 and 2006, the incidence of CDI in adults increased from 1.3/10,000 to 2.4/10,000, with higher rates in older adults, up to 49/10,000, and even up to 112/10,000 in those over 85 years old [29]. Rates of nosocomial CDI have increased but also cases of community-acquired CDI, although the incidence of community-acquired *C. difficile* diarrhea appears to be substantially lower than rates observed in hospitals [30]. Recurrent CDI (RCDI) is CDI that recurs after treatment, usually within 1–2 weeks of stopping antibiotic treatment. Risk factors associated with RCDI include older age, renal insufficiency, acquisition during hospitalization, concomitant antibiotics, acid-suppressing medications, and antacids [31,32].

Rarely, other enteric infections can cause epidemics in hospitals. A Camembert cheese contaminated with *Listeria monocytogenes* caused a hospital outbreak – 17 patients were infected and 3 died [33]. Immunodeficiency conditions, inherited or acquired, increase the risk of infections. Human immunodeficiency virus (HIV) infection is frequently complicated by diarrhea. In patients with <200 CD4 cells mm^{-3}, the occurrence of chronic diarrhea ranges from 8–10 % per year and can be due to cytomegalovirus (CMV), cryptosporidiosis, microsporidia, or *Mycobacterium avium* complex (MAC). Worldwide, the most common causes of diarrhea in HIV-infected patients are enteric bacteria including *Shigella*, *Salmonella*, and *Campylobacter*. Common variable immunodeficiency (CVID) presents with frequent bacterial infections, persistent diarrhea, and malabsorption caused by *Giardia lamblia* infection.

Nutrition

The most important epidemiologic risk factor for chronic diarrhea is malnutrition. In children, other associated conditions are zinc deficiency, lack of breastfeeding, and a history of intrauterine growth retardation.

Natural history, prognosis, and mortality

The natural history and the prognosis of infectious diarrhea depend on etiology and host factors. In adult patients with viral gastroenteritis, the illness usually lasts 24–48 hours, but the shedding of the virus in stools may continue for longer than a week and symptoms can present for up to a week. Norovirus infection may lead to a longer duration of diarrhea and severe consequences in the elderly, patients with cardiovascular disease, and recipients of renal transplant or immunosuppressive therapy [34]. Asymptomatic infection is very common. The case-fatality rate varies from 0.01 % in the USA to 0.075 % in England or as high as 2 % in Israel [35]. The highest mortality rate is associated with outbreaks in hospitals and residential-care facilities so infection in hospitalized persons may be more severe than that in other groups. Rotaviruses, enteric adenoviruses, and astroviruses sometimes cause gastroenteritis in adults; however, they are less common, perhaps because protective immunity for these viruses often develops in childhood, whereas the immune response to noroviruses is generally short-lived or ineffective. Patients with bacterial infection are usually ill for longer, typically 3–5 days but can be up to two weeks. Listeriosis can be life-threatening in neonates, pregnant women, and immunocompromised patients.

Most childhood diarrhea is mild, with complete recovery. However, about 50,000 hospitalizations in the USA may be attributable to rotavirus [36]. Rotavirus gastroenteritis requiring hospitalization occurs most frequently in infants and young children, aged from 6 months to 2 years. In developed countries, rotavirus infection rarely results in death (a total of about 20–40 cases in the USA). In developing countries overall, diarrhea accounts for 21 % of all deaths of children aged under 5 years, with 2.5 million deaths per year [37]. Of these, about 500,000 children die every year from rotavirus gastroenteritis, with >80 % of these deaths occurring in these countries [38]. For any etiology, interventions that can decrease morbidity and mortality rates are breastfeeding up to 6 months, improved sanitation, and use of oral

rehydration therapy. The number of deaths of children that could be prevented worldwide each year by breastfeeding alone has been estimated to be more than 1 million [39]. Oral rehydration reduces the diarrheal mortality, especially among children under 1 year of age, in whom acute dehydration is the greatest cause of death.

Disability and quality of life

Acute diarrhea

Acute diarrhea variably affects quality of life, from the inconvenience of having symptoms to the inability to leave home and function normally. Uncontrolled diarrhea can lead to dehydration and chemical imbalances that might be life-threatening in infants and the elderly. Other post-infectious complications of diarrhea are reactive arthritis, associated with 1–4 % of the gastroenteritis caused by *Shigella*, *Salmonella*, *Campylobacter*, and *Yersinia*. Guillain–Barré syndrome occurs in 1–2 cases per 100,000 population per year but its incidence is <1 per 1000 infections with *Campylobacter* [43]. HUS is caused in almost all cases by STEC or Shigella infection in developing countries. In the USA, Shiga toxin-producing *E. coli* (STEC) is implicated in 72 % of cases of HUS, and *E. coli* 0157:07 was the pathogen in over 80 % of patients with STEC infection [40].

Chronic diarrhea

Chronic diarrhea can have a substantial impact on the quality of life and health overall. In children, morbidity is related to malnutrition, and physical and cognitive abnormalities. Diarrheal illnesses occurring during the first two years of life can have a profound impact on growth, cognitive development and educational performance [41]. In adults, the condition can be an inconvenience or it can be disabling, causing malnutrition, weight loss, and increased morbidity.

Chronic diarrhea is rarely infectious, but chronic diarrhea can be associated with intestinal parasites, or bacteria such as, *Y. enterocolitica*, *Aeromonas*, and *Plesimonas*. The populations at risk for this are travelers returning from tropical countries and immunocompromised patients, especially patients with HIV infection whose CD4 cell counts are below 200 μL^{-1}. However, the introduction of highly active antiretroviral therapy (HAART) has decreased the diarrhea caused by organisms such as microsporidia and cryptosporidia by improving the immune system. In children, persistent diarrhea accounts for 36–54 % of all diarrhea-related deaths [41].

Post-infectious IBS (PI-IBS) is a well-recognized complication of infectious gastroenteritis. Post-infectious irritable bowel syndrome (PI-IBS) is defined as a change in bowel habits or an onset of new abdominal pain or discomfort following a recent exposure to infectious organisms. PI-IBS has been reported to occur after traveler's diarrhea with ETEC and EAEC and after gastroenteritis caused by *Campylobacter*, *Shigella*, and *Salmonella*, in 8–10 % of cases [42], and even after viral gastroenteritis. The development of PI-IBS is influenced by host and microbial factors, and the duration and severity of the acute infection. Known risk factors have been described as female gender, long duration of diarrhea with inciting episode, younger age, and a history of prior anxiety or depression. Several studies have shown that many patients with PI-IBS remain symptomatic many years later. A severe epidemic of infectious gastroenteritis in Walkerton, a small town in Canada whose water supply became contaminated with bacteria has provided important epidemiologic information about PI-IBS. The Walkerton Health Study was established to enable the long-term follow-up of a cohort of hundreds of individuals who had IGE [43]. The prevalence of PI-IBS was 28.3 % at two to three years, 21.4 % at four years, and 14.3 % at six years. At six years 14.3 % had PI-IBS compared to 5.4 % IBS in controls. PI-IBS was greater in those with acute bacterial gastroenteritis compared to controls.

Healthcare utilization and costs

Diarrhea is an extremely costly disease. The yearly cost of acute diarrhea in the USA is $5–6 billion in direct medical expenses and lost productivity [44].

Indirect costs

With an estimate of 50,000 hospitalizations attributable to rotavirus each year in the USA, the average nonmedical cost per case of rotavirus disease is about $448.77. The nonmedical costs of severe

rotavirus infections may exceed $22 million annually, with 80 % of the cost attributable to missed work [45].

Medication

According to an Australian study, the average cost of prescribed medication per visit was A$6.83 and the estimated cost of over-the-counter medication was A$8.76 [46]. It has been estimated that in the USA more than 30 % of the population with infectious diarrhea receives antidiarrheal medication. There are no data available for medication costs in the USA.

Ambulatory care

A survey of a sample of the US population found that each year 12 % of persons with diarrhea (about 25 million) visited a medical provider for their illness, and another 20 % (about 42 million) consulted a provider by telephone. The estimate for the median cost of diarrhea outpatient visits is $47–57 [47].

Hospitalization

Diarrhea-associated hospitalization has been decreasing, with an estimate of 1.5 % of all hospitalizations among adults in the USA. In children under 5 years of age, diarrhea accounts for 2 % of all hospitalizations. The median cost of diarrhea-associated hospitalization can add several thousand dollars to the cost. *C. difficile*-associated diarrhea causes death in 1–2 % of affected patients, and the estimate for the US healthcare costs attributable to *C. difficile*-associated diarrhea is over $600 million in excess healthcare costs and over 600,000 excess hospital days in nonfederal facilities [48].

Prevention

For developed countries, the use of vaccination against rotavirus has decreased the incidence of diarrheal illnesses in children. Improvement in hygiene, use of available vaccines and self-medication with bismuth subsalicylate are the best prevention for traveler's diarrhea. Vaccines to prevent traveler's diarrhea are in development. For developing countries, prevention should target vaccination and sanitation. A study in the Netherlands looked at attack rates for hepatitis A, Shigellosis, and typhoid fever in those traveling to high-risk areas between 1995 and 2006. These are all fecal-orally transmitted infections, and declining attack rates may reflect improved hygiene in these countries [49] although better education and prevention strategies may have played a role. Oral killed whole cell vaccines for cholera can prevent 50–60 % of cholera cases for up to two years [50]. These would have a great impact on acute illness, as well as on the high rate of mortality and morbidity.

Topical issues

Certain aspects of infectious diarrhea continue to pose intriguing questions for the epidemiologist.

1 Food safety:
- Research and epidemiologic tools such as electronic databases, for example FoodNet, are excellent tools for epidemiology. The Food-borne Viruses in Europe Network can enhance understanding of transmission of enteric viruses as there is often little documentation of food- and water-borne epidemics. Improved detection methods such as in viruses in food will be valuable. Person-to-person transmission is important. Molecular typing is an important tool in evaluation of epidemics. Pulse Net (the Foodborne Disease Outbreak Surveillance System) is a national US network to detect food-borne disease outbreaks, launched in 1996, in which all 50 states now participate. They estimate that for every case reported, many more go unreported.
- Expert advice and education are crucial. One good example is the Codex Alimentarius (a WHO website) [51].
- Current food safety standards in developed countries to protect bacterial infection are not adequate to protect from viral contamination, nor from all bacterial infections.

2 The epidemiology of tropical sprue remains obscure. Antibiotic treatment in combination with folate normalizes the mucosal structure and favors the hypothesis of an infectious etiology. Further studies may clarify its cause and the risk factors associated with the high incidence of relapses.

3 Recently, community-acquired *C. difficile* diarrhea has been increasing in incidence. Given its cost and its impact on quality of life, early diagnostic strategies and better therapy for severe cases are needed.

4 "Brainerd" diarrhea is another example of diarrhea that is thought to be infectious but for which no agent

has been identified. Six outbreaks have been reported in the USA, but this may be an underestimate. Further investigations to identify the etiology and the risk factors would help in treating and preventing the condition.

Recommendations for future studies

There is still much to do in areas of clinical research, epidemiology, and public health in order to combat infectious diarrhea more effectively.

- Research into understanding the biology of existing and emerging pathogens continues to be important.
- Controlling the incidence of infectious diarrhea in developing countries through educational programs to promote hygienic behavior and food processing.
- Developing an enteric vaccine coupled with studies of immunoprophylaxis.
- Defining rates and risk factors for PI-IBS as well as the most common associated pathogens, and determining whether medical treatment makes a difference to incidence or prevention.
- Investigating probiotics as a promising option in treating a variety of diarrheal disorders, including rotavirus diarrhea, *C. difficile* diarrhea, and traveler's diarrhea. Future studies should determine their efficacy over the long term for prevention and treatment.

Conclusions

There are 2–4 billion cases of food-borne and diarrheal disease worldwide, resulting in 3.1 million deaths, mostly of children living in developing countries. Despite expensive stool tests, most are of unknown etiology. Identifying the risk factors and the susceptible persons can lead to better strategies of prevention. In recent years, a global effort has tried to decrease the burden of this disease by improving hygiene, developing vaccines, and formulating prophylaxis guidelines. However, diarrheal diseases are still a major challenge.

Multiple choice questions

Case 1

A group attends a school picnic. Six hours after the picnic, several children and adults develop nausea and vomiting that lasts for several hours. No one has fever or chills, or any significant diarrhea.

1 Of the following foods, which is a likely pathogen?
- **A** *Staphylococcus aureus*
- **B** *Shigella flexneri*
- **C** *Campylobacter*
- **D** Norovirus

2 Of the following foods, which is most likely to be implicated?
- **A** Stir-fried beefsteak
- **B** Coleslaw with mayonnaise
- **C** Green beans with a vinaigrette dressing
- **D** Apple pie with ice cream

3 The disease in this case is usually self-limited.
- **A** True
- **B** False

4 In this case there is a risk of spread to other family members over the next few days.
- **A** True
- **B** False

Case 2

A 3-year-old boy complains of a stomach ache; he has no fever or chills, no one else is sick at home. The family had been to the county fair three days earlier and visited the petting zoo, and eaten hamburgers and French fries. The boy develops diarrhea; two days later there is blood in the stools and he is hospitalized.

1 Which is the best next step in therapy?
- **A** This is likely a rotavirus infection and he should be given a vaccine
- **B** He should be given loperamide and ciprofloxacin
- **C** Other family members should be given a vaccine
- **D** Intravenous fluids are the best first step

2 CT scan shows colon wall thickening, right colon worse than left. In the hospital, the intern notes petechiae and the nurses note anuria. Labs show a hematocrit 32, WBC 12,500 with a left shift, platelet count of 82,000, BUN of 45, and creatinine of 2.3. Which statement is correct?
- **A** Mortality is unlikely
- **B** Seizures can be a complication
- **C** Antibiotics would have prevented this complication
- **D** Children are less likely to develop this complication

References

1 World Health Organization. Media centre fact sheet: Diarrhoeal disease. Available at www.who.int/mediacentre/factsheets/fs330/en/index.html (last accessed June 12, 2013).

2 Mead PS, Slutsker L, Dietz V, et al. Food-related illness and death in the United States. *Emerg Infect Dis* 1999;5:607–25.

3 U.S. Department of Health & Human Services. The price of foodborne illness in the USA. *Lancet* 2010;375:866.

4 Flint JA, et al. Estimating the burden of acute gastroenteritis, foodborne disease, and pathogens commonly transmitted by food: an international review. *Clin Infect Dis* 2005;41:698.

5 Anon. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food – 10 states, 2004. *MMWR: Morb Mortal Wkly Rep* 2005;54:352.

6 Parashar UD, Gibson C, Bresee JS, Glass RI. Rotavirus and severe childhood diarrhea. *Emerg Infect Dis* 2006;12:304–6.

7 Palumbo E, Branchi M, Malorgio C, et al. Diarrhea in children: etiology and clinical aspects. *Minerva Pediatricia* 2010;62:347–51.

8 Yen C, Tate JE, Wenk JD, et al. Diarrhea-associated hospitalizations among US children over 2 rotavirus seasons after vaccine introduction. *Pediatrics* 2011;127:e9–e15. Epub.

9 Clark HF, Lawley D, Mallette LA, et al. Decline in cases of rotavirus gastroenteritis presenting to the Children's Hospital of Philadelphia after introduction of a pentavalent rotavirus vaccine. *Clin Vaccine Immunol* 2009;16:382–6.

10 Cortese MM, Tate JE, Simonsen L, et al. Reduction in gastroenteritis in United States children and correlation with early rotavirus vaccine uptake from national medical claims databases. *Pediatr Infect Dis J* 2010;29:489–94.

11 Kosec M, et al. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *B World Health Organ* 2003;81.

12 Allos BM. Campylobacter jejuni infections: update on emerging issues and trends. *Clin Infect Dis* 2001;32:1201.

13 Ford-Jones EL, et al. Hospitalization for community-acquired rotavirus-associated diarrhea: a prospective, longitudinal, population-based study during the seasonal outbreak. The Greater Toronto Area/Peel Region PRESI Study Group. Pediatric Rotavirus Epidemiology Study for Immunization. *Arch Pediatr Adolesc Med* 2000;154:578.

14 Harris JB, et al. Blood group, immunity, and risk of infection with *Vibrio cholerae* in an area of endemicity. *Infect Immun* 2005;73:7422.

15 Hutson AM, et al. Norwalk virus infection and disease is associated with ABO histo-blood group type. *J Infect Dis* 2002;185:1335.

16 Jiang Z-D, et al. Genetic susceptibility to enteroaggregative *Escherichia coli* diarrhea: polymorphism in the interleukin-8 promotor region. *J Infect Dis* 2003;188:506.

17 Villani AC, Lemire M, Thabane M, et al. Genetic risk factors for post-infectious irritable bowel syndrome following a waterborne outbreak of gastroenteritis. *Gastroenterology* 2010;138:1502–13.

18 Newell DG, Koopmans M, Verhoef L, et al. Food-borne diseases – the challenges of 20 years ago still persist while new ones continue to emerge. *Int J Food Microbiol* 2010;139:S3–S15.

19 Centers for Disease Control and Prevention (CDC). FoodNet Annual Report. Atlanta, USA, 2005. Available at: http://www.cdc.gov/foodnet/annual/2005/2005_AR_Report.pdf (last accessed June 13, 2013).

20 Centers for Disease Control and Prevention (CDC). Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food – 10 States, 2008. Atlanta, USA 2009. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5813a2.htm (accessed June 13, 2013).

21 Swaminathan A, Torresi J, Schlagenhauf P, et al. A global study of pathogens and host risk factors associated with infectious gastrointestinal disease in returned international travelers. *J Infection* 2009;59:19–27.

22 Steffen R. Epidemiology of traveler's diarrhea. *Clin Infect Dis* 2005;41:S536.

23 Shah N, DuPont HL, Ramsey DJ. Global etiology of traveler's diarrhea: systematic review from 1973 to the present. *Am J Trop Med Hyg* 2009;80:609–14.

24 Taylor M, MacDougall L, Li M, Galanis E, BC Enteric Policy Working Group. The impact of international travel on the epidemiology of enteric infections, British Columbia, 2008. *Can J Public Health* 2010;101:332–6.

25 Aseeri M, Schroeder T, Kramer J, Zackula R. Gastric acid suppression by proton pump inhibitors as a risk factor for *Clostridium difficile*-associated diarrhea in hospitalized patients. *Am J Gastroenterol* 2008;103:2308–13.

26 Cunningham R, Dial S. Is over-use of proton pump inhibitors fueling the current epidemic of *Clostridium difficile*-associated diarrhoea? *J Hosp Infect* 2008;70:1–6.

27 Leonard J, Marshall JK, Moayyedi P. Systematic review of the risk of enteric infection in patients taking acid suppression. *Am J Gastroenterol* 2007;102:2047–56.

28 Bavishi C, DuPont HL. Systematic review: the use of proton pump inhibitors and increased susceptibility to enteric infection. *Aliment Pharmacol Ther* 2011; 34:1269–81.
29 Elixhauser A, Jhung M. *Clostridium difficile*-associated disease in U.S. hospitals, 1993–2005. Agency for Healthcare Research and Quality, Statistical brief #50. April 2008. Available at: www.hcup-us.ahrq.gov/reports/statbrief/Sb50.pdf (accessed June 13, 2013).
30 Khanna S, Pardi DS, Aronson SL, et al. *Am J Gastroenterol* 2012;107:89–95.
31 Johnson S. Recurrent *Clostridium difficile* infection: a review of risk factors, treatments, and outcomes. *J Infect* 2009;58:403–10.
32 Howell MD, Novack V, Grgurich P, et al. Iatrogenic gastric acid suppression and the risk of nosocomial *Clostridium difficile* infection. *Arch Intern Med* 2010;170: 784–90.
33 Johnsen BO, Lingaas E, Torfoss D, et al. A large outbreak of *Listeria monocytogenes* infection with short incubation period in a tertiary care hospital. *J Infect* 2010;61:465–70.
34 Mattner F, et al. Risk groups for clinical complications of norovirus infections: an outbreak investigation. *Clin Microbiol Infect* 2006;12:69.
35 Calderon-Margalit R, et al. A large-scale gastroenteritis outbreak associated with Norovirus in nursing homes. *Epidemiol Infect* 2005;133:35.
36 Zimmerman CM, et al. Cost of diarrhea-associated hospitalizations and outpatient visits in an insured population of young children in the United States. *Pediatr Infect Dis J* 2001;20:14.
37 Kosec M, et al. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *B World Health Organ* 2003;81:391.
38 Parashar UD, Hummelman EG, Bresee JS, et al. Global illness and deaths caused by Rotavirus disease in children. *Emerg Infect Dis* 2003;9:565–72.
39 Morrow AL, et al. Human milk protection against infectious diarrhea: implications for prevention and clinical care. *Semin Pediatr Infect Dis* 2004;15: 221.
40 Banatvala N, et al. The United States National Prospective Hemolytic Uremic Syndrome Study: microbiologic, serologic, clinical, and epidemiologic findings. *J Infect Dis* 2001;183:1063.
41 Bhutta ZA, et al. Persistent and chronic diarrhea and malabsorption: Working Group Report of the Second World Congress of Pediatric Gastroenterology, Hepatology, and Nutrition. *J Pediatr Gastr Nutr* 2004;39(Suppl 2):S711.
42 Okhuysen PC, et al. Post-diarrhea chronic intestinal symptoms and irritable bowel syndrome in North American travelers to Mexico. *Am J Gastroenterol* 2004;99:1774.
43 Marshall JK, Thabane M, Garg AX, et al. Walkerton Health Study Investigators. Eight-year prognosis of postinfectious irritable bowel syndrome following waterborne bacterial dysentery. *Gut* 2010;59: 605–11.
44 National Institute of Allergy and Infectious Disease. Foodborne diseases, February 2005; http://www.niaid.nih.gov/topics/foodborne/Pages/default.aspx (accessed June 13, 2103).
45 Lee BP, et al. Nonmedical costs associated with rotavirus disease requiring hospitalization. *Pediatr Infect Dis J* 2005;24:984.
46 Hellard ME. Cost of community gastroenteritis. *J Gastroenterol Hepatol* 2003;18:322.
47 Zimmerman CM, et al. Cost of diarrhea-associated hospitalizations and outpatient visits in an insured population of young children in the United States. *Pediatr Infect Dis J* 2001;20:14.
48 McDonald LC, et al. *Clostridium difficile* infection in patients discharged from US short-stay hospitals, 1996–2003. *Emerg Infect Dis* 2006;12(3):409–15.
49 Baaten GG, Sonder GJB, Schim van der Loeff MF, et al. Fecal-orally transmitted diseases among travelers are decreasing due to better hygienic standards at travel destination. *J Travel Med* 2010;17:322–8.
50 Sinclair D, Abba K, Zaman K, et al. Oral vaccines for preventing cholera. *Cochrane Database Syst Rev* 2011;3:CD008603.
51 Codex Alimentarius Commission (2003) Recommended International Code of Practice General Principles of Food Hygiene, CAC/RCP 1-1969, Rev 4-2003. Food and Agriculture Organization/World Health Organization, Rome, Italy.

Answers to multiple choice questions

Case 1

1. A

The onset of vomiting several hours after eating a meal suggests a preformed toxin; most likely would be *Staphylococcus aureus*. If fried rice had been eaten, *Bacillus cereus* would be another possibility. Both *Shigella flexneri* and *Campylobacter* are invasive pathogens so there would be a longer duration between eating the contaminated food and the onset of illness. Norovirus is another possibility but would typically have both a longer duration and incubation period.

2. B
The most likely implicated food would be coleslaw as it has dairy in it. Cooked beef that is not ground beef is an unlikely cause for pathogens. The other foods would not likely be implicated but something with a dairy base is a good medium for the Staphylococcus.
3. A
The typical illness lasts less than 24 hours. The major complication is volume depletion.
4. B
This is unlikely to spread to family members from the person who is sick, unless they eat the same contaminated food that contains the preformed toxin.

Case 2
1. C
The illness suggests an invasive colonic pathogen, thus rotavirus infection which affects the small bowel is unlikely. This is very likely Shigatoxin *E. coli*, thus antidiarrheals and antibiotics are not indicated as they can increase the risk of hemolytic uremic syndrome. This organism can spread from person to person by the fecal-oral route.
2. B
The colon wall thickening and the development of symptoms of hemolytic uremic syndrome again suggest Shiga toxin *E. coli*. Seizures can be a complication of this due to the cerebral vasculitis. Children and the elderly are more likely to develop this complication. Antibiotics would not prevent this complication and might have made it more likely. This can be associated with significant morbidity and mortality.

24 Epidemiology of inflammatory bowel disease

Edward V. Loftus, Jr.
Division of Gastroenterology & Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

Key points

- Crohn's disease and ulcerative colitis, the two major subtypes of inflammatory bowel disease, remain idiopathic.
- The incidence of these conditions continues to rise in both industrialized and developing countries, and as many as 1 in 200 persons has a form of inflammatory bowel disease.
- Cigarette smoking is protective against the development of ulcerative colitis and a risk factor for Crohn's disease, while former smokers are at increased risk of ulcerative colitis.
- A history of appendectomy for appendicitis is protective against the development of ulcerative colitis. Appendectomy may be a risk factor for Crohn's disease but this may be confounded by the initial presentation of the latter.
- The hypothesis that growing up in a clean environment with lack of exposure and "tolerization" to pathogens causes inflammatory bowel disease is intriguing but unproven.
- Crohn's disease and ulcerative colitis result in substantial morbidity, with up to 30 % of colitis patients requiring colectomy and between 60 % and 80 % of Crohn's disease patients requiring at least one intestinal resection.

Clinical summary

Ulcerative colitis is a chronic inflammatory bowel disease characterized by mucosal inflammation of the rectum and/or colon. Crohn's disease, the other major subtype of idiopathic IBD, is characterized by transmural and sometimes granulomatous inflammation of the gastrointestinal tract, most commonly in the ileum and colon. Typical symptoms of ulcerative colitis include diarrhea, rectal bleeding, tenesmus, and increased stool urgency and frequency, while the most common symptoms of Crohn's disease are abdominal pain, fatigue, and diarrhea. Complications of ulcerative colitis include toxic megacolon and colorectal dysplasia or cancer. Complications of Crohn's disease include intestinal stenosis, fistulas, abscesses, and, less commonly, intestinal cancer. The pathogenesis of these conditions remains unclear, but they are thought to arise due to a combination of defective mucosal immune regulation in the gut combined with exposure to as-yet-undetermined environmental factors or luminal antigens. The diagnosis of ulcerative colitis is typically made by endoscopy and biopsy of the colorectum. The diagnosis of Crohn's disease may be difficult, as there are no pathognomonic features, and a host of modalities may be required, including colonoscopy with ileoscopy, small bowel radiography (barium-, computed

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

tomography-, or magnetic resonance-based), capsule endoscopy, or serologic markers. Treatment of these diseases consists of one or more of the following medication classes: 5-aminosalicylates, corticosteroids, antibiotics, immunosuppressive agents (thiopurines, methotrexate, or calcineurin inhibitors), or biologic (anti-cytokine or anti-adhesion molecule) agents. Survival in ulcerative colitis is essentially similar to the general population, although symptoms can contribute to decreased quality of life, and, if symptoms are refractory to medical therapy, colectomy may be required. Most patients with Crohn's disease require at least one intestinal resection (due to intestinal complications or to medically refractory disease) during the course of their illness. Overall survival in Crohn's disease is somewhat diminished compared to the general population, with relative mortality rates ranging from 20–100 % higher than expected.

Disease definition

The idiopathic inflammatory bowel diseases (IBD) consist of two major subtypes, ulcerative colitis and Crohn's disease (aka "regional enteritis" or "granulomatous colitis"). Disease classification can be troublesome, since these diagnoses are clinical ones and there is no single test that is pathognomonic for either condition. Ulcerative colitis is characterized by a continuous, confluent mucosal inflammation of the large intestine, almost always with rectal involvement, in the absence of infection, ischemia, and radiation exposure [1]. Crohn's disease, a more heterogeneous disorder, requires one or more of the following for diagnosis: (i) granulomatous transmural inflammation of the gastrointestinal tract (anywhere from mouth to anus); (ii) discontinuous involvement with "skip areas"; (iii) and a propensity for intestinal stenosis and/or fistula [1]. One of the challenges in interpreting epidemiologic research in IBD is that disease definitions have not been consistent. Further compounding the classification problem is the use of "indeterminate colitis" in some epidemiologic studies. Originally a term used by pathologists to describe surgical resection specimens with features of both ulcerative colitis and Crohn's disease, "indeterminate colitis" is now used by some clinicians and epidemiologists to describe chronic colitis that does not readily fall into either one of the two classic IBD subtypes. The Montreal Working Party modification of the Vienna classification of IBD proposed renaming indeterminate colitis to "inflammatory bowel disease unclassified" or IBDU [2]. In some population-based investigations, cases of indeterminate colitis are tracked separately, while in other studies these cases are "forced" into one of two categories.

Incidence and prevalence

In high-incidence areas such as North America, the incidence of ulcerative colitis ranges from 8.8 cases per 100,000 person-years [3] to 14.6 per 100,000 [4], and the incidence of Crohn's disease ranges from 7.9 per 100,000 [3] to 14.8 per 100,000 [4]. In other words, if we assume that the combined population of the United States (USA) and Canada is 333 million persons, between 29,000 and 49,000 Americans and Canadians are diagnosed with ulcerative colitis annually, while between 26,000 and 49,000 are diagnosed each year with Crohn's disease.

The prevalence of ulcerative colitis in North America in 2001 ranged from 191 cases per 100,000 persons [5] to 241 per 100,000 [4], and the prevalence of Crohn's disease ranged from 129 cases per 100,000 [5] to 270 per 100,000 [4]. The overall prevalence of IBD was approximately 0.25–0.3 % in a northern California health maintenance organization (HMO) [6], 0.4 % in multiple US settings (Olmsted County, Minnesota [3], a study of nine HMOs in the USA [5], and in a geographically diverse insurance claims database from the USA [7]), and approximately 0.5 % in Canada [8]. If these estimates from 2001 are extrapolated to the current population of the USA and Canada (348 million), they imply that between 665,000 and 839,000 American and Canadians suffer from ulcerative colitis and that between 449,000 and 940,000 carry a diagnosis of Crohn's disease, for a combined total of between 1.1 and 1.8 million persons.

In general, the incidence of inflammatory bowel disease has continued to rise since these clinical entities were first recognized. There was a suggestion in some high-incidence areas of a rapid postwar rise in incidence until the late 1960s and early 1970s, and then a stabilizing of the incidence rate in the 1980s and 1990s [9,10], but the finding of a "plateau in incidence" was not universal. Furthermore, several recent studies in areas with excellent longitudinal data have

demonstrated a continued increase in incidence rates [11–13]. As a consequence of rising incidence rates and normal or near-normal life expectancy with these conditions, the prevalence of both Crohn's disease and ulcerative colitis has continued to gradually rise.

Risk factors for disease

Age and gender

In a systematic review of the epidemiology of Crohn's disease from population-based cohorts from North America, the mean age at diagnosis ranged from 33 to 39 years [14]. The median age at diagnosis of Crohn's disease among Olmsted County residents as of 2000 was 29 years (range, 4–91) [3]. Whether there still exists a "bimodal distribution" in the age of onset of Crohn's disease is controversial, since a number of population-based studies no longer demonstrate this [3,15]. For ulcerative colitis, the average age at diagnosis tends to be slightly higher – in Olmsted County, the median age at colitis diagnosis was 33 years (range, 1–88) [3]. Some (but not all) recent studies have demonstrated a gender divergence in incidence of ulcerative colitis later in life [3,15], in that men are significantly more likely to be diagnosed with colitis in the sixth or seventh decades of life than women. The mechanism for the divergence remains undiscovered, but some have speculated that differential patterns of cigarette smoking status might play a role.

Slight differences in IBD incidence by gender may exist. Males are more likely than females to develop ulcerative colitis [10], while there may be a very slight female predominance in Crohn's disease.

Geography

Ulcerative colitis and Crohn's disease have been classically described most often in developed countries in northern climates, such as northern Europe and Scandinavia, the United Kingdom, and North America. However, both ulcerative colitis and Crohn's disease are being described more often in other regions, such as southern and eastern Europe [16], Asia [17,18], Africa, and Latin America [19]. The major subtypes of IBD are truly worldwide diseases. A north–south gradient of incidence has been described in Europe, but a multicenter study from the early 1990s suggested that this has to some extent dissipated [15]. Geographic differences in incidence from east to west have recently been identified in Canada – in general, the highest incidence rates and prevalence are noted in the Maritime province of Nova Scotia, while the lowest are seen in the far western province, British Colombia [8]. However, these differences may be attributable to the higher proportion of ethnic minorities residing in British Colombia. One interesting phenomenon that has been observed in several areas with formerly low incidence rates is that ulcerative colitis is first observed in a region, followed a decade or two later by Crohn's disease [17].

Race/ethnicity

Differences in incidence and prevalence of IBD between Caucasians and other racial and ethnic groups seem to have lessened over time. A 2009 systematic review of 28 publications concluded that IBD incidence and prevalence were rising in Hispanics and Asians, that Hispanics and Asians were much more likely to be diagnosed with ulcerative colitis while African Americans were more likely to be diagnosed with Crohn's disease, and that fistulizing disease was common among Crohn's disease patients from all three ethnic groups [20]. In Manitoba, Canada, it is clear that aboriginal Canadians are significantly less likely to develop IBD [21].

In military veteran studies that are now almost 50 years old, those of Jewish ancestry had a markedly increased risk of IBD relative to non-Jewish Caucasians. A study from Wales, UK later confirmed this difference [22]. Population-based investigations in Israel suggested that Ashkenazi Jews from Europe and the USA were more likely than Sephardic Jews from the Mediterranean region to develop IBD, but these differences have narrowed in succeeding generations [23,24].

Studies of migrant populations yield clues that environmental factors and/or lifestyle play a role in risk differences. South Asians who move to the United Kingdom are within one generation actually at higher risk to develop ulcerative colitis than those of European descent, and within the South Asian population Sikhs may be at higher risk for colitis than Hindus and Muslims [25,26]. Furthermore, South Asians who move to Singapore are at increased risk of colitis relative to ethnic Chinese and Malays [27].

Socioeconomic factors

The data are mildly conflicting, but in general there is a positive correlation between socioeconomic class and the risk of inflammatory bowel disease. For example, in a study from Manitoba, incidence rates from postal codes in the top tertile of income were 20 % higher than rates from postal codes in the lowest tertile of income [21]. Another study from Manitoba incorporating census data on education and income could not demonstrate a relationship between these variables and IBD risk [28]. Nevertheless, the bulk of available data suggests that those of higher socioeconomic status are at increased risk for IBD.

Familial aggregation/genetics

Both familial aggregation studies and twin studies suggest that genetic factors play a role in susceptibility to these conditions. The relative risk of IBD for a sibling of a proband with IBD ranges from 15 to 35 for Crohn's disease and from 7 to 17 for ulcerative colitis [29]. Twin studies from Scandinavia and the United Kingdom demonstrate a concordance for Crohn's disease ranging from 20 % to 50 % for monozygotic twins versus 0 % to 7 % for dizygotic twins [30]. For ulcerative colitis, the concordance ranges from 14 % to 19 % for monozygotic twins and from 0 % to 7 % for dizygotic twins. Both types of studies suggest that genetic influences are stronger in Crohn's disease than in ulcerative colitis.

The search for susceptibility genes in IBD has employed both candidate gene investigations and genome-wide scans. Several susceptibility loci have been identified. In 2001, several groups reported that NOD2/CARD15 was the gene at the IBD1 susceptibility locus, and this association has been confirmed in numerous populations [31–33]. Up to 30–40 % of Crohn's disease patients carry at least one of three polymorphisms of this gene, which encodes for a protein that recognizes muramyl dipeptide, a bacterial antigen. The relative risk of Crohn's disease in heterozygotes for one of the mutations is 2 to 3, while the risk in homozygotes or compound heterozygotes may be 40 times that of the general population. However, the exact way in which gene expression results in disease must be complex, since the vast majority of persons with these mutations do not develop Crohn's disease. The advent of large genome-wide association scans has resulted in the identification of over 70 susceptibility loci in Crohn's disease and over 45 in ulcerative colitis [34]. The implicated genes are involved in a number of pathways including regulation of innate immunity (pattern recognition), autophagy (self-breakdown of cellular components), epithelial barrier function, regulation of B- and T-lymphocytes, oxidative stress, and immune tolerance [34]. It has been estimated that the susceptibility loci identified to date explain at most one quarter of the heritability of these conditions, so the remainder must be explained through rare loci, epigenetics, or pure environmental factors [34].

Cigarette smoking

The curious inverse relationship between cigarette smoking and ulcerative colitis has been recognized for 25 years. Current smokers are 20–90 % less likely to develop ulcerative colitis than nonsmokers. A recent meta-analysis using rigorous criteria pooled the results of 13 studies and estimated a pooled risk reduction of 42 % [35]. Conversely, former smokers have an 80 % higher risk of ulcerative colitis than never-smokers [35]. The mechanism behind this association remains unclear, but effects on rectal blood flow, colonic mucus production, mucosal IgA production, and synthesis of prostaglandins, leukotrienes, and cytokines have been variously implicated. Smoking status may affect the clinical course of ulcerative colitis. Current smokers are half as likely to require hospitalization, while ex-smokers are twice as likely to undergo colectomy [36]. Transdermal nicotine is superior to placebo for clinical improvement of ulcerative colitis, but it does not appear to be superior to placebo for induction of clinical remission [37,38].

A number of studies have suggested that cigarette smoking is a risk factor for Crohn's disease. The aforementioned meta-analysis pooled the results of nine studies and estimated that current smokers are 76 % more likely than nonsmokers to develop Crohn's disease [35]. Moreover, active smokers who have Crohn's disease have a more severe clinical course than nonsmokers, as measured by need for additional surgery after resection and need for immunosuppressive drugs [39,40]. Indeed, patients who quit smoking actually have an improved clinical course, with fewer exacerbations and less need for corticosteroids or

immunosuppressives than those who continued to smoke [41].

Appendectomy

Next to cigarette smoking, a history of appendectomy is the best-established risk modifier in IBD. The inverse association between appendectomy and ulcerative colitis was first noted 20 years ago [42] and this relationship has been confirmed numerous times. A 2002 meta-analysis of 17 case-control studies yielded a pooled relative risk of 0.3; in other words, a nearly 70 % risk reduction in ulcerative colitis following an appendectomy [43]. A large Swedish cohort study suggested that the indication for appendectomy influenced the magnitude of protective effect [44]. The incidence of ulcerative colitis among the 212,000 people who had undergone appendectomy was approximately 75 % that of the controls who had not undergone the procedure, but there was no protective effect if appendectomy had been performed for abdominal pain (i.e. no clear-cut evidence of appendicitis). Several reports have also suggested that ulcerative colitis occurring after appendectomy has a milder clinical course, with lower likelihood of requiring immunosuppressive therapy or colectomy [45,46], but data are conflicting [47].

The relationship between appendectomy and risk of Crohn's disease is less clear [48]. Although most studies suggest that appendectomy increases the risk of Crohn's disease, one has to take into account the fact that the highest risk is seen in the first year after appendectomy, suggesting that the results may be confounded by patients presenting with acute abdominal pain, undergoing appendectomy, but in actuality having Crohn's ileitis. If patients developing Crohn's disease within one year of appendectomy are eliminated from analyses, the relative risks are still elevated, but not to the same magnitude [49].

Oral contraceptives

Analyses of oral contraceptives as a risk factor for IBD have yielded conflicting results, with some studies suggesting as high as a fivefold elevated risk of Crohn's disease in women who had used oral contraceptives for at least 6 years [50], but other studies demonstrating no association. An outdated meta-analysis estimated the pooled relative risk to be 1.4 for Crohn's disease and 1.3 for ulcerative colitis (the latter value was not statistically significant). Most subsequent studies have demonstrated similarly weak associations, but the most notable recent study, employing 444 incident cases of IBD and 10,000 controls from the United Kingdom, suggested a stronger association [51]. Users of oral contraceptives were two to three times more likely than nonusers to develop IBD. In the same study, women on hormone replacement therapy were more than twice as likely as women not on such therapy to develop Crohn's disease, but no association with ulcerative colitis was detected [51]. A 2008 meta-analysis of 14 observational studies involving over 75,000 patients yielded a pooled relative risk of 1.51 for Crohn's disease for women currently taking an oral contraceptive (95 % CI 1.17–1.96) and 1.53 for ulcerative colitis (1.21–1.94) [52]. Data on the effect of oral contraceptives on the clinical course of IBD are conflicting. Somewhat reassuring is a 2010 systematic review of five cohort studies examining the relationship between oral contraceptive use and clinical course of IBD [53]. The authors concluded that no significant association between contraceptive use and IBD relapses existed.

Antibiotics

Perhaps alteration of the intestinal microenvironment could serve as a trigger for inflammation in susceptible individuals. The role of antibiotics in the risk of IBD has been explored in several studies, and significant associations have been observed [42,54]. The strongest of these studies was a nested case-control study from a large database in the United Kingdom, where all prescriptions were recorded prospectively [55]. The use of antibiotics in the preceding 2–5 years increased the risk of Crohn's disease by 32 %, after adjusting for age, gender, and smoking [55]. A nested case-control study within a cohort of IBD patients from Manitoba, Canada (where prescription data could also be linked to individual patients) found that 58 % of cases with pediatric-onset IBD had received at least one prescription for antibiotics in the first year of life compared to 39 % of age- and sex-matched controls [56]. Overall, patients receiving at least one antibiotic prescription had a threefold increased risk of IBD, while those patients receiving at least five prescriptions were at a fivefold increase in IBD risk [56].

Using the same database, Shaw and colleagues examined the association between antibiotic use and IBD risk in the entire population (not restricted to pediatric patients) [57]. Among 2234 IBD cases, 12 % had received at least three prescriptions for antibiotics 2–5 years before diagnosis, while only 7 % of controls had, yielding a 50 % increase in IBD risk for those patients receiving at least three prescriptions for antibiotics 2 years prior to the index date [57]. The effect persisted even with antibiotic usage 5 years prior to the index date. These studies indeed suggest that manipulation of the enteric microflora may predispose certain individuals to the development of Crohn's disease and ulcerative colitis.

Diet

Although it is logical and tempting to blame dietary factors on the increasing incidence of IBD, to date there is no definitive proof that a particular diet is protective or a risk factor. Studies examining this relationship are difficult to perform, because attempts to recall pre-diagnosis dietary intake are inaccurate. The most consistent relationship has been between Crohn's disease and sugar – numerous case-control studies have detected significant associations between Crohn's disease and intake of refined sugar [58], but there has always been concern that this association may be skewed by the fact that patients with Crohn's disease may have altered their diet in attempt to control their symptoms. A recent Japanese case-control study suggested that higher consumption of sweets was significantly associated with the risk of ulcerative colitis, while higher consumption of sugars, sweets, fats, and fish were associated with Crohn's disease risk [59]. The quality of such dietary studies has improved of late because several have been performed within the context of large prospective cohorts [60–62]. A 2011 systematic review of 19 studies involving over 2600 IBD cases and over 4000 controls, noted that increased intake of total fats, polyunsaturated fatty acids, omega-6 fatty acids, and meat were associated with IBD risk [63]. In addition, a significant association between mono- and disaccharides and Crohn's disease risk was noted. In contrast, high intake of fiber and fruit were inversely associated with Crohn's disease risk and high vegetable intake was inversely associated with ulcerative colitis risk [63]. It is interesting to speculate that diet might exert its effects on IBD risk indirectly via changes in the fecal microbiome.

Hygiene hypothesis

The incidence of allergic and immune-mediated diseases (e.g. asthma, multiple sclerosis) has risen in industrialized countries. There exists in the asthma and diabetes literature a "hygiene hypothesis", such that lack of exposure to pathogens predisposes one to disease, perhaps because of a failure to induce tolerance [64]. The reverse corollary to this is that certain infections in childhood, such as helminthic infestation and *Helicobacter pylori* infection, might be protective. Within the field of IBD, the term "hygiene hypothesis" could be used as an umbrella term to encompass several aspects of IBD epidemiology such as vaccination, early antibiotic use, breastfeeding, family size, birth order, domestic hygiene, and socioeconomic status [65]. Such a hypothesis might explain a higher incidence of IBD in developed countries as well as the association between higher socioeconomic status and risk of IBD. Several studies have examined the influence of a clean household environment during childhood on the risk of IBD and have yielded conflicting results [66–69].

Infection

The role of infection in promoting IBD risk is confusing, since a number of studies suggest that certain childhood infections may actually increase, not decrease, IBD risk. Recurrent respiratory infections in childhood, perinatal infections, and recurrent pharyngitis in childhood, periodontitis, and hand-foot-mouth disease have all been associated with Crohn's disease [42,54,70,71]. Two large cohort studies have found that the risk of IBD, especially Crohn's disease, was between 40 % and 100 % higher among those developing acute gastroenteritis compared to unaffected controls [72,73].

Mycobacterium avium paratuberculosis (MAP) causes a granulomatous wasting disease in cattle called Johne's disease. MAP was first cultured from the intestinal tissue of Crohn's disease patients over 20 years ago, but confirmatory reports have been inconsistent. Concerns about a lack of specificity of this finding have been raised, because atypical mycobacteria can be recovered from healthy controls.

Polymerase chain reaction technology has been utilized to recover mycobacterial DNA from the intestinal tissue of Crohn's disease patients, but the relatively high rate of recovery from controls again raises questions about the specificity of this finding. A 2000 meta-analysis of seven relatively small randomized trials of antimycobacterial therapy for Crohn's disease suggested a modest benefit [74]. However, in a placebo-controlled randomized study of 2 years of clarithromycin, rifabutin, and clofazimine in 213 patients with Crohn's disease, the rates of remission at any of the three predetermined endpoints were not significantly different between the two treatment arms [75]. In a provocative study from Florida, MAP DNA could be extracted from buffy coat preparations of 46 % of Crohn's patients compared with 20 % in controls, and viable MAP could be cultured from the blood of 50 % of Crohn's disease patients versus none of the controls [76]. What cannot be explained by the MAP theory of Crohn's disease is why patients seem to improve, not worsen, with anti-TNF agents, which are known to cause reactivation of latent *Mycobacterium tuberculosis*.

Fecal microbiome

The science of understanding the constituents and roles of the fecal microbiome in intestinal homeostasis and disease pathophysiology is still in its infancy. If IBD represents the end result of an abnormal interplay between luminal antigens and the gut immune response in a genetically susceptible individual, then it stands to reason that "intestinal dysbiosis" could increase the risk of developing IBD. The fecal microbiome in patients with IBD appears to differ from that of normal subjects in several ways, including a reduction in diversity and numbers of commensal microbes, reduction in members of the phylum Firmicutes, and in particular *Faecalibacterium prausnitzii* [77–79]. The secreted products of *F. prausnitzii* have been shown to exhibit immunomodulatory properties.

Bernstein and Shanahan speculated that several aspects of our modern lifestyle, including improved sanitation, decline in endemic parasitism, decreased exposure to soil microbes, decline in *Helicobacter pylori* infection, increased antibiotic usage, less-crowded living conditions, refrigeration, and dietary changes, have all contributed to changes in the fecal microbiome which predispose to the development of IBD [80].

Natural history and mortality

Ideally, natural history studies should be performed in a population-based fashion, patients should be followed from time of diagnosis, and follow-up should be long enough and complete enough to measure the outcome of interest. Since ulcerative colitis and Crohn's disease are diseases of a lifetime, and important events such as surgery occur relatively infrequently, well-designed natural history studies are scarce. The natural history of Crohn's disease, including need for hospitalization and surgery, intestinal cancer, extraintestinal manifestations, and mortality, has recently been reviewed [81,82].

For Crohn's disease, after the first 2 years following diagnosis, the course is waxing and waning [83]. At any given point in time after the first 2 years, approximately 30 % of patients have moderate to high disease activity, 15 % have low disease activity, and 55 % are in symptomatic remission. For individual patients, of course, the disease activity may change from year to year. Roughly 25 % have continuously active disease, 20 % remain in prolonged remission, and the other 55 % have a waxing and waning course [83]. The cumulative probability of at least one surgical resection ranges from 64 % at 30 years from diagnosis [84] to 82 % at 20 years [85].

For ulcerative colitis, at any given point in time, 40 to 50 % of patients who have not undergone colectomy are in remission, and the remainder have active disease [86]. Only about 10 % of patients will experience prolonged symptomatic remission, but only 1 % have continuously active disease. Ninety percent have a waxing and waning course. For ulcerative colitis, the cumulative risk of requiring colectomy ranges from 28 % at 30 years from diagnosis [87] to 30 % at 30 years [88].

Colorectal cancer risk is increased in ulcerative colitis, but recent studies indicate that the risk has decreased markedly [88,89]. It is not yet clear if this is due to more widespread use of aminosalicylates, more frequent colonoscopic surveillance, judicious use of colectomy, or other factors [90]. In Crohn's disease, the relative risk of small bowel cancer is elevated, as is the risk of colorectal cancer in those with colonic involvement [89,91].

Recent mortality rates in Crohn's disease, with few exceptions, are higher than those seen in the general population, ranging from 20 % higher than expected [92] to almost twice as high as expected [93]. There is as yet no satisfactory explanation for these differences in relative mortality. Approximately one third of Crohn's disease patients die from disease-related complications. In ulcerative colitis, mortality rates have decreased significantly over time, such that several recent studies have found either similar or decreased mortality compared to the general population [92,94,95]. Roughly 20 % of patients die from colitis-related complications, and there is a suggestion that colitis patients are less likely to die from cardiovascular causes, perhaps due to the inverse association with cigarette smoking [92,95].

Disability and quality of life

Patients with IBD have higher rates of disability than the general population, and quality of life for many patients is diminished. A case-control study from the Netherlands showed a full-time employment rate of 61 % for male Crohn's disease patients and 65 % for those with ulcerative colitis, compared to 75 % for controls [96]. Overall disability rates for males were 33 % for Crohn's, 28 % for colitis, and 12 % for the controls, resulting in disability rates that were more than twice that expected. In Norway, 25 % of women with Crohn's disease were collecting disability pensions [97]. Among participants in the ACCENT I trial of maintenance infliximab for Crohn's disease, 39 % were unemployed and 25 % were receiving disability compensation [98]. Numerous studies have indicated that health-related quality of life is diminished in both Crohn's disease and ulcerative colitis, and that this is largely driven by disease activity.

Healthcare utilization and costs

Several reports indicate that IBD patients are more likely than members of the general population to have an outpatient visit, to see a specialist, and to require an emergency room visit or overnight hospitalization [99,100]. They are more likely than non-IBD patients to use prescription medication. Much of the resource utilization seems to occur in the first 5 years after diagnosis [99,100]. The average annual direct costs of Crohn's disease have ranged from $6,561 in 1990 to $12,417 in 1994 [101]. Furthermore, in Sweden it was estimated that indirect costs of these conditions (lost work productivity, early retirement) account for two-thirds of the total costs [102].

Recommendations for future studies

Many of the observations made in these retrospective studies ideally should be confirmed in different populations. A better understanding of the reasons for disparities in outcomes between different regions or countries is needed – are these due to differences in disease severity (and if so, why), treatment strategy, or quality of care? There are hints in recent literature that biologic agents may be able to alter the occurrence of "hard outcomes" such as hospitalization and surgery, and further studies are required to definitively confirm these observations.

Conclusions

Although etiopathogenic hypotheses abound, Crohn's disease and ulcerative colitis remain idiopathic. The incidence and prevalence of these conditions continue to increase. Cigarette smoking and appendectomy are well-established risk modifiers. The hygiene hypothesis is intriguing but remains unproven. Over time these conditions result in hospitalization, surgery, disability, and sometimes mortality. There is hope that newer treatment agents may be able to alter the natural history of IBD.

Multiple choice questions

1 Which of the following statements is most accurate regarding cigarette smoking and inflammatory bowel disease?

A Ulcerative colitis is more likely to be diagnosed in never-smokers and former smokers than in current smokers

B Patients with ulcerative colitis who smoke are more likely to require hospitalization than those who do not smoke

C Former smokers are more likely than current smokers to be diagnosed with Crohn's disease

D Transdermal nicotine patches were shown in randomized trials to be modestly effective in the induction of clinical remission in Crohn's disease

2 Which of the following statements regarding the descriptive epidemiology of Crohn's disease is most accurate?

A The prevalence of Crohn's disease in North America in the early 2000s was approximately 0.5 %

B The average age at diagnosis of Crohn's disease is approximately 25 years

C The female-to-male ratio in Crohn's disease is 2:1

D The incidence rate of Crohn's disease in North America ranges from approximately 8 cases per 100,000 person-years to 15 per 100,000

3 Which of the following statements regarding the natural history of Crohn's disease is most accurate?

A The cumulative incidence of bowel resection in Crohn's disease ranges from 60 % to 80 % at 20 to 30 years after diagnosis

B The mortality rate in Crohn's disease is lower than that expected in the general population

C Up to 50 % of Crohn's disease patients have a continuously active disease course

D The risk of colorectal cancer in Crohn's disease is similar to that of the general population

References

1 Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol, Supplement* 1989;170:2–6.

2 Silverberg MS, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol* 2006;19(Suppl A):5A–36A.

3 Loftus CG, et al. Update on the incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota, 1940–2000. *Inflamm Bowel Dis* 2007;13:254–61.

4 Green C, et al. A population-based ecologic study of inflammatory bowel disease: searching for etiologic clues. *Am J Epidemiol* 2006;164:615–23.

5 Herrinton LJ, et al. Estimation of the period prevalence of inflammatory bowel disease among nine health plans using computerized diagnoses and outpatient pharmacy dispensings. *Inflamm Bowel Dis* 2007;13:451–61.

6 Herrinton LJ, Liu L, Lewis JD, et al. Incidence and prevalence of inflammatory bowel disease in a northern California managed care organization, 1996–2003. *Am J Gastroenterol* 2008;103:1998–2006.

7 Kappelman MD, et al. The prevalence and geographic distribution of Crohn's disease and ulcerative colitis in the United States. *Clin Gastroenterol Hepatol* 2007;5:1424–9.

8 Bernstein CN, *et al.* The epidemiology of inflammatory bowel disease in Canada: a population-based study. *Am J Gastroenterol* 2006;101:1559–68.

9 Loftus EV, Jr., et al. Crohn's disease in Olmsted County, Minnesota, 1940–1993: incidence, prevalence, and survival. *Gastroenterology* 1998;114:1161–8.

10 Loftus EV, Jr., et al. Ulcerative colitis in Olmsted County, Minnesota, 1940–1993: incidence, prevalence, and survival. *Gut* 2000;46:336–43.

11 Lapidus A. Crohn's disease in Stockholm County during 1990–2001: an epidemiological update. *World J Gastroenterol* 2006;12:75–81.

12 Vind I, et al. Increasing incidences of inflammatory bowel disease and decreasing surgery rates in Copenhagen City and County, 2003–2005: a population-based study from the Danish Crohn colitis database. *Am J Gastroenterol* 2006;101:1274–82.

13 Ingle SB, et al. Increasing incidence and prevalence of inflammatory bowel disease in Olmsted County, Minnesota, 2001–2004. *Gastroenterology* 2007;132(4 Suppl 2):A19–A20.

14 Loftus EV, Jr., et al. The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. *Aliment Pharmacol Therap* 2002;16:51–60.

15 Shivananda S, et al. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Gut* 1996;39:690–7.

16 Lakatos L, et al. Is the incidence and prevalence of inflammatory bowel diseases increasing in Eastern Europe? *Postgrad Med J* 2006;82:332–7.

17 Ouyang Q, et al. The emergence of inflammatory bowel disease in the Asian Pacific region. *Curr Opin Gastroenterol* 2005;21:408–13.

18 Thia KT, et al. An update on the epidemiology of inflammatory bowel disease in Asia. *Am J Gastroenterol* 2008;103:3167–82.

19 Loftus EV, Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 2004;126:1504–17.

20 Hou JK, El-Serag H, Thirumurthi S. Distribution and manifestations of inflammatory bowel disease in Asians, Hispanics, and African Americans: a systematic review. *Am J Gastroenterol* 2009;104:2100–9.

21 Blanchard JF, et al. Small-area variations and sociodemographic correlates for the incidence of Crohn's disease and ulcerative colitis. *Am J Epidemiol* 2001;154:328–35.
22 Mayberry JF, et al. Crohn's disease in Jewish people – an epidemiological study in south-east Wales. *Digestion* 1986;35:237–40.
23 Fireman Z, et al. Epidemiology of Crohn's disease in the Jewish population of central Israel, 1970–1980. *Am J Gastroenterol* 1989;84:255–8.
24 Odes HS, et al. Epidemiology of Crohn's disease in southern Israel. *Am J Gastroenterol* 1994;89:1859–62.
25 Probert CS, et al. Epidemiological study of ulcerative proctocolitis in Indian migrants and the indigenous population of Leicestershire. *Gut* 1992;33:687–93.
26 Carr I, et al. The effects of migration on ulcerative colitis: A three-year prospective study among Europeans and first- and second-generation South Asians in Leicester (1991–1994). *Am J Gastroenterol* 1999;94:2918–22.
27 Lee YM, et al. Racial differences in the prevalence of ulcerative colitis and Crohn's disease in Singapore. *J Gastroen Hepatol* 2000;15:622–5.
28 Bernstein CN, et al. The relationship between inflammatory bowel disease and socioeconomic variables. *Am J Gastroenterol* 2001;96:2117–25.
29 Tamboli CP, et al. What are the major arguments in favour of the genetic susceptibility for inflammatory bowel disease? *Eur J Gastroenterol Hepat* 2003;15:587–592.
30 Halme L, et al. Family and twin studies in inflammatory bowel disease. *World J Gastroenterol* 2006;12:3668–72.
31 Hugot JP, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599–603.
32 Ogura Y, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;411:603–6.
33 Hampe J, et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001;357:1925–8.
34 Khor B, Gardet A, Xavier RJ. The genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;474:307–17.
35 Mahid SS, et al. Smoking and inflammatory bowel disease: A meta-analysis. *Mayo Clin Proc* 2006;81:1462–71.
36 Boyko EJ, et al. Effects of cigarette smoking on the clinical course of ulcerative colitis. *Scand J Gastroenterol* 1988;23:1147–52.
37 Pullan RD, et al. Transdermal nicotine for active ulcerative colitis. *New Engl J Med* 1994;330:811–15.
38 Sandborn WJ, et al. Transdermal nicotine for mildly to moderately active ulcerative colitis – a randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 1997;126:364–71.
39 Timmer A, et al. Oral contraceptive use and smoking are risk factors for relapse in Crohn's disease. The Canadian Mesalamine for Remission of Crohn's Disease Study Group. *Gastroenterology* 1998;114:1143–50.
40 Cosnes J, et al. Effects of current and former cigarette smoking on the clinical course of Crohn's disease. *Aliment Pharmacol Therap* 1999;13:1403–11.
41 Cosnes J, et al. Smoking cessation and the course of Crohn's disease: An intervention study. *Gastroenterology* 2001;120:1093–9.
42 Gilat T, et al. Childhood factors in ulcerative colitis and Crohn's disease. An international cooperative study. *Scand J Gastroenterol* 1987;22:1009–24.
43 Koutroubakis IE, et al. Role of appendicitis and appendectomy in the pathogenesis of ulcerative colitis: a critical review. *Inflamm Bowel Dis* 2002;8:277–86.
44 Andersson RE, et al. Appendectomy and protection against ulcerative colitis. *New Engl J Med* 2001;344:808–14.
45 Radford-Smith GL, et al. Protective role of appendicectomy on onset and severity of ulcerative colitis and Crohn's disease. *Gut* 2002;51:808–13.
46 Cosnes J, et al. Effects of appendicectomy on the course of ulcerative colitis. *Gut* 2002;51:803–7.
47 Selby WS, et al. Appendectomy protects against the development of ulcerative colitis but does not affect its course. *Am J Gastroenterol* 2002;97:2834–8.
48 Radford-Smith GL. The role of the appendix and appendectomy in patients with IBD. *IBD Monitor* 2003;4:120–8.
49 Andersson RE, et al. Appendectomy is followed by increased risk of Crohn's disease. *Gastroenterology* 2003;124:40–6.
50 Boyko EJ, et al. Increased risk of inflammatory bowel disease associated with oral contraceptive use. *Am J Epidemiol* 1994;140:268–78.
51 García Rodríguez LA, et al. Risk factors for inflammatory bowel disease in the general population. *Aliment Pharmacol Ther* 2005;22:309–15.
52 Cornish JA, et al. The risk of oral contraceptives in the etiology of inflammatory bowel disease: a meta-analysis. *Am J Gastroenterol* 2008;103:2394–400.
53 Zapata LB, et al. Review article: contraception use among women with inflammatory bowel disease: a systematic review. *Contraception* 2010;82:72–85.
54 Wurzelmann JI, et al. Childhood infections and the risk of inflammatory bowel disease. *Digest Dis Sci* 1994;39:555–60.

55 Card T, et al. Antibiotic use and the development of Crohn's disease. *Gut* 2004;53:246–50.
56 Shaw SY, et al. Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease. *Am J Gastroenterol* 2010;105:2687–92.
57 Shaw SY, et al. Association between the use of antibiotics and new diagnoses of Crohn's disease and ulcerative colitis. *Am J Gastroenterol* 2011;106:2133–42.
58 Riordan AM, et al. A review of associations between Crohn's disease and consumption of sugars. *Eur J Clin Nutr* 1998;52:229–38.
59 Sakamoto N, et al. Dietary risk factors for inflammatory bowel disease – A multicenter case-control study in Japan. *Inflamm Bowel Dis* 2005;11:154–63.
60 Hart AR, et al. Diet in the aetiology of ulcerative colitis: a European prospective cohort study. *Digestion* 2008;77:57–64.
61 John S, et al. Dietary n-3 polyunsaturated fatty acids and the aetiology of ulcerative colitis: a UK prospective cohort study. *Eur J Gastroenterol Hepatol* 2010;22:602–6.
62 IBD in EPIC Study Investigators, et al. Linoleic acid, a dietary n-6 polyunsaturated fatty acid, and the aetiology of ulcerative colitis: a nested case-control study within a European prospective cohort study. *Gut* 2009;58:1606–11.
63 Hou JK, et al. Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. *Am J Gastroenterol* 2011;106:563–73.
64 Bach JF. The effect of infections on susceptibility to autoimmune and allergic diseases. *New Engl J Med* 2002;347:911–20.
65 Koloski NA, et al. Hygiene hypothesis in inflammatory bowel disease: a critical review of the literature. *World J Gastroenterol* 2008;14:165–73.
66 Baron S, et al. Environmental risk factors in paediatric inflammatory bowel diseases: a population-based case-control study. *Gut* 2005;54:357–63.
67 Amre DK, et al. Investigating the hygiene hypothesis as a risk factor in pediatric onset Crohn's disease: A case-control study. *Am J Gastroenterol* 2006;101:1005–11.
68 Bernstein CN, et al. A population-based case control study of potential risk factors for IBD. *Am J Gastroenterol* 2006;101:993–1002.
69 Lashner BA, et al. True or false? The hygiene hypothesis for Crohn's disease. *Am J Gastroenterol* 2006;101:1003–4.
70 Ekbom A, et al. Perinatal risk factors for inflammatory bowel disease: a case-control study. *Am J Epidemiol* 1990;132:1111–19.
71 Van Kruiningen HJ, et al. Environmental factors in familial Crohn's disease in Belgium. *Inflamm Bowel Dis* 2005;11:360–5.
72 García Rodríguez LA, et al. Acute gastroenteritis is followed by an increased risk of inflammatory bowel disease. *Gastroenterology* 2006;130:1588–94.
73 Porter CK, et al. Infectious gastroenteritis and risk of developing inflammatory bowel disease. *Gastroenterology* 2008;135:781–6.
74 Borgaonkar MR, et al. A meta-analysis of antimycobacterial therapy for Crohn's diseaes. *Am J Gastroenterol* 2000;95:725–9.
75 Selby W, et al. Two-year combination antibiotic therapy with clarithromycin, rifabutin, and clofazimine for Crohn's disease. *Gastroenterology* 2007;132:2313–9.
76 Naser SA, et al. Culture of *Mycobacterium avium* subspecies *paratuberculosis* from the blood of patients with Crohn's disease. *Lancet* 2004;364:1039–44.
77 Frank DN, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U.S.A.* 2007;104:13780–5.
78 Sokol H, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U.S.A.* 2008;105:16731–6.
79 Sokol H, et al. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm Bowel Dis* 2009;15:1183–9.
80 Bernstein CN, et al. Disorders of a modern lifestyle: reconciling the epidemiology of inflammatory bowel diseases. *Gut* 2008;57:1185–91.
81 Peyrin-Biroulet L, et al. The natural history of Crohn's disease in population-based cohorts. *Am J Gastroenterol* 2010;105:289–97.
82 Peyrin-Biroulet L, et al. Long-term complications, extraintestinal manifestations, and mortality in adult Crohn's disease in population-based cohorts. *Inflamm Bowel Dis* 2011;17:471–8.
83 Munkholm P, et al. Disease activity courses in a regional cohort of Crohn's disease patients. *Scand J Gastroenterol* 1995;30:699–706.
84 Dhillon S, et al. The natural history of surgery for Crohn's disease in a population-based cohort from Olmsted County, Minnesota. *Am J Gastroenterol* 2005;100:S305 [abstract].
85 Munkholm P, et al. Intestinal cancer risk and mortality in patients with Crohn's disease. *Gastroenterology* 1993;105:1716–23.
86 Langholz E, et al. Course of ulcerative colitis: analysis of changes in disease activity over years. *Gastroenterology* 1994;107:3–11.
87 Dhillon S, et al. The natural history of surgery for ulcerative colitis in a population-based cohort from Olmsted County, Minnesota. *Am J Gastroenterol* 2005;100:S303 [abstract].

88 Winther KV, et al. Long-term risk of cancer in ulcerative colitis: a population-based cohort study from Copenhagen County. *Clin Gastroenterol Hepatol* 2004;2:1088–95.
89 Jess T, et al. Risk of intestinal cancer in inflammatory bowel disease: a population-based study from Olmsted County, Minnesota. *Gastroenterology* 2006;130:1039–46.
90 Loftus EV. Epidemiology and risk factors for colorectal dysplasia and cancer in ulcerative colitis. *Gastroenterol Clin North Am* 2006;35:517–31.
91 Jess T, et al. Intestinal and extra-intestinal cancer in Crohn's disease: follow-up of a population-based cohort in Copenhagen County, Denmark. *Aliment Pharmacol Ther* 2004;19:287–93.
92 Jess T, et al. Survival and cause specific mortality in patients with inflammatory bowel disease: a long-term outcome study in Olmsted County, Minnesota, 1940–2004. *Gut* 2006;55:1248–54.
93 Wolters FL, et al. Crohn's disease: increased mortality 10 years after diagnosis in a Europe-wide population based cohort. *Gut* 2006;55:510–18.
94 Winther KV, et al. Survival and cause-specific mortality in ulcerative colitis: follow-up of a population-based cohort in Copenhagen County. *Gastroenterology* 2003;125:1576–82.
95 Masala G, et al. Divergent patterns of total and cancer mortality in ulcerative colitis and Crohn's disease patients: the Florence IBD study 1978–2001. *Gut* 2004;53:1309–13.
96 Boonen A, et al. The impact of inflammatory bowel disease on labor force participation: Results of a population sampled case-control study. *Inflamm Bowel Dis* 2002;8:382–9.
97 Bernklev T, et al. Relationship between sick leave, unemployment, disability, and health-related quality of life in patients with inflammatory bowel disease. *Inflamm Bowel Dis* 2006;12:402–12.
98 Feagan BG, et al. Unemployment and disability in patients with moderately to severely active Crohn's disease. *J Clin Gastroenterol* 2005;39:390–5.
99 Longobardi T, et al. Utilization of health care resources by individuals with inflammatory bowel disease in the United States: a profile of time since diagnosis. *Am J Gastroenterol* 2004;99:650–5.
100 Longobardi T, et al. Health care resource utilization in inflammatory bowel disease. *Clin Gastroenterol Hepatol* 2006;4:731–43.
101 Bodger K. Cost of illness of Crohn's disease. *Pharmacoeconomics* 2002;20:639–52.
102 Blomqvist P, et al. Inflammatory bowel diseases: health care and costs in Sweden in 1994. *Scand J Gastroenterol* 1997;32:1134–9.

Answers to multiple choice questions

1. A
2. D
3. A

25 Epidemiology of fecal incontinence

Adil E. Bharucha
Division of Gastroenterology & Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

Key points

- Fecal incontinence is a common symptom among nursing home residents and also in the community, where the prevalence varies from 2.2 % to 15 %.
- Though most attention has focused on women, the prevalence of fecal incontinence in men is comparable to that in women.
- Fecal incontinence results from weakness of the pelvic floor muscles (i.e. the anal sphincters and/or the levator ani) and/or diarrhea. Diarrhea and rectal urgency are the strongest risk factors for fecal incontinence among women in the community, in whom the symptom typically begins several decades after vaginal delivery.
- Though the symptom significantly impacts quality of life and is associated with psychosocial distress, only a minority of patients will discuss the symptom with family members or a physician, partly due to embarrassment.

Introduction

Fecal incontinence (FI) is the involuntary loss of feces – solid or liquid. Anal incontinence includes involuntary loss of feces and flatus. While incontinence for flatus can be embarrassing, patients find it difficult to quantify flatus incontinence, and there is no cut-off to discriminate inadvertent expulsion of gas from incontinence. Most epidemiologic studies and the Rome criteria [1] are based on fecal rather than anal incontinence. Some epidemiologic studies on FI excluded leakage during short-term diarrheal illnesses (e.g. acute gastroenteritis) [2,3].

Our understanding of the epidemiology and pathophysiology of FI is predominantly derived from selected populations, (e.g. tertiary care centers), rather than community patients. These studies suggest that FI occurs in conditions associated with pelvic floor weakness and/or altered bowel habits, particularly diarrhea [4], and can impact nearly every aspect of daily life [5]. At the extreme, individuals with FI may withdraw from social contact and remain tethered to the toilet in an attempt to minimize incontinence [6]. FI may also contribute to institutionalization: up to 50 % of nursing home residents in one survey had FI [7]. Despite these potentially devastating consequences, it is unclear why only a small proportion of incontinent patients discuss the symptom with a physician [2,8,9]. Therefore, physicians tend to underestimate the personal impact of FI [6]. Moreover, results of clinic-based studies on FI cannot be extrapolated to the entire population, and community-based studies are essential to understand the risk factors, clinical spectrum, and personal impact of FI. Where possible, this chapter will focus on evidence derived from large population-based studies on the epidemiology of FI. See Table 25.1.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 25.1 Epidemiology of fecal incontinence: community-based studies

Survey	Respondents; Instrument	Response rate (number respondents)	Prevalence
Talley (1990) [52]	Olmsted County residents ≥65 years; Mailed questionnaire	66 % (328)	FI once per week over past year: 3.1 % (F); 4.5 % (M)
Drossman [8]	US householder marketing list; Mailed questionnaire	66 % (5430)	Soiling: 6.9 %(F); 7.4 %(M) Gross incontinence: 0.9 % (F); 0.5 % (M)
Nelson (1993) [10]	Wisconsin residents of all ages; Phone interview with *one member* in each household	73 % (6959)	Any FI over past year: 2.2 % (overall); 7.5 % (aged ≥65)
Reilly (1994) [3]	Olmsted County residents ≥50 years; Mailed questionnaire	64 % (1540)	Any FI: 17.8 % (F); 12.8 % (M)
Walter (2002) [53]	County of Ostergotland (Sweden); aged 31–76 years; Mailed questionnaire	81 % (1610)	Liquid FI >1 month^{-1}: 10.9 % (F); 9.7 % (M) Solid FI >1 month^{-1}: 1.4 % (F); 0.4 % (M)
Perry (2002) [24]	Leicestershire Health Authority (UK) patient register; Mailed questionnaire	70 % (10,226)	Any FI: 5.7 % (F); 6.2 % (M)
Bharucha (2005) [2]	Olmsted County residents ≥20 years; Mailed questionnaire	53 % (2800)	Any FI: 14 % (F)
Melville (2005) [26]	HMO population, 30–90 years, Washington State; Mailed questionnaire	64 % (3536)	Loss of liquid or stolid stool month^{-1}: 7.7 % (F)
Quander (2006) [11]	Chicago Health and Aging Project, ≥65 years, door-to-door survey (? 1 member of household)	79 % (6158)	
Varma (2006) [27]	Reproductive Risks for Incontinence Study at Kaiser, ≥40 years with ≥50 % of deliveries at Kaiser, mailed questionnaire	2109	Any FI in past 12 months: 25 % (F)
Nygaard (2008) [54]; Whitehead (2009) [12]	National Health and Nutrition Examination Survey (2005–2006), men and women ≥20 years, door-to-door interview	64 % (4308)	Loss of liquid or stolid or mucus stool during the last 30 days: 8.9 % (F); 7.7 % (M)

Prevalence rates for males (M) and females (F) are provided separately where available.
FI, fecal incontinence.

Methodological considerations

Survey techniques

Most studies on the epidemiology of FI have used a mailed questionnaire. Two large studies were conducted by telephone [8,10]. In the Chicago Health and Aging Project (CHAP) and National Health and Nutrition Examination Survey (NHANES) surveys, subjects were interviewed at home [11,12]. Patients with FI are reluctant, perhaps embarrassed to discuss the symptom [9] not only with physicians but also with family members and friends, perhaps explaining why the prevalence of FI was lower in surveys conducted by interviewing only one member of the household by phone (e.g. the Wisconsin survey) [10] compared to a mailed questionnaire [2,13].

Assessing severity of fecal incontinence and its impact on quality of life

There are several instruments for rating the severity of FI. However, these scales suffer from one or more limitations and there is no agreed threshold to identify clinically significant FI. Most scales for rating symptom severity in FI incorporate the frequency and type, but not the amount of leakage [14–17]. Without the latter, FI severity would be identical for two subjects, one of whom had minor staining and the other a large liquid bowel movement once a week. Second, only one questionnaire (i.e. the St. Marks severity rating system) incorporates urgency, assigning a score of 0 to 4 for patients who can or cannot defer defecation for 15 minutes, respectively [16]. It is important to incorporate rectal urgency in assessing the severity of FI because patients with urge FI and rectal hypersensitivity have more frequent stools, use more pads, and report more lifestyle restrictions compared to patients with normal rectal sensation [18]. However, this threshold (i.e. 15 minutes) for discriminating normal from excessive rectal urgency is relatively liberal, since clinical observations suggest a majority of incontinent patients are unable or reluctant to defer defecation for 15 minutes. Third, concerns have been raised about the weighting of variables in existing scales, which assume that different components (e.g. amount and frequency) are equally important in determining the severity of FI [17]. However, patients and colorectal surgeons disagree on the relative impact of different symptoms. For example, patients assigned a higher severity score to incontinence for flatus compared to physicians; conversely, physicians assigned a higher severity score for solid stool incontinence compared to patients. Finally, most symptom severity scales do not shed light on the impact of FI on quality of life (QOL). Thus, separate scales have been devised for assessing the impact of FI on QOL [19].

To overcome these limitations, we developed and validated a scale for rating symptom severity in FI that includes four components (i.e. frequency of FI, type of FI, amount of FI, and circumstances surrounding FI (i.e. urge or passive FI)) derived from a self-report questionnaire (the Fecal Incontinence and Constipation Assessment (FICA)) [20] (Table 25.2). The symptom-severity score was devised a priori to be user-friendly by assigning arbitrary weights (i.e. 0, 1, 2, and 3) for symptoms within each category (e.g. frequency of FI). Subjects who reported they often (i.e. >25 % of time) or usually (i.e. >75 %) experienced an "urgent need to empty their bowels" making them rush to the toilet were considered to have rectal urgency. Subjects who often (i.e. >25 % of time) or usually (i.e. >75 %) "leaked liquid or solid stool without any warning" were considered to have passive incontinence. Patients who did not report symptoms of urge or passive incontinence were classified as "neither," while those who had symptoms of urge and passive incontinence were classified as "combined" incontinence. In contrast to urinary incontinence [21], this FI symptom-severity scale was strongly correlated with a QOL-weighted symptom-severity score suggesting that the symptom-severity score, which is simple to use in the office, is a reasonable indication not only of the physical manifestations of FI (i.e. symptom severity), but also its impact on quality-of-life [22]. This strong correlation dispels the concern that measures of stool leakage may underestimate the severity of FI in people who avoid FI by staying close to a toilet (e.g. by staying at home) [23].

Perry et al. characterized the severity of FI as rare or no FI, minor FI, and major FI [24]. Those who leaked several times a year or less were characterized as rare incontinence regardless of the extent of soiling. Infrequent leakage was attributed to a coincident acute illness rather than a chronic condition. Minor incontinence was defined as staining of underwear several times a month or more often. Major FI was defined as soiling of underwear, outer clothing,

Table 25.2 Symptom-Severity Scale in Fecal Incontinence [19]

	Score			
Symptoms	1	2	3	4
Frequency	<1 month^{-1}	>1 month^{-1} to several times per week	Daily	
Composition	Mucus/Liquid stool	Solid stool	Liquid and solid stool	
Amount	Small (i.e. staining only)	Moderate (i.e. requiring change of underwear)	Large (i.e. requiring change of all clothes)	
Urgency or passive incontinence	Neither	Passive incontinence	Urge incontinence	Combined urge and passive incontinence

The symptom-severity score (maximum score = 13) is calculated by summing scores for individual components in this scale. Source: Rockwood et al. 2000 [19]. Reproduced with permission of Lippincott Williams & Wilkins.

furnishing, or bedding several times a month or more often. However, the reliability and validity of this simple and rational approach has not been evaluated.

Perineal protective devices

It is possible to quantify the use of devices worn to protect underclothes from FI by evaluating the type of device (i.e. panty liner, pad or diaper), the duration for which the device was worn (i.e. all the time, when awake away from home, when awake at home or when asleep), and the number of devices worn when awake (i.e. none, about 1 device per day, 2–4 devices per day, 5 or more devices per day) [2]. Because FI is associated with urinary incontinence, it is important to specify that devices worn only to protect against leakage of urine be excluded when responding to these items. Because the use of perineal protective devices may reflect coping strategies rather than severity of FI per se, this factor should not be used to gauge the severity of FI. For example, it is conceivable that fastidious people are more likely to use perineal protective devices even with mild FI.

Prevalence of fecal incontinence in the community

Nelson comprehensively reviewed epidemiologic studies in FI up to 2004 [25]. Only eight of 34 surveys in that review were community-based and sampled the entire population, that is, were unrestricted by age, residence, or underlying disease. However, four of these eight studies surveyed <750 subjects, and only two studies, conducted in a market mailing sample and Wisconsin households [8,10], were from the United States. Since that review, there have been five large studies on the epidemiology of FI [2,11,26,27].

The prevalence of FI in the population has varied among studies. Estimates range from 2.2 % in Wisconsin households and 7 % in a sample of US householders, to ~11–15 % in Australia, in Sweden, and in Olmsted County, Minnesota (Table 25.1). Different prevalence rates among studies probably reflect varying definitions of FI, differences in survey methods, and in the age distribution of the population surveyed. For example, the minimum duration of FI required to define cases was undefined in some studies, 1 month in the NHANES study, and 1 year in most Olmsted County studies [2,12]. While most attention has focused on FI in women, two studies suggest that the prevalence in men is comparable to women [12,24].

The prevalence of FI among nursing home residents (i.e. up to 50 %) is much higher compared to the general population [25]. Community-based studies demonstrate that the prevalence of FI increases with age. However, age-related trends in the prevalence of FI vary across studies. For example, in Leicestershire, UK, the prevalence increased steadily from ~4 % for

any incontinence in women aged between 40–49 years old to 7.8 % in subjects aged 70–79 years old and sharply thereafter to 11.6 % in women aged 80 years and older [24]. However, in Olmsted County and in the NHANES, the age-specific prevalence increased with age up to 22 % in the sixth decade and 15 % in the eighth decade, respectively [2,12].

Onset of fecal incontinence

Studies conducted in specific populations followed over relatively short periods of time who did not sustain an iatrogenic insult (e.g. anal sphincterectomy or obstetric injury) reported an onset rate of 7.5 % in institutionalized elderly people at 1 year [28]and 5.4 % at 1 year after rehabilitation from acute brain injury [29]. In a population-based study from Olmsted County, 15 % of the population over 50 years of age had FI at baseline and an additional 7 % developed FI during the following 10 years [30]. In a multivariate analysis which was limited by a relatively small sample size, rectal urgency was the only independent risk factor for development of FI. However, the natural history of women who had FI in the initial survey was not assessed.

Severity of fecal incontinence and its impact on quality of life

Severity of fecal incontinence

A majority of people with FI in the community have mild symptoms. In Olmsted County, most women with FI reported infrequent symptoms (55 % less than monthly), and most reported only staining of underwear (60 % of those with FI) [2]. Thus, 50 % of women had mild, 45 % had moderate, and 5 % had severe symptoms. In contrast to the prevalence of the condition, the severity of FI was not related to age.

Impact of fecal incontinence on quality of life

FI was associated with anxiety, depression, and physical disability in a community-based study of subjects aged >65 years from the United Kingdom [31]. In the Wisconsin Family Health Study, 33 % of subjects restricted their activities due to incontinence [10]. Studies from Leicestershire, UK and Olmsted County, indicate that FI impacts quality of life (QOL) in the community. In Leicestershire, 32 % of all subjects with FI and over 50 % of those reporting major FI (i.e. soiling of underwear, outer clothing, furnishing, or bedding several times a month or more often) reported that the symptom had "a lot of impact" on their QOL [24]. In Olmsted County, 23 % of women with FI reported that the symptom had a moderate to severe impact on one or more domains of QOL [2]. Moreover, the impact of FI on QOL was clearly related to symptom severity. Thus, 6 % of women with mild symptoms, 35 % of women with moderate symptoms, and 82 % of women with severe symptoms reported a moderate or severe impact on $\geq$1 domain of QOL. The proportion reporting moderate to severe impact for a given domain ranged from 3–4 % (e.g. for family relationships, employment, sex life) to 12 % (for the ability to eat outside home or going out to eat). However, differences in the impact of FI on specific domains of QOL were not significant. FI was also associated with worse health-related QOL at 6 months after vaginal delivery [32].

Risk factors for fecal incontinence

FI occurs in conditions associated with pelvic floor weakness and/or altered bowel habits, particularly diarrhea [4] (Table 25.3). Several epidemiologic studies have evaluated the multiple putative risk factors for FI by questionnaires (Table 25.4). In these studies, advancing age, diarrhea, rectal urgency, cholecystectomy, anal fistula, nonchildbirth anal injury, urinary incontinence, chronic illnesses (e.g. diabetes mellitus or stroke), and psychoactive medications were associated with FI [3,10,11,13].

We also used a questionnaire to assess individual risk factors and the interaction among risk factors (e.g. between risk factors for anal sphincter injury, rectal urgency, and bowel symptoms) for FI in Olmsted County [33]. The symptom of rectal urgency was the single most important risk factor for FI in women. The risk of FI was higher among women with rectal urgency whether or not they also had bowel disturbances (i.e. constipation, diarrhea or abdominal pain) (OR 8.3; 95 % CI 4.8–14.3) or had a vaginal delivery with forceps or stitches (OR 9.0; 95 % CI 5.6–14.4). Though rectal urgency was associated with

Table 25.3 Common causes of fecal incontinence

Anal sphincter weakness
Traumatic: obstetric, surgical (e.g. hemorrhoidectomy, internal sphincterotomy)
Nontraumatic: scleroderma, internal sphincter degeneration of unknown etiology
Neuropathy: peripheral (e.g. pudendal) or generalized (e.g. diabetes mellitus)
Disturbances of pelvic floor: rectal prolapse, descending perineum syndrome
Inflammatory conditions: radiation proctitis, Crohn's disease, ulcerative colitis
Central nervous system disorders: dementia, stroke, brain tumors, multiple sclerosis, spinal cord lesions
Diarrhea: irritable bowel syndrome, postcholecystectomy diarrhea
Other: fecal retention with overflow, behavioral disorders

loose stools as previously reported [34], this symptom was an independent, and much stronger risk factor for FI compared to loose stools (i.e. functional diarrhea), extending previous observations that in patients with functional bowel disorders, rectal urgency is not always associated with loose stools [34]. Indeed, the symptom of rectal urgency is associated with reduced rectal capacity and reduced rectal capacity is associated with rectal hypersensitivity among women with FI [35]. The NHANES (2005–2006) confirmed that age (OR 1.20; 95 % CI 1.10–1.31), loose or watery stools (OR 2.82; 95 % CI 1.95–4.08), more than 21 stools per week (OR 2.36; 95 % CI 1.09–5.12), multiple chronic illnesses (OR 2.20; 95 % CI 1.19–4.05), and urinary incontinence (OR 1.62; 95 % CI 0.99–2.66) were independent risk factors in women. In men, age (OR 1.24; 95 % CI 1.09–1.41), loose or watery stools (OR 4.76; 95 % CI 1.94–11.69), poor self-rated health (OR 1.78; 95 % CI 1.18–2.66), and urinary incontinence (OR 2.60; 95 % CI 1.44–4.67) were independent risk factors. The NHANES survey did not specifically ask about rectal urgency. Also, with one exception [33], these studies did not address the issue of obstetric trauma as a risk factor for FI.

However, vaginal delivery can damage the anal sphincters and the pudendal nerve, and up to 10 % of women develop FI after a vaginal delivery [4]. A review of the literature observed that the incidence of postpartum FI was considerably higher (i.e. 15–59 %) in women who sustained a third-degree (i.e. anal sphincter disruption) or a fourth-degree tear (i.e. a third-degree tear with anal epithelial disruption) [36]. Similarly, FI was more prevalent in women who delivered vaginally with (17 %) than those without (8.2 %) recognized anal sphincter tears, or women who delivered by cesarean section prior to labor (7.6 %) in the Childbirth and Pelvic Symptoms (CAPS) Study [37]. Moreover, the severity of FI and the prevalence of fecal urgency were also more pronounced in women who had anal sphincter tears. A prospective study demonstrated that anal sphincter defects and pudendal nerve injury after vaginal delivery were often clinically occult and that forceps delivery was the single independent factor associated with anal sphincter damage during vaginal delivery [38]. A systematic Cochrane review concluded that maternal morbidity was lower for assisted deliveries conducted with a vacuum extractor than with forceps [39]. Another Cochrane review concluded that restrictive episiotomy policies were beneficial (i.e. less posterior perineal trauma, less suturing, and fewer complications) compared to routine episiotomy policies [40]. However, there is an increased risk of anterior perineal trauma with restrictive episiotomy.

In contrast to urinary incontinence, the risk of FI was not significantly lower among women who had a cesarean section only compared to a vaginal delivery [25,41]. Further studies are necessary to clarify the risk of pelvic floor injury relative to the type of cesarean section (i.e. emergency or elective) because women who have an emergency cesarean section for stalled labor may not, in contrast to women who have an elective section, be protected against pelvic floor injury.

While obstetric anal sphincter injury can cause FI, FI typically begins 2–3 decades after vaginal delivery among unselected women in the community. For example, in one study, FI began before the age of 40 years in 31 %, between 41 and 60 years in 37 %, and between 61 and 80 years in 32 % [2], suggesting that obstetric pelvic floor injury is not the only risk factor for FI among women in the community. Only

Table 25.4 Risk factors for fecal incontinence (FI) in community-based studies

Survey	Risk factors significantly associated with FI	Risk factors not significantly associated with FI
Talley [52]	None	Age and gender
Drossman [8]	Employment (OR, 0.8; 95 % CI 0.6 −1.0).	Risks associated with other sociodemographic features (e.g. income) not specified
Nelson [10]	Age, male sex, poor general health, physical limitations	Race, marital status, employment status, educational level, launderer respondent
Reilly [3]	Urgency, pelvic radiation, and rectal/anal trauma	Unclear: published in abstract form only
Kalantar [13]*	Poor general health, perianal injury, perianal surgery, sense of incomplete evacuation, loose or watery motions, urgency	Radiation treatment to abdomen and pelvis (OR 2.7; 95 % CI 0.8–8.9), diabetes mellitus (OR 2.1; 95 % CI 0.7–6.3).
Bharucha [33]	Age, rectal urgency, prior anal surgery, history of anal fissure, cholecystectomy	Vaginal delivery with forceps/stitches alone (i.e. without bowel symptoms), hysterectomy (OR 1.3; 95 % CI 1.0–1.7), contraceptive use (OR 1.4; 95 % CI 1.0–1.9)
Melville [26]	Age, major depression, urinary incontinence, medical comorbidity, operative vaginal delivery	Body mass index, h/o cesarean delivery only, nulliparity
Quander [11]	Age, low income and education, diabetes, stroke, certain medications	Gender, certain medications
Varma [27]	Obesity, chronic obstructive pulmonary disease, IBS, urinary incontinence, colectomy	Age, diabetes mellitus, parity, pelvic organ prolapse surgery, hysterectomy, cholecystectomy
Whitehead (2009) [12]	In women: age, loose or watery stools, >21 stools week^{-1}, multiple chronic illnesses, and urinary incontinenceIn men: age, loose or watery stools, poor self-rated health, and urinary incontinence	Race/ethnicity, education, income, or marital status after adjusting age
Bharucha 2010 [45]	Current smoking, BMI, diarrhea, IBS, cholecystectomy, rectocoele and stress urinary incontinence	Obstetric events

Odds ratios are specified when the mean risk factor is >1.0 but the lower bound of the 95 % CI is ≤1.0.
*Multiple variable analysis not performed.

three truly population-based studies have evaluated the relationship between obstetric events and FI, and that was also by questionnaires; operative vaginal deliveries were [26] or were not [33,42] risk factors for FI. However, maternal recall of distant pregnancy events is variable. Recall is excellent for certain items (e.g. cesarean section) but weaker for other items (e.g. induced labor or problems during delivery) [43]. Moreover a State-of-the-Science Conference in Prevention of Fecal and Urinary Incontinence in Adults highlighted the limitation that most studies of fecal and urinary incontinence used a cross-sectional design. "Such studies let us examine associations with incontinence but not cause. We cannot be sure that the associated factor comes before the recurrence of incontinence or determine whether

it is the cause of the incontinence and therefore whether changing the associated factor would reduce to eliminate the incontinence" [44]. To address these limitations, we conducted a nested case-control study of 176 randomly selected women with FI (cases; mean age, 58 y) and 176 age-matched community controls in a population-based cohort from Olmsted County [45]. Risk factors for FI were evaluated by reviewing inpatient and outpatient medical (including original obstetric) records rather than by questionnaires. Analyses focused on conditions that preceded the incidence date of FI for case in each matched pair. In 88 % of cases, FI began at age ≥40 y; severity was mild (37 %), moderate (58 %) or severe (5 %). By multivariable analysis, current smoking (OR 4.7; 95 % CI 1.4–15), body mass index (OR per unit, 1.1; 95 % CI 1.004–1.1), diarrhea (OR 53; 95 % CI 6.1–471), IBS (OR 4.8; 95 % CI 1.6–14), cholecystectomy (OR 4.2; 95 % CI 1.2–15), rectocoele (OR 4.9; 95 % CI 1.3–19), and stress urinary incontinence (OR 3.1; 95 % CI 1.4–6.5), but not obstetric events, were independent risk factors for FI. Taken together, these findings demonstrate that bowel disturbances rather than obstetric events are a primary determinant of late-onset FI and suggest that, similar to urinary incontinence [46], obstetric trauma (e.g. forceps use) is a stronger risk factor for postpartum FI [38] than for delayed onset FI [47]. Hence, measures to ameliorate bowel disturbances and other potentially reversible risk factors should be implemented before anal imaging in women with FI.

In addition to obstetric trauma, other risk factors for FI are also influenced by the age distribution of the population. In a population aged 65 years and older, self-reported diabetes mellitus (OR 1.7; 95 % CI 1.4–2.1), self-reported stroke (OR 2.8; 95 % CI 2.2–3.5), and certain medications were also risk factors for FI after adjusting for age, sex, and race [11]. It is unclear if FI preceded or followed diabetes mellitus or stroke. Because other medical conditions and other putative risk factors for FI were not assessed, it is unclear if the increased risk was attributable to diabetes mellitus or stroke, or if these conditions were merely markers for other risk factors. In the same study, anti-Parkinsonian, hypnotic, and antipsychotic medications were also associated with a three- to fourfold increased risk for FI even after adjusting for age, sex, race, stroke, and diabetes. On the other hand, calcium channel blockers decreased the risk of FI while estrogens, diuretics, antacids, β-blockers, and benzodiazepines did not affect the risk of FI.

FI is well documented to occur even after "minor" operations (e.g. lateral internal anal sphincterotomy) [4]. In Olmsted County, prior anal surgery, a history of anal inflammation (e.g. abscess, fistula), and a cholecystectomy increased the risk for FI [25].

Health-seeking for fecal incontinence

In one study, only 10 % of women with FI had discussed the symptom with a physician in the preceding year [2]. Although this estimate may not include subjects who had discussed the symptom with a physician at an earlier time, it confirms other studies in which only ~20–25 % of subjects with FI or IBS had discussed the symptom with a physician [8,48]. However 48 % of women with severe FI had consulted a physician for the symptom. In addition to symptom severity, general health status also independently predicted physician consulting behavior for FI. Taken together, these factors explained 15 % of the variance in consulting behavior, which is similar compared to previous population-based studies in IBS that have addressed this issue [48,49].

Impact of fecal incontinence on institutionalization and mortality

The contribution of FI to institutionalization was recently assessed in a 10-year follow-up of 9008 community-dwelling participants in the Canadian Study of Health and Aging aged 65 years and older [50]. Among subjects who had FI at baseline (i.e. 4 %), mortality was higher (hazard ratio 1.19; 95 % CI 1.00–1.41) even after adjusting for age, sex, cognition, and functional independence but not after adjusting for self-reported health. This suggests that FI increased mortality by virtue of its association with poor health status rather than independently. Moreover, while individuals with FI had an increased risk of institutionalization (OR 1.79; 95 % CI 1.00–3.20) independent of age and sex, this association was not significant after adjusting for cognition, ADL dependence, and self-reported health. Hence, these findings suggest that while FI is often cited as a cause for institutionalization, this risk is largely explained

by other factors, particularly cognitive and functional impairment.

Summary and a look to the future

Population-based studies in FI are important because they (i) avoid the bias accompanying studies on the epidemiology of FI in selected populations, (ii) underscore that the symptom is common not only in nursing homes but also in the community, (iii) quantify the impact of FI on quality of life, and (iv) demonstrate that the symptom generally begins 2–3 decades after the initial insult to the pelvic floor, that is, vaginal delivery. While most studies have focused on women, there are limited data to suggest that the prevalence of FI is comparable in men and in women. Epidemiologic studies have also provided insights into the etiology of FI. These findings are particularly important because as the population ages, the number of women with FI in the United States is anticipated to increase by 59 % from 10.6 to 16.8 million between 2010 and 2050 [51]. Further studies are necessary to define the relationship between obstetric history, pelvic floor injury and FI, to evaluate the incidence and natural history of FI, and to explore the factors that influence health-seeking behavior in FI.

Acknowledgment

This work was supported in part by Grant R01 HD41129 from the National Institutes of Health, U.S. Public Health Service.

Multiple choice questions

1 Which of the following components are critical for assessing the severity of fecal incontinence?

A Type of leakage (i.e. solid or liquid stool or both)
B Rectal urgency
C Quantity of stool leakage
D Frequency of leakage
E All the above

2 With respect to fecal incontinence in the community, which of the following statements is not accurate?

A The prevalence across most studies ranges from 2–15 %
B The prevalence is much higher in women than in men
C The prevalence is higher among nursing home residents than in the community
D Severity of fecal incontinence is correlated with its impact on quality of life

3 Which of the following statements regarding risk factors for fecal incontinence among women in the community is not accurate?

A Rectal urgency is an independent and important risk factor, even after adjusting for diarrhea
B Obstetric history is an independent risk factor in women who develop fecal incontinence in the seventh decade or after
C BMI
D Current smoking

References

1 Bharucha AE, Wald A, Enck P, Rao S. Functional anorectal disorders. *Gastroenterology* 2006;130:1510–18.
2 Bharucha AE, Zinsmeister AR, Locke GR, et al. Prevalence and burden of fecal incontinence: A population-based study in women. *Gastroenterology* 2005;129: 42–9.
3 Reilly W, Talley N, Pemberton J. Fecal incontinence: prevalence and risk factors in the community. *Gastroenterology* 1995;108:A32.
4 Bharucha A. Fecal incontinence. *Gastroenterology* 2003;124:1672–85.
5 Norton NJ. The perspective of the patient. *Gastroenterology* 2004;126:S175–9.
6 Miner PB, Jr. Economic and personal impact of fecal and urinary incontinence. *Gastroenterology* 2004;126: S8–13.
7 Nelson R, Furner S, Jesudason V. Fecal incontinence in Wisconsin nursing homes: prevalence and associations. *Dis Colon Rectum* 1998;41:1226–9.
8 Drossman DA, Li Z, Andruzzi E, et al. U.S. householder survey of functional gastrointestinal disorders. Prevalence, sociodemography, and health impact. *Digest Dis Sci* 1993;38:1569–80.
9 Leigh RJ, Turnberg LA. Faecal incontinence: the unvoiced symptom. *Lancet* 1982;1:1349–51.
10 Nelson R, Norton N, Cautley E, Furner S. Community-based prevalence of anal incontinence. *JAMA* 1995;274:559–61.
11 Quander CR, Morris MC, Melson J, et al. Prevalence of and factors associated with fecal incontinence in a large community study of older individuals. *Am J Gastroenterol* 2005;100:905–9.

12 Whitehead WE, Borrud L, Goode PS, et al.; Network PFD. Fecal incontinence in US adults: epidemiology and risk factors. *Gastroenterology* 2009;137:512–17.
13 Kalantar JS, Howell S, Talley NJ. Prevalence of faecal incontinence and associated risk factors; an underdiagnosed problem in the Australian community? *Med J Australia* 2002;176:54–7.
14 Jorge JM, Wexner SD. Etiology and management of fecal incontinence. *Dis Colon Rectum* 1993;36:77–97.
15 Pescatori M, Anastasio G, Bottini C, Mentasti A. New grading and scoring for anal incontinence. Evaluation of 335 patients. *Dis Colon Rectum* 1992;35:482–7.
16 Vaizey CJ, Carapeti E, Cahill JA, Kamm MA. Prospective comparison of faecal incontinence grading systems. *Gut* 1999;44:77–80.
17 Rockwood TH, Church JM, Fleshman JW, et al. Patient and surgeon ranking of the severity of symptoms associated with fecal incontinence: the fecal incontinence severity index. *Dis Colon Rectum* 1999;42:1525–32.
18 Chan CL, Scott SM, Williams NS, Lunniss PJ. Rectal hypersensitivity worsens stool frequency, urgency, and lifestyle in patients with urge fecal incontinence. *Dis Colon Rectum* 2005;48:134–40.
19 Rockwood TH, Church JM, Fleshman JW, et al. Fecal incontinence quality of life scale: quality of life instrument for patients with fecal incontinence. *Dis Colon Rectum* 2000;43:9–16; discussion 16–17.
20 Bharucha AE, Locke GR, Seide B, Zinsmeister AR. A new questionnaire for constipation and fecal incontinence. *Aliment Pharmacol Therap* 2004;20:355–64.
21 Naughton MJ, Donovan J, Badia X, et al. Symptom severity and QOL scales for urinary incontinence. *Gastroenterology* 2004;126:S114–23.
22 Bharucha AE, Zinsmeister AR, Locke GR, et al. Symptoms and quality of life in community women with fecal incontinence. *Clin Gastroenterol Hepatol* 2006;4(8): 1004–9.
23 Rockwood TH. Incontinence severity and QOL scales for fecal incontinence. *Gastroenterology* 2004;126:S106–13.
24 Perry S, Shaw C, McGrother C, et al. Prevalence of faecal incontinence in adults aged 40 years or more living in the community. *Gut* 2002;50:480–4.
25 Nelson RL. Epidemiology of fecal incontinence. *Gastroenterology* 2004;126:S3–7.
26 Melville JL, Fan MY, Newton K, Fenner D. Fecal incontinence in US women: a population-based study. *Am J Obstet Gynecol* 2005;193:2071–6.
27 Varma MG, Brown JS, Creasman JM, et al.; Reproductive Risks for Incontinence Study at Kaiser Research Group. Fecal incontinence in females older than aged 40 years: who is at risk? *Dis Colon Rectum* 2006;49:841–51.
28 Chassagne P, Landrin I, Neveu C, et al. Fecal incontinence in the institutionalized elderly: incidence, risk factors, and prognosis. *Am J Med* 1999;106:185–90.
29 Foxx-Orenstein A, Kolakowsky-Hayner S, Marwitz JH, et al. Incidence, risk factors, and outcomes of fecal incontinence after acute brain injury: findings from the Traumatic Brain Injury Model Systems national database. *Arch Phys Med Rehab* 2003;84:231–7.
30 Rey E, Choung RS, Schleck CD, et al. Onset and risk factors for fecal incontinence in a US community. *Am J Gastroenterol* 2010;105:412–9.
31 Edwards NI, Jones D. The prevalence of faecal incontinence in older people living at home. *Age Ageing* 2001;30:503–7.
32 Handa VL, Zyczynski HM, Burgio KL, et al.; Pelvic Floor Disorders Network. The impact of fecal and urinary incontinence on quality of life 6 months after childbirth. *Am J Obstet Gynecol* 2007;197:636.e1–6.
33 Bharucha AE, Zinsmeister AR, Locke GR, et al. Risk factors for fecal incontinence: a population-based study in women. *Am J Gastroenterol* 2006;101:1305–12.
34 Heaton KW, Ghosh S, Braddon FE. How bad are the symptoms and bowel dysfunction of patients with the irritable bowel syndrome? A prospective, controlled study with emphasis on stool form. *Gut* 1991;32:73–9.
35 Bharucha AE, Fletcher JG, Harper CM, et al. Relationship between symptoms and disordered continence mechanisms in women with idiopathic fecal incontinence. *Gut* 2005;54:546–55.
36 Mostwin J, Bourcier A, Haab F, et al. (2005) Pathophysiology of urinary incontinence, fecal incontinence and pelvic organ prolapse, in *Incontinence*, Vol. 1 (Eds. P Abrams, L Cardozo, S Khoury, A Wein), Health Publications Ltd, Paris, pp. 425–84.
37 Borello-France D, Burgio KL, Richter HE, et al.; Pelvic Floor Disorders Network. Fecal and urinary incontinence in primiparous women. *Obstetrics & Gynecology* 2006;108:863–72.
38 Sultan AH, Kamm MA, Hudson CN, et al. Anal-sphincter disruption during vaginal delivery. *New Engl J Med* 1993;329:1905–11.
39 Johanson RB, Menon V. Vacuum extraction versus forceps for assisted vaginal delivery [systematic review]. *Cochrane Database Syst Rev* 2010, 11. Art. No.: CD000224. DOI: 10.1002/14651858.CD000224.pub2.
40 Carroli G, Mignini L. Episiotomy for vaginal birth. *Cochrane Database Syst Rev* 2009, 1. Art. No.: CD000081. DOI: 10.1002/14651858.CD000081.pub2.
41 Rortveit G, Daltveit AK, Hannestad YS, et al. Urinary incontinence after vaginal delivery or cesarean section. *New Engl J Med* 2003;348:900–7.
42 Fritel X, Ringa V, Varnoux N, et al. Mode of delivery and fecal incontinence at midlife: a study of 2,640

women in the Gazel cohort. *Obstetrics & Gynecology* 2007;110:31–8.
43 Yawn BP, Suman VJ, Jacobsen SJ. Maternal recall of distant pregnancy events. *J Clin Epidemiol* 1998;51:399–405.
44 Landefeld CS, Bowers BJ, Feld AD, et al. National Institutes of Health State-of-the-Science Conference Statement: Prevention of fecal and urinary incontinence in adults. *Ann Intern Med* 2008;148:449–58.
45 Bharucha AE, Zinsmeister AR, Schleck CD, Melton LJ, 3rd. Bowel disturbances are the most important risk factors for late onset fecal incontinence: a population-based case-control study in women. *Gastroenterology* 2010;139:1559–66.
46 Rortveit G, Hannestad YS, Daltveit AK, Hunskaar S. Age- and type-dependent effects of parity on urinary incontinence: the Norwegian EPINCONT study. *Obstetrics & Gynecology* 2001;98:1004–10.
47 Bollard RC, Gardiner A, Duthie GS, Lindow SW. Anal sphincter injury, fecal and urinary incontinence: a 34-year follow-up after forceps delivery. *Dis Colon Rectum* 2003;46:1083–8.
48 Talley NJ, Zinsmeister AR, Van Dyke C, Melton LJ, 3rd. Epidemiology of colonic symptoms and the irritable bowel syndrome. *Gastroenterology* 1991;101:927–34.
49 Koloski NA, Talley NJ, Huskic SS, Boyce PM. Predictors of conventional and alternative health care seeking for irritable bowel syndrome and functional dyspepsia. *Aliment Pharmacol Ther* 2003;17:841–51.
50 AlAmeel T, Andrew MK, MacKnight C. The association of fecal incontinence with institutionalization and mortality in older adults. *Am J Gastroenterol* 2010;105:1830–4.
51 Wu JM, Hundley AF, Fulton RG, Myers ER. Forecasting the prevalence of pelvic floor disorders in U.S. women: 2010 to 2050. *Obstetrics & Gynecology* 2009;114:1278–83.
52 Talley NJ, O'Keefe EA, Zinsmeister AR, Melton LJ. Prevalence of gastrointestinal symptoms in the elderly: a population-based study. *Gastroenterology* 1992;102:895–901.
53 Walter S, Hallbook O, Gotthard R, et al. A population-based study on bowel habits in a Swedish community: prevalence of faecal incontinence and constipation. *Scand J Gastroenterol* 2002;37:911–6.
54 Nygaard I, Barber MD, Burgio KL, et al.; Pelvic Floor Disorders Network. Prevalence of symptomatic pelvic floor disorders in US women. *JAMA* 2008;300:1311–6.

Answers to multiple choice questions

1. E
2. B
3. B

26 Epidemiology of gallstones and biliary tract cancers

Guy D. Eslick[1] *& Eldon A. Shaffer*[2]

[1]The Whiteley-Martin Research Centre, Discipline of Surgery, The University of Sydney, Sydney, Australia

[2]Division of Gastroenterology, University of Calgary, Calgary, AB, Canada

Key points

- For gallstone disease, ethnic differences vary from a high of 60–70 % in American Indians to prevalences of 20 % in northern Europe and 6–17 % overall in Europe and North American white adults. Very lowest rates occur in sub-Saharan black Africans.
- The risks for developing gallstones include both genetic and environmental factors. Modifiable factors such as diet, activity, rapid weight loss and obesity carry the potential for primary prevention. Immutable factors like age, female gender, genetics, and ethnicity cannot be altered.
- The risk factors for developing gallbladder cancer include ethnicity, gender, age, a family history, lifestyle, gallstones, chronic inflammation, and certain bacterial infections.
- Cholangiocarcinomas are uncommon in developed countries except in a setting of primary sclerosing cholangitis. In Asia, cholangiocarcinomas are more frequent, being associated with liver fluke infestations, hepatolithiasis, and chronic viral hepatitis.

Gallstone disease

Epidemiology

Gallstone disease is very common throughout the world and constitutes a major health burden in developed societies. The best epidemiology studies use ultrasonography to screen for gallstones in large populations. In the United States 10–15 % of the adult population (20–25 million Americans) will someday harbor gallstones [1], yet only 20 % ever develop symptoms [2]. Gallstone disease is a leading cause of hospital admissions for gastrointestinal problems [3], while the overall all-cause mortality for those with gallstone disease has increased compared with the population (hazard ratio (HR) 1.30; 95 % CI 1.10–1.50) [4]. The resultant direct and indirect cost of gallbladder disease consumes ~$6.2 billion annually in the United States, constituting a major health burden that has increased more than 20 % over the last three decades. In Europe, the prevalence of gallstone disease is similar, somewhere between 6 % and 22 % [5]. Since the introduction of laparoscopic cholecystectomy, rates of surgery have increased on both continents [6,7].

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 26.1 Characteristics of the types of stones that are associated with gallstone disease

	Cholesterol stones	Black pigment stones	Brown pigment stones
Childhood	Becoming more common	Rare	Rare
Ethnicity	Developed countries	Developed countries	Asia
Composition	50–100 % cholesterol	Calcium bilirubinate polymer	Unconjugated bilirubin, calcium soaps (palmitate, stearate), cholesterol and mucin
Size	0.3–3 cm	0.3–0.6 cm	0.5–1.0 cm
Radiodensity	Lucent (10 % opaque)	50 % opaque	Lucent
Low-protein diet	No	No	Yes
Hemolysis/cirrhosis	No	Yes	No
Color	Yellow-brown	Black	Brown
Consistency	Crystalline	Hard	Soft, greasy
Location	Gallbladder ± common duct (~10 %)	Gallbladder ± common duct	Bile ducts
Radiodensity	Lucent (85 %)	Opaque (>50 %)	Lucent (100 %)
Recurrence		Uncommon	Frequent
Clinical associations	Metabolic: family history (genetic traits), obesity, female sex, aging [excessive cholesterol secretion]	Increased red cell destruction (hemolysis), cirrhosis, cystic fibrosis, Crohn's disease, advanced age [excessive bilirubin excretion]	Infection, inflammation, infestation [stasis, strictures]

North American Indians have the highest prevalence of gallstone disease with 64 % of females and 30 % of males being affected [8]. Among Pima Indian females over the age of 30, the rates are as high as 73 % [9]. Other Aboriginal populations also exhibit extraordinary frequencies of gallstone disease, including the native Mapuche Indians of Chile: affecting 50 % of women and 13 % of men [10]. Conversely, the lowest prevalence rates are in sub-Saharan black Africans (less than 5 %); gallstones are almost nonexistent in the Masi and the Bantu [11].

Ethnicity is a key factor determining the type of stones that form, why they develop and where they reside in the biliary system [12] (Table 26.1). The majority of stones in developed countries arise in the gallbladder. Gallstones consist predominantly of cholesterol (>85 %), whereas the remainder constitute black pigment stones (i.e. composed of calcium bilirubinate). In East Asia, brown pigment stones form in bile ducts (choledocholithiasis) or hepatic ducts (hepatolithiasis), secondary to stasis and inflammation, largely due to parasitic infestation causing partial biliary obstruction. Stone type recently has shifted in developing Asian countries from brown pigment to cholesterol stones, likely from improved public health and the adoption of Western dietary habits.

Risk factors

The numerous risk factors associated with the development of gallstones in humans include both genetic and environmental aspects. Modifiable factors such as diet, activity, rapid weight loss, and obesity carry the potential for primary prevention. Immutable factors like age, female gender, genetics, and ethnicity cannot be altered.

Age

Gallstone disease historically affects those over the age of 40 years, earlier in populations at high risk like the Pima Indians. With advancing age, more gallstones are composed of black pigment material. Gallstones once rare in childhood except for associated conditions like chronic hemolysis are increasing, perhaps related to the epidemic of obesity predisposing to cholesterol stone formation [13].

Female gender and sex hormones

This important element for gallstone formation places women at high risk [1]. Women are twice as likely as men to form stones; this gap narrows following menopause after which men begin to catch up though older age increases the risk substantially in both. Additional risk factors for the development of gallstones among females include the use of oral contraceptives and estrogen replacement therapy [14].

During pregnancy, biliary sludge (particulate material that is composed of cholesterol, calcium bilirubinate, and mucin) appears in up to one-third of women. Resolution frequently transpires during the postpartum period: sludge and small (<1 cm) stones (microlithiasis) vanish in most, but definitive gallstones become established in as many as 5 %.

Obesity

Obesity (body mass index (BMI) $\geq$30 kg m^{-2}) is a well-established risk factor for development of gallstones and the risk is elevated in females, with those who are extremely obese (BMI >35) having a sixfold increased risk for developing gallstones [15]. Those who are obese at an earlier stage in life, particularly in their teenage years, are at the greatest risk of developing gallstones compared with those who are underweight, which appears to be protective [16].

Metabolic syndrome

Cholesterol gallstone formation and stone complications correlate with the metabolic syndrome: a combination of abdominal obesity, type 2 diabetes mellitus, and hypertriglyceridemia, the common denominator being insulin resistance [17]. Curiously, the development of cholesterol stones in the gallbladder is not associated with hypercholesterolemia; rather, low HDL increases the risk of developing stones. Any relationship between diabetes and gallstones is unclear, being confounded by age, obesity, and a family history of gallstones.

Rapid weight loss

Substantial weight loss (>1.5 kg $week^{-1}$) following bariatric surgery or low-calorie dieting is associated with the development of gallstones in 30–70 % of individuals. Gallstones may be identified in the early stages following bariatric surgery, when the rapidity of weight loss is most extreme. Only a small proportion, however, experience symptoms (<20 %) [16,18].

Diet

Dietary factors are complex and not easy to assess in terms of gallstone formation [19]. Diets high in cholesterol, fatty acids, carbohydrates, and/or legumes appear to increase the risk of gallstone development, whereas, the consumption of unsaturated fats, coffee, fiber, ascorbic acid (vitamin C), calcium, and moderate consumption of alcohol reduce the risk [1]. The shift to a more Western diet, high in refined carbohydrates and fat (triglycerides) and low in fiber, best explains the increase in cholesterol gallstones amongst American Indians (unmasking their presumed genetic burden), in Europe following World War II, and more recently in some developing Asian countries.

Socioeconomic status

The relationship between socioeconomic status and the development of gallstones is unclear. Socioeconomic status most likely represents a proxy marker for another factor.

Lifestyle factors

These are important modifiable risk factors, with increased physical activity reducing the risk of gallstone development and decreased physical activity increasing this menace [20]. Currently, the role of smoking in the development of gallstones is unclear.

Genetics and family history

The importance of genetics in the development of human gallstones is a complex issue due to the multifactorial nature of gallstone disease, which makes identifying genetic defects problematic [21]. Identification of *Lith* genes (*Lith 1*, *Lith 2*) in mouse models provides important information regarding genetic susceptibility, but is obscured by the complex interactions that humans experience with their environment. Familial and twin studies reveal that gallstone disease can cluster among relatives or be increased in monozygotic compared to dizygotic twins [22,23]. This genetic component accounts for about 25 % of the overall effect of the gallstone disease phenotype. The remainder is unique or exhibits shared environmental factors, accounting for 62 % and 13 % variances, respectively.

Underlying chronic diseases

Liver disease

Advanced cirrhosis is a well-established risk factor for gallstones with a prevalence of 25–30 % [24,25]. The majority of stones are the black pigment type due to a combination of altered bilirubin metabolism and abnormal gallbladder motility [26]. Hepatitis C also is associated with gallstone disease [27] along with non-alcoholic fatty liver disease (NAFLD), the connection being the metabolic syndrome and obesity [28].

Crohn's disease

A two- to threefold increased frequency of black pigment stones occurs in patients with ileal Crohn's disease [29]. The mechanism results from unabsorbed bile acids which enter the colon and operate as a biologic detergent to solubilize bilirubin, enhance its absorption and enterohepatic cycling, and thus increase pigment excretion, leading to stone formation [30].

Cystic fibrosis

The gallstone prevalence in cystic fibrosis is increased: 10–30 % versus <5 % in age-matched controls, the result of bile acid malabsorption [31].

Other diseases

A recent meta-analysis found that gallstone disease is associated with an increased risk of rectal cancer [32], colonic adenoma [33], and colon cancer. Irritable bowel syndrome (IBS) has been linked to gallstone disease and cholecystectomy; however, a recent large population-based study has found that while individuals with IBS have higher rates of cholecystectomy, it is not due to an increased risk of gallstones, rather a consequence of having abdominal pain [34].

Spinal cord injury is associated with a threefold increase in gallstones, presumably related to gallbladder stasis (i.e. forming sludge) and intestinal hypomotility (by augmenting secondary bile acids, like deoxycholic acid, which then adversely influence bile formation) [35].

Biliary tract cancers

Biliary tract cancers are best divided into malignancies of the gallbladder, the extrahepatic bile ducts, and the ampulla of Vater, whereas intrahepatic tumors are deemed primary liver cancers. Cholangiocarcinomas refer to bile duct cancers that arise in the intrahepatic, perihilar, or distal (extrahepatic) biliary tree, exclusive of the gallbladder or ampulla of Vater.

Gallbladder cancer

Epidemiology

Gallbladder cancer is a rare malignancy. Many patients are asymptomatic [36] often with advanced metastatic disease and hence, a dismal prognosis. Adenocarcinoma of the gallbladder exhibits great diversity both geographically and by ethnicity [37]. Prevention measures may be possible for this enigmatic malignancy but necessitate further research to reduce its mortality [38].

Geographic differences abound (Figure 26.1, Figure 26.2). The extraordinary frequency of gallstones in South America correlates with the high rates of gallbladder cancer in Chile for both genders. For males, the overall incidence of gallbladder cancer is highest in two Asian countries – Korea and Japan, globally ranked within the top10 incident countries. Females differ geographically; high incidence countries include Algeria, India, Korea, Peru, and Ecuador.

In the United States, the driving force behind the variations in the incidence of gallbladder cancer relate to ethnicity. High incidence areas for males include Los Angeles (Korean and Japanese), New Mexico (American Indian), Hawaii (Hawaiian), and Connecticut (African American). A similar situation exists for females with high incident areas for Los Angeles (Hispanic Whites, Koreans), and San Francisco (Hispanic Whites). There also have been changes over time for some ethnic groups like female Asian/Pacific Islanders and male Hispanics, but no real change for African Americans or non-Hispanic white females.

Mortality

The highest incidence countries also have the highest mortality for gallbladder cancer. In Chile this involves the Mapuche Indians and Hispanics, while in the United States the American Indians and Mexican Americans carry this burden [39]. There appears to be a decline in mortality related to gallbladder cancer in Europe, Canada, USA, and the United Kingdom,

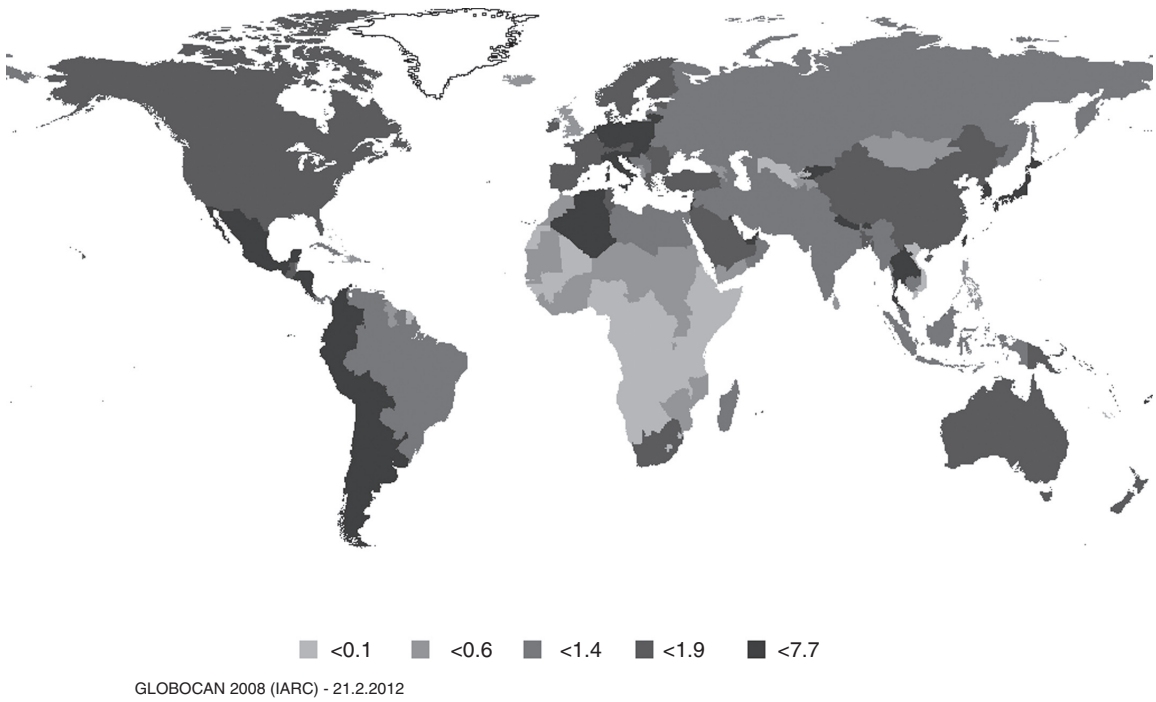


Figure 26.1 World standardized incidence rates for gallbladder cancer for males. Source: Ferlay et al. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet], International Agency for Research on Cancer, Lyon, France; 2010. Available from: http://globocan.iarc.fr.

whereas some countries like Iceland, Costa Rica, and Korea have experienced increases in mortality [40].

Risk factors

The risk factors for developing gallbladder cancer include ethnicity, gender, age, lifestyle, gallstones, chronic inflammation, and infections such as *Salmonella typhi* [41], *Helicobacter bilus*, and *Helicobacter hepaticus* [42]. Genetic susceptibility is likely with the risk being increased in those with a family history, but the responsible gene(s) are unknown at this time.

Age

Gallbladder cancer increases with age, traditionally affecting those over 65 years of age [43]. Recent analysis has found a substantial decrease in the age of those now being affected; those under 50 years of age are now developing gallbladder cancer [44]. The reasons for this change remain unknown.

Gender and sex hormones

Studies highlight females generally being at a substantially greater risk of developing gallbladder cancer compared to males. There is conflicting evidence linking female reproductive factors and the development of gallbladder cancer. Some reports suggest that parity, gravidity, oral contraceptive use, and hormone replacement therapy (HRT) may be risk factors, but other studies show no association [45].

Gallbladder disease

The relationship between gallstone disease and the development of gallbladder cancer is close but not absolute. Although gallstone disease increases the risk of cancer, not all individuals with gallstones will progress to develop gallbladder cancer [46].

Obesity

A meta-analysis exploring the relation between obesity and gallbladder cancer found that increasing BMI

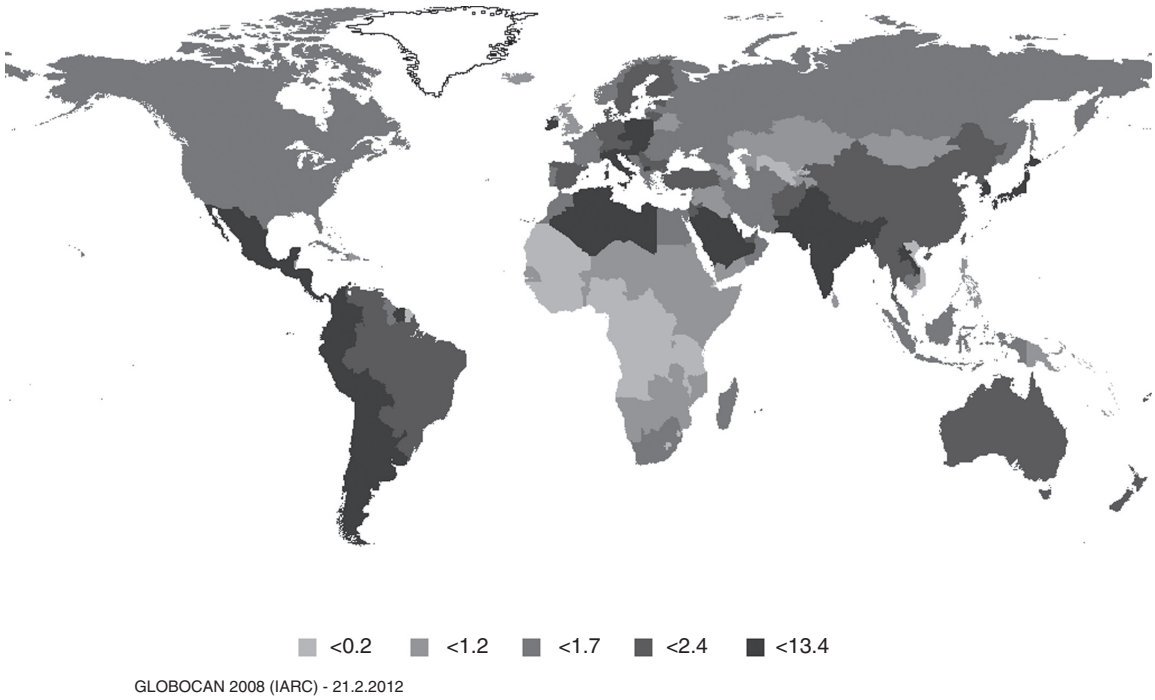


Figure 26.2 World standardized incidence rates for gallbladder cancer for females. Source: Ferlay et al. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet], International Agency for Research on Cancer, Lyon, France; 2010. Available from: http://globocan.iarc.fr.

is associated with an increased risk of gallbladder cancer, with females having a slightly higher risk (RR 1.88; 95 % CI 1.66–2.13) than males (RR 1.35; 95 % CI 1.09–1.68) [47].

Infection

Several bacteria are associated with gallbladder cancer. These include *Salmonella typhi*, *Salmonella paratyphi*, and the Helicobacter species: *Helicobacter bilis*, *Helicobacter hepaticus*, and *Helicobacter pylori* [42]. Currently, there is no evidence implicating viruses, parasites, yeasts, or fungi.

Cholangiocarcinoma

Epidemiology

Cholangiocarcinomas are not common, accounting for only ~3 % of all gastrointestinal malignancies, yielding an incidence of 1–2 cases per 100,000 population in developed countries like the United States. Due to the rare nature and geographic variation of this cancer, there is inconsistent data on its natural history and epidemiology.

The incidence of cholangiocarcinoma is highest in regions of northeastern Thailand where liver flukes are endemic [48]. There have been fluctuations in the incidence of cholangiocarcinoma in the United States with increases reported between 1973 and 1997. This perceived increase may relate to changes in the classification of cholangiocarcinoma which resulted in skewed reporting and inflated rates [49].

In terms of gender, the incidence of this cancer varies: intrahepatic cholangiocarcinoma is more common in men compared to women [50], whereas extrahepatic cholangiocarcinoma is similar in the two sexes [50]. Ethnicity is also important in the epidemiology of cholangiocarcinoma. In the United States, the incidence is highest among Asian Americans and Pacific Islanders, then American Indian and Hispanic, followed by African Americans and Caucasians.

Risk factors

Similar to gallbladder cancer, cholangiocarcinoma has some key risk factors. These include: liver fluke infestations, primary sclerosing cholangitis, chronic viral hepatitis, choledochal cysts, and hepatolithiasis.

Liver fluke infestation
Clonorchis sinensis (Chinese liver fluke) and *Opisthorchis viverrini* (Southeast Asian liver fluke) are the key liver flukes that increase the risk of developing cholangiocarcinoma. *Opisthorchis viverrini* is considered a definite carcinogen and *Clonorchis sinensis*, a probable carcinogen [51]. *Clonorchis sinensis* is predominantly found in Asia (southern China, Japan, Korea, and Taiwan), related to the custom of eating raw freshwater fish or shellfish. Chronic infestations, occurring over 20 to 30 years, induce an inflammatory response that progresses to cholangiocarcinoma. *Opisthorchis viverrini*, endemic in Thailand and Laos, results from ingesting undercooked fish.

Primary sclerosing cholangitis (PSC)
This progressive inflammatory disorder of the biliary tract is the major cause (~30 %) of cholangiocarcinoma in the Western world. Cholangiocarcinoma can develop within a few years from the time of primary sclerosing cholangitis diagnosis [52]. The term "primary" distinguishes PSC from other secondary cholangiopathies such as bacterial cholangitis. Its incidence is 1 per 100,000 person-years, higher in men than women. PSC is particularly associated with inflammatory bowel disease: 90 % have ulcerative colitis, more commonly in the form of a pancolitis than distal colitis. Among patients with PSC, 0.6 to 1.5 % develop cholangiocarcinoma on an annual basis, yielding a lifetime risk of 10–15 %.

Viral hepatitis
Both hepatitis B and C virus (HBV/HCV) infections are risk factors for developing cholangiocarcinoma. The carcinogenic mechanism is unknown [53].

Hepatolithiasis
Hepatolithiasis, another source for chronic, recurrent infection and inflammation, is associated with cholangiocarcinoma [53]. Up to 10 % of individuals with hepatolithiasis will develop cholangiocarcinoma [54]. The stones are characteristically located close to the site of the tumor [55].

Alcohol
Heavy alcohol consumption appears to convey an increased risk of developing cholangiocarcinoma, despite some conflicting reports [56].

Cirrhosis
Individuals with cirrhosis are at a significantly (13-fold) increased risk of developing cholangiocarcinoma [57]. The interaction of cirrhosis with chronic hepatitis viral infections further increases the risk for intrahepatic cholangiocarcinoma [58].

Conclusions

The epidemiology of gallstone disease, gallbladder cancer and cholangiocarcinoma is evolving with more rigorous studies that better illuminate changes in their frequency and attendant risk factors (Box 26.1). Gallstone disease is common; its prevalence varies substantially depending on geography and ethnicity. The relationship between gallstone disease and genetics remains unclear in humans. There are several modifiable risk factors for gallstone disease that provide opportunities for primary prevention. Biliary tract malignancies appear related to chronic inflammatory processes. Gallbladder cancer is considered rare, but can be endemic in certain populations that also have a high prevalence of gallstone disease, such as American Indians. Cholangiocarcinoma is a more geographically defined entity that affects predominantly Asians and those with primary sclerosing cholangitis. Recognition of the epidemiology of these conditions offers the opportunity to identify primary and secondary preventative measures to reduce their morbidity and mortality.

Box 26.1 Risk factors for gallstone disease and gallbladder cancer.

Gallstone disease
Obesity
Rapid weight loss
Low physical activity
Drugs

Total parenteral nutrition (TPN)
Use of female sex hormones
Diet
Metabolic syndrome
Cirrhosis
Crohn's disease
Family history of gallbladder disease
Ethnicity
Female gender

Gallbladder cancer
Increased parity
Salmonella typhi, *Salmonella paratyphi* infection
Helicobacter bilis, *Helicobacter pylori* infection
Family history of gallbladder disease or cancer
Ethnicity
Female gender

Cholangiocarcinoma
Liver flukes (*Clonorchis sinensis*, *Opistorchis viverrini*)
Primary sclerosing cholangitis
Choledochal cysts
Viral hepatitis B and C
HIV
Cirrhosis
Hepatolithiasis
Toxins
Lynch syndrome
Diabetes
Obesity

Multiple choice questions

1 Which one of the following represents risks for cholesterol gallstone formation?
A Old age
B Metabolic syndrome
C Black African
D Advanced cirrhosis
E Elevated total serum cholesterol

2 Gallbladder cancer is more common in which of the following?
(More than one answer may be correct.)
A American Indians
B Rapid weight loss
C Cholelithiasis
D Caucasians
E Black Africans

3 Cholangiocarcinoma is associated with which of the following?
(More than one answer may be correct.)
A Gallstone disease
B Primary sclerosing cholangitis
C Liver flukes
D Hepatolithiasis
E Cystic fibrosis

References

1 Shaffer EA. Gallstone disease: Epidemiology of gallbladder stone disease. *Best Pract Res Clin Gastroenterol* 2006;20:981–96.
2 Gracie WA, Ransohoff DF. The natural history of silent gallstones: the innocent gallstone is not a myth. *New Engl J Med* 1982;307:798–800.
3 Everhart JE, Ruhl CE. Burden of digestive diseases in the United States part I: overall and upper gastrointestinal diseases. *Gastroenterology* 2009;136:376–86.
4 Ruhl CE, Everhart JE. Gallstone disease is associated with increased mortality in the United States. *Gastroenterology* 2011;140:508–16.
5 Aerts R, Penninckx F. The burden of gallstone disease in Europe. *Aliment Pharmacol Ther* 2003;18(Suppl): 349–53.
6 Legorreta AP, Silber JH, Costantino GN, et al. Increased cholecystectomy rate after the introduction of laparoscopic cholecystectomy. *JAMA* 1993;270:1429–32.
7 Kang JY, Ellis C, Majeed A, et al. Gallstones – an increasing problem: a study of hospital admissions in England between 1989/1990 and 1999/2000. *Aliment Pharmacol Ther* 2003;17:561–9.
8 Everhart JE, Yeh F, Lee ET, et al. Prevalence of gallbladder disease in American Indian populations: findings from the Strong Heart Study. *Hepatology* 2002;35: 1507–12.
9 Sampliner RE, Bennett PH, Comess LJ, et al. Gallbladder disease in pima indians. Demonstration of high prevalence and early onset by cholecystography. *New Engl J Med* 1970;283:1358–64.
10 Everhart JE. Gallstones and ethnicity in the Americas. *J Assoc Acad Minor Phys* 2001;12:137–43.
11 Biss K, Ho KJ, Mikkelson B, et al. Some unique biologic characteristics of the Masai of East Africa. *New Engl J Med* 1971;284:694–9.
12 Trotman BW, Soloway RD. Pigment gallstone disease: Summary of the National Institutes of Health – International Workshop. *Hepatology* 1982;2:879–84.

13 Kaechele V, Wabitsch M, Thiere D, et al. Prevalence of gallbladder stone disease in obese children and adolescents: influence of the degree of obesity, sex, and pubertal development. *J Pediatr Gastroenterol Nutr* 2006;42: 66–70.
14 Cirillo DJ, Wallace RB, Rodabough RJ, et al. Effect of estrogen therapy on gallbladder disease. *JAMA* 2005; 293:330–9.
15 Maclure KM, Hayes KC, Colditz GA, et al. Weight, diet, and the risk of symptomatic gallstones in middle-aged women. *New Engl J Med* 1989;321:563–9.
16 Everhart JE. Contributions of obesity and weight loss to gallstone disease. *Ann Intern Med* 1993;119:1029–35.
17 Nervi F, Miquel JF, Alvarez M, et al. Gallbladder disease is associated with insulin resistance in a high risk Hispanic population. *J Hepatol* 2006;45: 299–305.
18 Li VK, Pulido N, Fajnwaks P, et al. Predictors of gallstone formation after bariatric surgery: a multivariate analysis of risk factors comparing gastric bypass, gastric banding, and sleeve gastrectomy. *Surg Endosc* 2009;23:1640–4.
19 Mendez-Sanchez N, Zamora-Valdes D, Chavez-Tapia NC, Uribe M. Role of diet in cholesterol gallstone formation. *Clin Chim Acta* 2007;376:1–8.
20 Leitzmann MF, Rimm EB, Willett WC, et al. Recreational physical activity and the risk of cholecystectomy in women. *New Engl J Med* 1999;341:777–84.
21 Wang HH, Portincasa P, Afdhal NH, Wang DQH. *Lith* genes and genetic analysis of cholestoerol gallstone formation. *Gastroenterol Clin North Am* 2010;39:185–207.
22 Lammert F, Miquel JF. Gallstone disease: from genes to evidence-based therapy. *J Hepatol* 2008;48(Suppl 1):S124–S135.
23 Sarin SK, Negi VS, Dewan R, et al. High familial prevalence of gallstones in the first-degree relatives of gallstone patients. *Hepatology* 1995;22:138–41.
24 Acalovschi M, Badea R, Dumitrascu D, Varga C. Prevalence of gallstones in liver cirrhosis: a sonographic survey. *Am J Gastroenterol* 1988;83:954–6.
25 Conte D, Barisani D, Mandelli C, et al. Cholelithiasis in cirrhosis: analysis of 500 cases. *Am J Gastroenterol* 1991;86:1629–32.
26 Alvaro D, Angelico M, Gandin C, et al. Physico-chemical factors predisposing to pigment gallstone formation in liver cirrhosis. *J Hepatol* 1990;10:228–34.
27 Acalovschi M, Buzas C, Radu C, Grigorescu M. Hepatitis C virus infection is a risk factor for gallstone disease: a prospective hospital-based study of patients with chronic viral C hepatitis. *J Viral Hepat* 2009;16:860–6.
28 Loria P, Lonardo A, Lombardini S, et al. Gallstone disease in non-alcoholic fatty liver: prevalence and associated factors. *J Gastroenterol Hepatol* 2005;20:1176–84.
29 Whorwell PJ, Hawkins R, Dewbury K, Wright R. Ultrasound survey of gallstones and other hepatobiliary disorders in patients with Crohn's disease. *Dig Dis Sci* 1984;29:930–3.
30 Vitek L, Carey MC. Enterohepatic cycling of bilirubin as a cause of 'black' pigment gallstones in adult life. *Eur J Clin Invest* 2003;33:799–810.
31 Curry MP, Hegarty JE. The gallbladder and bilary tract in cystic fibrosis. *Curr Gastroenterol Rep* 2005;7: 147–53.
32 Chiong C, Cox MR, Eslick GD. Gallstone disease is associated with rectal cancer: A meta-analysis. *Scand J Gastroenterol* 2012;47:553–64.
33 Chiong C, Cox MR, Eslick GD. Gallstones are associated with colonic adenoma: A meta-analysis. *World J Surg* 2012;36:2202–9.
34 Corazziari E, Attili AF, Angeletti C, De Santis A. Gallstones, cholecystectomy and irritable bowel syndrome (IBS): MICOL population-based study. *Dig Liver Dis* 2008;40:944–50.
35 Xia CS, Han YQ, Yang XY, Hong GX. Spinal cord injury and cholelithiasis. *Hepatobiliary Pancreat Dis Int* 2004;3:595–8.
36 Portincasa P, Moschetta A, Petruzzelli M, et al. Symptoms and diagnosis of gallbladder stones. *Best Prac Res Clin Gastroenterol* 2006;20:1017–29.
37 Eslick GD. Epidemiology of gallbladder cancer. *Gastroenterol Clin North Am* 2010;39:307–30.
38 Wistuba II, Gazdar AF. Gallbladder cancer: Lessons from a rare tumour. *Nat Rev Cancer* 2004;695–706.
39 Curado M P, Edwards B, Shin HR, et al. (eds.) (2007) *Cancer Incidence in Five Continents*, Vol. IX. IARC Scientific Publications No. 160, IARC, Lyon.
40 Hariharan D, Saied A, Kocher HM. Analysis of mortality rates for gallbladder cancer across the world. *HPB* 2008;10:327–31.
41 Schlenker C, Surawicz CM. Emerging infections of the gastrointestinal tract. *Best Prac Res Clin Gastroenterol* 2009;23:89–99.
42 Mishra RR, Tewari M, Shukla HS. *Helicobacter* species and pathogenesis of gallbladder cancer. *Hepatobiliary Pancreat Dis Int* 2010;9:129–34.
43 Zatonski W, La Vecchia C, Levi F, et al. Descriptive epidemiology of gall-bladder cancer in Europe. *J Cancer Res Clin Oncol* 1993;119:165–71.
44 Kiran RP, Pokala N, Dudrick SJ. Incidence pattern and survival for gallbladder cancer over three decades: an analysis of 10301 patients. *Ann Surg Oncol* 2007;14:827–32.
45 Randi G, Franceschi S, La Vecchia C. Gallbladder cancer worldwide: geographical distribution and risk factors. *Int J Cancer* 2006;118:1591–1602.

46 Dutta U, Nagi B, Garg PK, et al. Patients with gallstones develop gallbladder cancer at an earlier age. *Eur J Cancer Prev* 2005;14:381–5.
47 Larsson SC, Wolk A. Obesity and the risk of gallbladder cancer: a meta-analysis. *Brit J Cancer* 2007;96:1457–61.
48 Sripa B, Pairojkul C. Chlangiocarcinoma: lessons from Thailand. *Curr Opin Gastroenterol* 2008;24:349–56.
49 Khan SA, Emadossadaty S, Ladep NG, et al. Rising trends in cholangiocarcinoma: Is the ICD classification system misleading us? *J Hepatol* 2012;56:848–54.
50 Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012;62:10–29.
51 IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 61 (1994). *Schistosomes, Liver Flukes and* Helicobacter pylori.
52 Kitiyakara T, Chapman RW. Chemoprevention and screening in primary sclerosing cholangitis. *Postgrad Med J* 2008;84:228–37.
53 Srivatanakul P, Honjo S, Kittiwatanachot P, et al. Hepatitis viruses and risk of cholangiocarcinoma in Northeast Thailand. *Asian Pacific J Cancer Prev* 2010;11:985–8.
54 Suzuki Y, Mori T, Abe N, et al. Predictive factors for cholangiocarcinoma associated with hepatolithiasis determined on the basis of Japanese Multicenter study. *Heaptol Res* 2012;42:166–70.
55 Liu Z-Y, Zhou Y-M, Shi L-H, Yin Z-F. Risk factors for intrahepatic cholangiocarcinoma in patients with hepatolithiasis: a case-control study. *Hepatobiliary Pancreat Dis Int* 2011;10:626–31.
56 Tyson GL, El-Serag HB. Risk factors for cholangiocarcinoma. *Hepatology* 2011;54:173–84.
57 Kalaitzakis E, Gunnarsdottir SA, Josefsson A, Bjornsson E. Increased risk for malignant neoplasms among patients with cirrhosis. *Clin Gastroenterol Hepatol* 2011;9:168–74.
58 Tao L-T, He X-D, Qu Q, et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: a case-control study in China. *Liver Int* 2009;30:215–21.

Answers to multiple choice questions

1. B
2. A, C
3. B, C, D

27 Epidemiology of pancreatitis

DhirajYadav[1], *Santhi Swaroop Vege*[2], *& Suresh T. Chari*[3]
[1]Division of Gastroenterology & Hepatology, University of Pittsburgh, Pittsburgh, PA, USA
[2]Division of Gastroenterology & Hepatology, Mayo Clinic, Rochester, MN, USA
[3]Division of Gastroenterology & Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

Key points
- The incidence of acute pancreatitis appears to be increasing worldwide.
- The natural history of alcoholic chronic pancreatitis differs from idiopathic forms.
- Future studies on pancreatitis should focus on host–environment interactions, factors determining quality of life and healthcare costs.

Acute pancreatitis

Clinical summary

Patients with acute pancreatitis (AP) typically present with severe, continuous upper abdominal pain radiating to the back. Severe cases may develop organ failure, and local complications such as fluid collections and pancreatic necrosis may occur. Treatment is predominantly conservative in the form of pain relief by narcotics and aggressive fluid resuscitation. Patients with severe AP may require monitoring and treatment in the intensive care unit, prophylactic antibiotics, enteral nutrition, and debridement of infected necrotic pancreatic tissue.

Disease definition

Clinically, AP is diagnosed when patients with upper abdominal pain have a threefold elevation of serum pancreatic enzymes (amylase and/or lipase) and/or evidence of pancreatic inflammation on imaging studies (e.g. computerized tomography scan). AP can recur in up to one third of patients.

Incidence and prevalence

The annual incidence of AP over the past two decades has been in the range of ~20–40 per 100,000 population [1–5]. The incidence varies widely between countries due to differences in the distribution of risk factors and the type of study design. Incidence rates have been noted to be lower in Great Britain [3,6,7], the Netherlands [8], and Germany [9], somewhat higher in Denmark [10], Sweden [11], Norway [4], and Iceland [12], and the highest in the United States [1] and Finland [13]. Studies using administrative datasets report higher incidence rates than studies where the diagnosis is validated by review of records. Studies are beginning to examine the accuracy of administrative data for AP. Use of hospital admissions to capture incident cases is likely to be representative of true incidence rates. Because AP is not usually a chronic condition, all the population-based studies of AP describe incidence and not prevalence.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Temporal trends in incidence

A consistent observation in most population-based studies is the rise in incidence of AP [5]. As an example, the number of admissions for AP in the United States increased from 101,000 in 1988 to 210,000 in 2003 [1]. The rising incidence is likely due to multiple factors of which the most important are an increase in the rates of gallstone AP from rising obesity and increased testing for pancreatitis among patients with abdominal pain. Other contributing factors would include an increase in alcohol-related AP, medication-induced AP, and post-procedure pancreatitis.

Demographics and risk factors

Increased risk of AP has been linked to particular sections of the population or certain factors.

Gender and race

While most studies report a higher incidence in males [7–9,11,12,14,15], a Danish study [10] reported a higher incidence in women. Temporal trends in Danish and British studies show a more pronounced increase in incidence rates in females compared with males [7,10]; the British study also observed an increase in the proportion of women who drank >14 units of alcohol per week [7]. The rates of AP in Blacks have been noted to be higher than in Whites [1].

Age

Many studies have observed increasing incidence of AP with increasing age; the incidence reported per 100,000 population was <5–10, 10–30 and >20–30 in age groups <25 years, 25–60 years, and >60 years, respectively [7,8,15]. In southern England, the more pronounced increase in AP in younger men and women is, at least in part, due to an increase in alcohol-related AP [15].

Gallstones and alcohol

Gallstones are the commonest cause of AP [9,11, 12,14], and with alcohol abuse, account for >60 % of cases. Gallstone-induced AP is more common in women whereas alcohol-induced AP is more often seen in men. Lindkvist et al. observed an increase of 7.6 % per year in gallstone-induced AP in Sweden and correlated this finding with increased obesity and gallstone-related diseases [11]. The same authors also found a decrease in alcohol-related AP of 5.1 % per year and correlated this finding with a decrease in the incidence of delirium tremens and mortality from cirrhosis, both markers of alcohol-induced diseases [11].

Drugs

A Danish study showing increasing incidence of AP over time also observed an increase in the number of prescriptions for potentially pancreatitis-causing drugs such as azathioprine, estrogens, and estrogen-progesterone combinations during the study period [10]. However, it is not clear that the patients with AP in this population were exposed to these drugs.

Healthcare costs

The health costs of pancreatitis care are substantial. Using the Nationwide Inpatient Sample, a recent US study estimated the direct cost of AP in 2003 to be $2 billion [16]. In another analysis, the total cost of care for all pancreatitis at nonfederal US institutions in 2004 was estimated to be $3.2 billion [17]. Although this analysis did not separate AP from CP, a large fraction of this is likely related to AP due to higher incidence rates (several fold higher for AP then CP) and need for inpatient care (almost all AP patients).

Natural history and mortality

Approximately 80 % of patients with AP will have mild disease and recover without sequelae. The remaining 20 % with severe AP will have a prolonged hospital stay due to organ failure, local complications, and sepsis. While the overall mortality in AP is reported to be low (~2 %) [5], it increases with age and with severe disease. In severe AP, the mortality can be up to 15–25 %. Most deaths due to AP (65 %) occur in the first 14 days [10[. In recent studies the case-fatality rate (usually in the first 30 days) has ranged from 3 to 10 % [18], with some studies showing a decreasing case-fatality rate over time [8,10,15], especially in younger patients [8,15]. This decrease was not seen in population-based studies [7,8,15]. There was no difference in the mortality due to different etiologies. Recurrent attacks are associated with lower mortality compared with a first attack of AP [8,12].

Issues and gaps in the epidemiology

The very few prospective studies cover only a short period of time. Hospital-based studies using solely administrative data may overestimate incidence. When compared to community hospitals, patients treated at referral hospitals are sicker since many are transferred from smaller hospitals for expert care. Limited data exists on the epidemiology of AP in regions other than the United States and Europe.

Recommendations for future studies

Epidemiology of AP should be studied in regions with limited data to understand distribution and trends. In defined populations, excluding cases transferred from hospitals outside the area, recurrent attacks and flares of chronic pancreatitis (CP), and all cases of confirmed AP should be prospectively studied for temporal trends, etiology, clinical outcomes, and healthcare utilization costs. Such studies could perhaps be done in centers with well-defined catchment areas and a limited number of healthcare facilities.

Conclusions

Based on available epidemiologic studies (which are mostly retrospective and hospital-based), it appears that the incidence of AP is increasing (especially AP due to gallstones and alcohol in some areas), and that the case fatality rate is decreasing. Prospective, preferably population- based, studies are needed to confirm these findings.

Chronic pancreatitis (CP)

Disease definition

Chronic pancreatitis (CP) is a progressive fibroinflammatory disease of the pancreas that, in its end stages, is characterized by permanent loss of pancreatic parenchyma and consequent functional insufficiency (diabetes and steatorrhea) [19]. Three forms of CP are currently recognized.

Usual CP, or calcifying CP (CCP)

This is characterized by severe abdominal pain, recurrent bouts of clinical AP, and eventual development of intraductal calculi in a high proportion of cases. On histology, there is perilobular fibrosis and acinar destruction with acute and chronic inflammatory cells [20]. The most frequent cause of CCP is alcohol and tobacco use.

Obstructive CP

This form of CP develops upstream from an area of ductal obstruction, often due to a tumor or post-inflammatory AP pancreatic duct stricture. It is usually painless but occasionally causes clinical AP. Persistent obstruction leads to pancreatic atrophy upstream from the area of ductal narrowing. The development of steatorrhea and diabetes depends on the amount of pancreas that becomes atrophied. Intraductal calculi are generally absent.

Autoimmune CP

This systemic autoimmune fibro-inflammatory disorder afflicts the pancreas as well as other organs [21]. It is a relatively painless disorder and clinical AP is not a common presentation. Affected organs show infiltration of IgG4-positive cells and it dramatically responds to steroid therapy. Intense fibrosis in AIP may lead to permanent structural damage and functional insufficiency. Intraductal calculi are uncommon, but may develop in the late "burnt-out" stage. More recently, two forms of the disease have been recognized. Type I or lymphoplasmacytic sclerosing pancreatitis is usually seen in older individuals, presents as obstructive jaundice with focal or diffuse mass in the pancreas and elevated serum IgG4 levels. On histology there is lymphoplasmacytic infiltrate rich in IgG4-positive cells, intense fibrosis and destruction of venules. Type II is seen in younger individuals, is often associated with inflammatory bowel disease and patients may not have elevated serum IgG4 levels. On histology, granulocyte infiltration of the duct wall with/without acinar inflammation is present. IgG4 cell infiltration is less prominent or absent.

Epidemiology of usual CP (or CCP)

Incidence and prevalence
Almost all literature on the epidemiology of CP relates to CCP, mostly from Western countries and Japan, with little information on the epidemiology of other

forms. The annual incidence in Western countries ranges from 4–9 per 100,000 [22]. The variability in incidence rates depends on the study design, year of study, and risk factor prevalence. Prevalence estimates are available from Copenhagen (27.4 per 100,000 in 1979) [23], Japan (28.9 per 100,000 in 1994) [24], and more recently from Olmsted County, MN, USA (41.8 per 100,000 in 2006) [25].

Temporal trends

Longitudinal studies indicate a trend toward an increase in incidence over time [25].

Risk factors

In most reports from Western countries between the 1960s–1990s, heavy alcohol use accounted for the majority of CP cases (70–90 %). In recent studies, the etiologic spectrum of CP has been noted to be wider and the proportion of patients directly attributed to alcohol is somewhat lower (~50 %) [25–27]. Smoking has been recognized as an independent dose-dependent risk factor [28]. Alcohol and smoking together likely have a multiplicative effect on the risk of CP [29]. The risk of developing pancreatitis increases significantly with consumption of 4–5 drinks or more per day [29–31]. The absolute risk of pancreatitis with heavy drinking is ~2–3 % [32]. The exact role of diet in CP is still unclear. Structural abnormalities of the pancreas, specifically pancreatic ducts are associated with CP in some patients. Other less common associations of CP are hypertriglyceridemia and hypercalcemia.

Genetic susceptibility to CCP is conferred by mutations in the cationic trypsinogen, *CFTR* and *SPINK1* genes. Hereditary pancreatitis is an autosomal dominant disorder with high (80 %) penetrance caused by mutations in the cationic trypsinogen gene. Mutations in the *CFTR* and *SPINK1* genes are associated with apparently idiopathic CCP [33].

Demographics and presentation

Alcoholic CCP, which is more common in middle-aged men with a long history of heavy alcohol and tobacco use, usually presents in the fifth decade of life with attacks of pain or AP. The presentation of the idiopathic form of CCP, which affects both sexes equally, is bimodal: the juvenile form (early-onset) is painful, while over 50 % of subjects with senile-onset idiopathic CP have painless disease [34]. Most patients with hereditary pancreatitis are symptomatic by age 20 years, with pain and clinical AP [19]. Tropical pancreatitis is an early-onset form of idiopathic CP that is endemic in south Asia, particularly southern India, and in Africa and South America. It is characterized by a high prevalence of pancreatic calcification, diabetes, and pancreatic cancer [35]. Data on racial predisposition for CP is limited but important observations indicate that black people may have a greater risk for alcoholic pancreatitis compared with white people [36].

Natural history

Described mainly from centers specializing in pancreatic disease, the natural history of alcoholic CP may be different from idiopathic CP [34,37,38]. Patients with early-onset idiopathic CP have a much slower progression toward requiring pain relief, and to experiencing exocrine and endocrine insufficiency compared with alcoholic and late-onset idiopathic CP [34]. Approximately 50–60 % of patients undergo surgery at some point, primarily to achieve pain relief or to treat complications from CP [34,37]. The mortality in CP subjects is two- to fourfold higher compared to background population and is mostly from nonpancreatic causes [25,38]. The cumulative risk of developing pancreatic cancer is much greater in hereditary pancreatitis (40 %) than in other forms of CP [39].

Disability and quality of life

Abdominal pain, which can be continuous and intractable, is the most important determinant of quality of life in CP [40]. About 50 % of patients regularly use pain medications and about a third have disability from pancreatitis [41].

Prevention

Recent data suggest that progression from AP to CP is not inevitable and occurs only in a subset of patients (10–30 %) [42–44]. The risk is increased with continued alcohol consumption and smoking. Thus, the natural history of CP can be altered by aggressive counseling and appropriate measures for abstinence from

alcohol and smoking cessation. Currently, no preventive strategies are available for other forms of CP.

Issues and gaps in the epidemiology

Despite the progress made in recent years in understanding the pathogenesis of CP, especially its inherited forms, several questions remain unanswered. For example, what is the role of cofactors (host, environmental, or both) in individual susceptibility to develop CP? What is the mechanism of pain in CP? Why do only a small proportion of heavy-drinking alcoholics develop CP? What is the role of alcohol at intermediate levels of consumption? How does smoking increase the risk and modify disease progression? What are the determinants of resource utilization and disability in subjects with CP? How accurate are administrative datasets in identifying cases of CP? The current classification of CP is based on morphology rather than etiology. Because biopsy from the pancreas is rarely obtained, classification systems providing an improved understanding of both the etiology and progression of the disease are needed.

Recommendations for future studies

Future studies should focus on establishing incidence and prevalence estimates and trends in general and high-risk groups in different populations, factors determining individual susceptibility to CP, mechanisms of pain in CP, and determinants of quality of life and healthcare utilization by subjects with CP.

Conclusions

Significant advances have been made in our understanding of the etiology, mechanisms and natural history of CP. Most epidemiologic studies on CP have originated from specialized centers and are not population-based. Well-designed, preferably prospective, population-based studies are needed to understand better the disease estimates and trends.

Multiple choice questions

1 The factors contributing to the increasing incidence of acute pancreatitis include:
- **A** Gallstones
- **B** Increased use of serum pancreatic enzymes
- **C** Alcohol
- **D** Medications
- **E** All

2 The overall mortality in patients with acute pancreatitis is:
- **A** 1–2 %
- **B** 5–10 %
- **C** 10–15 %
- **D** More than 20 %

3 The risk of developing pancreatitis in heavy alcoholics is:
- **A** 1–3 %
- **B** 5–10 %
- **C** 10–20 %
- **D** >20 %

4 All of the following are true for chronic pancreatitis except:
- **A** Alcohol and smoking are independent risk factors
- **B** The effects of alcohol and smoking on the pancreas may be multiplicative
- **C** Cessation of alcohol and smoking does not prevent progression from acute to chronic pancreatitis
- **D** Only a small fraction of patients progress from acute to chronic pancreatitis

References

1 Fagenholz PJ, et al. Increasing United States hospital admissions for acute pancreatitis, 1988–2003. *Ann Epidemiol* 2007;17:491–7.

2 Frey CF, et al. The incidence and case-fatality rates of acute biliary, alcoholic, and idiopathic pancreatitis in California, 1994–2001. *Pancreas* 2006;33:336–44.

3 O'Farrell A, et al. Hospital admission for acute pancreatitis in the Irish population, 1997–2004: could the increase be due to an increase in alcohol-related pancreatitis? *J Public Health (Oxf)* 2007;29:398–404.

4 Omdal T, et al. Time trends in incidence, etiology, and case fatality rate of the first attack of acute pancreatitis. *Scand J Gastroenterol* 2011;46:1389–98.

5 Yadav D, et al. Trends in the epidemiology of the first attack of acute pancreatitis: a systematic review. *Pancreas* 2006;33:323–30.

6 Roberts SE, et al. Incidence and case fatality for acute pancreatitis in England: geographical variation, social deprivation, alcohol consumption and aetiology – a record linkage study. *Aliment Pharmacol Ther* 2008;28:931–41.

7 Tinto A, et al. Acute and chronic pancreatitis – diseases on the rise: a study of hospital admissions in England 1989/90–1999/2000. *Aliment Pharmacol Ther* 2002;16: 2097–105.

8 Eland IA, et al. Incidence and mortality of acute pancreatitis between 1985 and 1995. *Scand J Gastroenterol* 2000;35:1110–6.

9 Lankisch PG, et al. Temporal trends in incidence and severity of acute pancreatitis in Luneburg County, Germany: a population-based study. *Pancreatology* 2009;9:420–6.

10 Floyd A, et al. Secular trends in incidence and 30-day case fatality of acute pancreatitis in North Jutland County, Denmark: a register-based study from 1981–2000. *Scand J Gastroenterol* 2002;37:1461–5.

11 Lindkvist B, et al. Trends in incidence of acute pancreatitis in a Swedish population: is there really an increase? *Clin Gastroenterol Hepatol* 2004;2:831–7.

12 Birgisson H, et al. Acute pancreatitis: a prospective study of its incidence, aetiology, severity, and mortality in Iceland. *Eur J Surg* 2002;168:278–82.

13 Jaakkola M, et al. Pancreatitis in Finland between 1970 and 1989. *Gut* 1993;34:1255–60.

14 Gislason H, et al. Acute pancreatitis in Bergen, Norway. A study on incidence, etiology and severity. *Scand J Surg* 2004;93:29–33.

15 Goldacre MJ, et al. Hospital admission for acute pancreatitis in an English population, 1963–98: database study of incidence and mortality. *BMJ* 2004;328:1466–9.

16 Fagenholz PJ, et al. Direct medical costs of acute pancreatitis hospitalizations in the United States. *Pancreas* 2007;35:302–7.

17 Everhart JE, et al. Burden of digestive diseases in the United States part III: liver, biliary tract, and pancreas. *Gastroenterology* 2009;136:1134–44.

18 Andersson R, et al. Incidence, management and recurrence rate of acute pancreatitis. *Scand J Gastroenterol* 2004;39:891–4.

19 Etemad B, et al. Chronic pancreatitis: diagnosis, classification, and new genetic developments. *Gastroenterology* 2001;120:682–707.

20 Kloppel G, et al. Pathology of acute and chronic pancreatitis. *Pancreas* 1993;8:659–70.

21 Sugumar A, et al. Autoimmune pancreatitis. *J Gastroenterol Hepatol* 2011;26:1368–73.

22 Yadav D, et al. The role of alcohol and smoking in pancreatitis. *Nat Rev Gastroenterol Hepatol* 2010;7: 131–45.

23 Copenhagen pancreatitis study. An interim report from a prospective epidemiological multicentre study. *Scand J Gastroenterol* 1981;16:305–12.

24 Lin Y, et al. Nationwide epidemiological survey of chronic pancreatitis in Japan. *J Gastroenterol* 2000;35: 136–41.

25 Yadav D, et al. Incidence, prevalence, and survival of chronic pancreatitis: a population-based study. *Am J Gastroenterol* 2011;106:2192–9.

26 Cote GA, et al. Alcohol and smoking as risk factors in an epidemiology study of patients with chronic pancreatitis. *Clin Gastroenterol Hepatol* 2011;9:266–73.

27 Frulloni L, et al. Chronic pancreatitis: report from a multicenter Italian survey (PanCroInfAISP) on 893 patients. *Digest Liver Dis* 2009;41:311–7.

28 Andriulli A, et al. Smoking as a cofactor for causation of chronic pancreatitis: a meta-analysis. *Pancreas* 2010;39:1205–10.

29 Yadav D, et al. Alcohol consumption, cigarette smoking, and the risk of recurrent acute and chronic pancreatitis. *Arch Intern Med* 2009;169:1035–45.

30 Irving HM, et al. Alcohol as a risk factor for pancreatitis. A systematic review and meta-analysis. *JOP* 2009;10:387–92.

31 Kristiansen L, et al. Risk of pancreatitis according to alcohol drinking habits: a population-based cohort study. *Am J Epidemiol* 2008;168:932–7.

32 Lankisch PG, et al. What is the risk of alcoholic pancreatitis in heavy drinkers? *Pancreas* 2002;25:411–2.

33 LaRusch J, et al. Genetics of pancreatitis. *Curr Opin Gastroenterol* 2011;27:467–74.

34 Layer P, et al. The different courses of early- and late-onset idiopathic and alcoholic chronic pancreatitis. *Gastroenterology* 1994;107:1481–1487.

35 Mohan V, et al. Tropical chronic pancreatitis: an update. *J Clin Gastroenterol* 2003;36:337–46.

36 Lowenfels AB, et al. Racial factors and the risk of chronic pancreatitis. *Am J Gastroenterol* 1999;94:790–4.

37 Ammann RW, et al. Course and outcome of chronic pancreatitis. Longitudinal study of a mixed medical-surgical series of 245 patients. *Gastroenterology* 1984;86:820–8.

38 Lowenfels AB, et al. Prognosis of chronic pancreatitis: an international multicenter study. *International Pancreatitis Study Group. Am J Gastroenterol* 1994;89: 1467–71.

39 Raimondi S, et al. Epidemiology of pancreatic cancer: an overview. *Nat Rev Gastroenterol Hepatol* 2009;6:699–708.

40 Pezzilli R, et al. The quality of life in patients with chronic pancreatitis evaluated using the SF-12 questionnaire: a comparative study with the SF-36 questionnaire. *Digest Liver Dis* 2006;38:109–15.

41 Mullady DK, et al. Type of pain, pain-associated complications, quality of life, disability and resource utilisation in chronic pancreatitis: a prospective cohort study. *Gut* 2011;60:77–84.

42 Lankisch PG, et al. Natural history of acute pancreatitis: a long-term population-based study. *Am J Gastroenterol* 2009;104:2797–805.

43 Nojgaard C, et al. Progression from acute to chronic pancreatitis: prognostic factors, mortality, and natural course. *Pancreas* 2011;40:1195–200.
44 Takeyama Y. Long-term prognosis of acute pancreatitis in Japan. *Clin Gastroenterol Hepatol* 2009;7: S15–17.

Answers to multiple choice questions

1. E
2. A
3. A
4. C

28 Epidemiology of pancreatic cancer

Aravind Sugumar & Santhi Swaroop Vege
Division of Gastroenterology & Hepatology, Mayo Clinic, Rochester, MN, USA

Key points

- An estimated 44,000 people were diagnosed with pancreatic cancer in the United States, and over 36,000 died from the disease in 2010.
- Symptoms of pancreatic cancer are nonspecific, unreliable, and occur late in the disease presentation.
- Treatment options for pancreatic cancer are limited. Surgical removal of the tumor affords the best prognosis, but is possible in less than 20 % of patients. Chemotherapy and radiation are not very effective and are palliative.
- Pancreatic cancer has the highest mortality rate of all major cancers. Ninety-four percent of pancreatic cancer patients will die within five years of diagnosis. Life expectancy with metastatic disease is just 3–6 months without treatment.
- Currently there are no screening strategies to detect the disease early when surgical removal of the tumor is still possible. Pancreatic cancer survival has not improved substantially in the last 40 years.

Disease definition

Neoplasms of the pancreas can be broadly classified into exocrine and endocrine cancers. The vast majority of pancreatic neoplasms tend to originate in the exocrine pancreas (>95 %) with a small number originating in the endocrine part (~5 %) [1]. The bulk of the cancers originating in the exocrine pancreas tend to be ductal adenocarcinomas (>90 %). The small numbers of nonductal cancers are constituted by acinar cell cancer, solid pseudopapillary tumors, and pancreatoblastomas. The term "pancreatic cancer" usually pertains to pancreatic ductal adenocarcinoma and hence in this chapter we will focus solely on pancreatic ductal adenocarcinoma, the most common cancer of the pancreas.

Incidence and prevalence

In 2010 in the United States, there were approximately 43,000 new cases of pancreatic cancer and 36,000 deaths due to pancreatic cancer (Figure 28.1). These rates have been relatively stable since 1980 (Figure 28.2). Pancreatic cancer is the tenth leading cause of cancer in men as well as women, comprising

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Estimated new cases*

Males				Females		
Prostate	217,730	28%		Breast	207,090	28%
Lung & bronchus	116,750	15%		Lung & bronchus	105,770	14%
Colon & rectum	72,090	9%		Colon & rectum	70,480	10%
Urinary bladder	52,760	7%		Uterine corpus	43,470	6%
Melanoma of the skin	38,870	5%		Thyroid	33,930	5%
Non-Hodgkin lymphoma	35,380	4%		Non-Hodgkin lymphoma	30,160	4%
Kidney & renal pelvis	35,370	4%		Melanoma of the skin	29,260	4%
Oral cavity & oharynx	25,420	3%		Kidney & renal pelvis	22,870	3%
Leukemia	24,690	3%		Ovary	21,880	3%
Pancreas	21,370	3%		Pancreas	21,770	3%
All sites	789,620	100%		All sites	739,940	100%

Estimated death

Males				Females		
Lung & bronchus	86,220	29%		Lung & bronchus	71,080	26%
Prostate	32,050	11%		Breast	39,480	15%
Colon & rectum	26,580	9%		Colon & rectum	24,790	9%
Pancreas	18,770	6%		Pancreas	18,030	7%
Liver & intraheptic bile duct	12,720	4%		Ovary	13,850	5%
Leukemia	12,660	4%		Non-Hodgkin lymphoma	9,500	4%
Esophagus	11,650	4%		Leukemia	9,180	3%
Non-Hodgkin lymphoma	10,710	4%		Uterine corpus	7,950	3%
Urinary bladder	10,410	3%		Liver & intraheptic bile duct	6,190	2%
Kidney & renal pelvis	8,210	3%		Brain & othernervous system	5,720	2%
All sites	299,200	100%		All sites	270,290	100%

Figure 28.1 Estimated new cases and deaths from pancreatic cancer in 2010. Source: Jemal et al. [2]. Reproduced with permission of John Wiley & Sons.

approximately 6 % of all new cancers [2]. Despite this, it is the fourth leading cause of cancer death in the United States, a testament to its poor prognosis [2]. Only lung, breast, and colorectal cancers supersede it and the incidence of those cancers vastly outnumbers cases of pancreatic cancer, which has the lowest 5-year survival of any known cancer [2]. Worldwide there are 278,684 new cases and 266,669 deaths per year based on a 2008 estimate. It was the eighth leading cause of cancer death in the world [3].

Although the reasons for such a high case fatality rate are many, a major source is the late stage at which most pancreatic cancers are diagnosed. Despite the availability of modern imaging techniques, the vast majority of pancreatic cancers are inoperable at the time of presentation (>80 %). Sudden onset obstructive jaundice with change in color of urine (dark) and stool (pale) is the most common presentation. Back pain, loss of appetite, weight loss, onset of diabetes mellitus, and cachexia are the other major symptoms of the disease. Such symptomatic patients either have locally advanced cancer (invasion of major blood vessels) or distant metastasis which precludes surgery. Therefore this pattern of presentation assumes importance in the majority of patients as surgery affords the best long-term survival for pancreatic cancer. The median survival of inoperable pancreatic cancer is 3–6 months. Surgery improves the median survival by 18–24 months. Although there is some correlation of tumor size at the time of diagnosis and survival times, the 5-year survival for even early stage cancer (i.e. resection that is lymph node-negative for cancer) is only 24–30 %. Five-year survival drops to 10 % for resection that is lymph node-positive for cancer

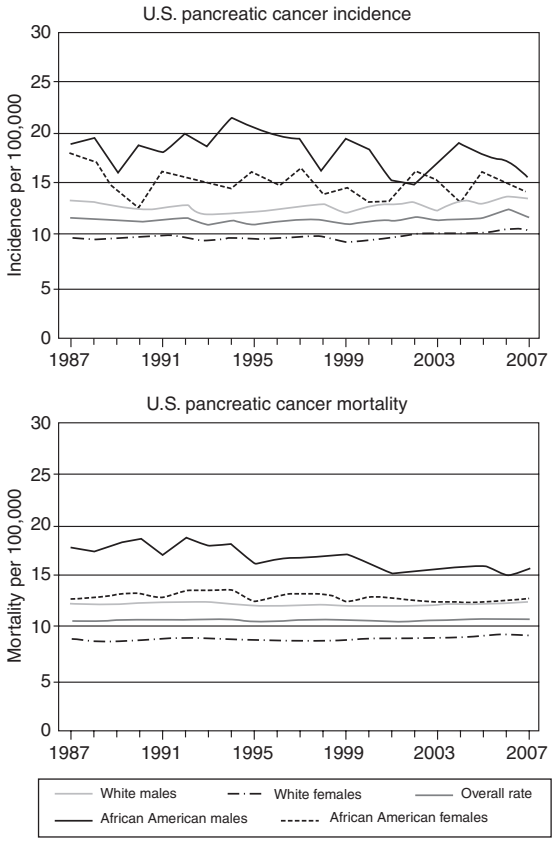


Figure 28.2 Incidence and mortality from pancreatic cancer 1987–2007. Source: Incidence and Mortality data: Surveillance, Epidemiology and End results (SEER) and the National Center for Health Statistics.

[4–6]. Recently, a chemotherapy regimen with a multidrug combination almost doubled the survival time of patients with advanced pancreatic cancer when compared to standard therapy alone [7]. Despite this, the results of chemotherapy and radiation therapy continue to be dismal at best. The few times that asymptomatic pancreatic cancer is diagnosed, is when it is serendipitously seen on abdominal imaging done for an unrelated reason. Such patients have the best long-term prognosis. Thus, cancer-specific symptoms such as pain and dramatic weight loss portend a poor prognosis even at the outset.

Most pancreatic adenocarcinomas arise in the head and neck of the pancreas (>75 %) with the rest arising in the body and tail of the pancreas. Most if not all pancreatic cancers arise from precursor lesions called PanIN (pancreatic intra-epithelial neoplasias) which are further classified as PanIN 1, PanIN 2, and PanIN 3, with increasing levels of dysplasia and consequently higher risk of cancer. Some pancreatic cancers arise in the setting of cystic lesions such an IPMN (intrapancreatic mucinous neoplasm) or MCN (mucinous cystic neoplasm). The cancer is classified according to the TNM classification and staged from stage 0–IV [8,9].

Risk factors

The risk factors for pancreatic cancer can be broadly classified as major and minor risk factors. The major risk factors for pancreatic cancer are smoking, duration and presence of chronic pancreatitis, genetic factors, obesity, and physical inactivity. The minor risk factors include age, sex, ABO blood group, diabetes mellitus, dietary factors, postsurgical states (cholecystectomy and partial gastrectomy), and *Helicobacter pylori* infection.

Major risk factors

Numerous case-control and cohort studies have demonstrated smoking to be an independent risk factor for pancreatic cancer. Studies demonstrate that when compared with never smokers, current smokers had a significantly elevated risk, with odds ratio (OR) estimates ranging from 1.5–5.7, of developing pancreatic cancer [10–15]. In addition, these studies also demonstrated a dose effect, that is, the risk of pancreatic cancer increased with the number of cigarettes smoked. In one recent study, when compared with never-smokers, current smokers had an OR of 1.7 (95 % CI 1.38–2.26). The risk increased significantly with more smoking (duration and quantity), that is, with greater than 30 cigarettes per day the OR was 1.75(95 % CI 1.27– 2.42), duration greater than 50 years the OR was 2.13, (95 % CI 1.25–3.62), and cumulative smoking dose greater than 40 pack-years the OR was 1.78 (95 % CI 1.35–2.34) [13]. Studies have also demonstrated that the risk of pancreatic cancer comes down with smoking cessation. The estimates for the time to return to baseline risk vary from 10–15 yrs [14–16].

The other well-associated risk factor for developing pancreatic cancer is presence of hereditary pancreatitis and duration of chronic pancreatitis regardless of the

cause. Hereditary pancreatitis is mostly due to an autosomal dominant mutation in the serine protease 1 gene (*PRSS1*) which codes for a cationic trypsinogen [17]. The rates of pancreatic cancer in patients with hereditary pancreatitis are increased 50-fold, although in absolute numbers, this is a rare cause of pancreatic cancer [18–22]. The cumulative risk of pancreatic cancer by age 70 in patients with hereditary pancreatitis is as high as 40 % [22]. The rates are even higher when such patients smoke [19].

Chronic pancreatitis due to any cause also increases the risk of pancreatic cancer. A multicenter cohort study of 2000 patients in 1993 showed that the risk of pancreatic cancer was significantly elevated in patients with chronic pancreatitis independent of other known risk factors [23]. This risk was as high as 26-fold when compared to the general population. Subsequent studies also showed an elevated risk in patients with chronic pancreatitis, albeit at a much lower level of around three- to fourfold elevation when compared to the general population [24–28].

The pathways by which chronic pancreatitis leads to pancreatic cancer are unknown. It is likely that DNA damage due to chronic inflammation leads to formation of cancer precursor lesions such as PanINs, which then turn into pancreatic cancer with continued inflammation [29,30]. Mutation in the gene *Kras*, which is present in most if not all pancreatic adenocarcinomas, may be induced by chronic inflammation or the chronic inflammation may accelerate the development of pancreatic cancer in patients with a mutated *Kras* gene [29,30]. A recent meta-analysis looking at risk of pancreatic cancer and all types of chronic pancreatitis concluded that over a 20-year period, pancreatic cancer will develop in only about 5 % or less of all patients with usual chronic pancreatitis [27]. However, it noted that risk is markedly increased in those patients with hereditary pancreatitis or tropical pancreatitis [27,31]. The most likely reason for this is the early onset of pancreatitis in these two groups, and consequently more at-risk years of chronic inflammation [27].

A wide array of genetic abnormalities increases the risk of pancreatic cancer. These genetic alterations can be divided into syndromic and nonsyndromic abnormalities. Syndromic causes include germline mutation in the *STK-1* gene seen in Peutz–Jeghers syndrome that increases the relative risk of pancreatic cancer by as much as 132 times [32,33]. Other known associations include *BRCA2* mutation, atypical multiple-mole melanoma (FAMMM) syndrome, familial adenomatous polyposis (FAP), Lynch syndrome, von Hippel–Lindau syndrome and multiple endocrine neoplasia [34–40]. In addition to the above syndromes, a mutation in the *PRSS1* gene, the cause of hereditary pancreatitis, is a genetic risk factor for pancreatic cancer. That said, these genetic factors account for less than 10 % of all cases of pancreatic cancer [40]. The nonsyndromic causes account for "high-risk families" which have an increased incidence of pancreatic cancer. These families have been defined as families having at least two first-degree relatives with pancreatic cancer. People with such family histories are being enrolled in high-risk screening programs across a few centers in the United States.

Two well-known cohort studies (the Nurses' Health Study and the Health Professionals Follow-up Study) showed that the risk of pancreatic cancer increased with obesity and decreased with more physical activity. The obese cohort (BMI of 30 kg m^{-2}) had a 1.72 times increased relative risk (RR) of pancreatic cancer as compared with the normal weight cohort (BMI of 23 kg m^{-2}). An inverse relation was observed for moderate physical activity (RR 0.45) for the highest versus lowest categories [41]. These conclusions corroborate with other studies which also show an increased risk of pancreatic cancer with obesity and a protective effect with physical activity [42–44].

Minor risk factors and association

A multitude of minor risk factors have been identified to increase the rate of pancreatic cancer. A well-known association is diabetes mellitus [45–47]. Multiple studies have shown that the presence of long-standing diabetes mellitus modestly increases the risk of pancreatic cancer, and this was demonstrated in a meta-analysis of these studies from 1974–1994 [48]. The pooled risk ratio of pancreatic cancer for diabetics relative to nondiabetics was 2.1 (95 % CI 1.6–2.8). This risk was similar when only patients with diabetes for more than 5 years were considered [48]. It is important to note that the association between diabetes mellitus and pancreatic cancer may be confounded by the fact that new-onset diabetes may be an early manifestation of pancreatic cancer (i.e. the cancer causes the diabetes) [48–50]. When tested, up to 80 % of pancreatic cancer patients have glucose intolerance [50,51]. Thus, most of the diabetes seen

in pancreatic cancer patients is due to the cancer. This form of diabetes is called pancreatic cancer-associated diabetes (PacDM). It is unique in that this diabetes continues to worsen despite the dramatic weight loss seen in pancreatic cancer [50].

It has been observed that people with blood group O may have a lower risk of pancreatic cancer than those with groups A or B [52–55]. This association is consistent and may hold some insights in to inheritance pattern of pancreatic cancer. Postsurgical states such as partial gastrectomy and history of cholecystectomy, to varying degrees, have been associated with pancreatic cancer [56–58]. Finally, *H. pylori* infection has been shown in some studies to increase the risk of pancreatic cancer, especially the CagA strain [59,60].

Studies looking at red meat intake have been conflicting when it comes to added risk of pancreatic cancer. Increased fruit and vegetable intake has not been shown to be protective in a prospective study [61–64]. Coffee consumption has been debunked as a causal factor for pancreatic cancer after a flurry of initial reports [16]. Alcohol consumption is unlikely to increase the risk of pancreatic cancer. Multiple large studies suggest that there is no added risk including a recent large European study of 478,400 subjects [65–68].

Areas for further study

One of the biggest challenges in the field of pancreatic cancer, as of today, is the need to diagnose/screen for asymptomatic pancreatic cancer. As alluded to earlier, by the time cancer-specific symptoms such as abdominal pain, loss of appetite, or weight loss occur, the cancer is too far advanced to alter its natural course. Conversely, it has been shown that the smaller the size of the tumor (<2 cm) at diagnosis, in addition to being surgically resectable, the 5-year survival is also better [69]. Unfortunately there is no established screening modality for detecting asymptomatic pancreatic cancer. This need is of great significance to high-risk kindred with multiple first-degree relatives affected by pancreatic cancer. Studies of such families have demonstrated an increasing risk profile with more first-degree relatives afflicted by pancreatic cancer [70]. Screening for pancreatic cancer is further compounded by the fact that there is no reliable serum marker for resectability. In this regard, new-onset diabetes, defined as diabetes diagnosed within 24 months of a diagnosis of pancreatic cancer, may offer a potential lead. It has long been recognized that new-onset diabetes/impaired fasting glucose often precedes any symptoms of pancreatic cancer such as weight loss or abdominal pain. Despite this observation, screening protocols outside the purview of clinical trials do not exist as of today.

Multiple choice questions

1 Which of the following best define the symptoms of pancreatic cancer?

A Pancreatic cancer symptoms occur late in the disease

B Pancreatic cancer symptoms are unreliable for early diagnosis

C Pancreatic cancer symptoms are vague and can mimic other conditions

D Severe daily abdominal pain needing narcotic pain medications is a bad prognostic factor

E All of the above

2 Mr. John Smith is a 55-year-old gentleman who presents to you because he is worried about getting pancreatic cancer. His father and aunt died of pancreatic cancer at age 83 and 64, respectively. He has smoked one pack of cigarettes a day for the past 30 years and has consumed two mixed alcoholic drinks a day for a similar amount of time. He has hypertension and type 2 diabetes mellitus, which are well controlled. He does not exercise on a regular basis and says he does not enjoy eating a lot of fruits and vegetables. His BMI is 34.

Which of the following interventions has the greatest impact in altering the patient's risk for pancreatic cancer?

A Schedule him for an endoscopic ultrasound now and check serum Ca 19-9

B Schedule him for a CT scan of the abdomen and check serum Ca 19-9

C Smoking cessation

D Increase his fruit and vegetable intake and avoid red meat

E Diet and exercise regime to decrease his BMI to 24

3 Mrs. Jane Dune is a 48-year-old female who was recently diagnosed with metastatic pancreatic cancer. She is completely asymptomatic except 2/10

abdominal pain which is relieved by two tablets of extra strength Tylenol. She wants to do everything in her power to "beat the cancer" as she has two young children.

The treatment option most likely to prolong her survival is

A Neoadjuvant chemotherapy followed by a pylorus-preserving Whipple operation

B Chemotherapy with single agent Gemcitabine with radiation

C Chemotherapy with FOLFIRNOX with no radiation

D Chemotherapy with Erlotinib along with radiation

E Celiac Ganglia neurolysis followed by chemotherapy with Gemcitabine, Erlotinib, and radiation therapy

4 James Smith is a 45-year-old male who has a family history of pancreatic cancer. His mother and paternal grandfather died of pancreatic cancer. He is asymptomatic but wants to know if you "can run some test" to see if he has pancreatic cancer. Which of the following would you recommend:

A Endoscopic ultrasound (EUS)

B EUS and serum Ca 19-9

C EUS, serum Ca 19-9, and blood sugar

D No specific testing

E EUS, serum Ca 19-9, and HbA1c

5 Mary Joe is a healthy 76-year-old who was recently hospitalized for acute pancreatitis. She had a cholecystectomy 35 years ago, does not smoke or consume alcohol in excess. Her calcium, Ca 19-9 and triglyceride levels were normal and she does not take any prescription medications or supplements. Thus a preliminary work-up during her hospitalization did not reveal a cause for the pancreatitis. She had mild intestinal pancreatitis on the CT scan on the day of her admission. She is now ready to be discharged after spending 3 days in the hospital. What follow-up test would you recommend?

A EUS/CT with repeat Ca-19-9 in 6 weeks

B PET scan in 6 weeks

C Recheck amylase and lipase in 6 weeks

D No specific test at this time

References

1 Klimstra DS. Nonductal neoplasms of the pancreas. *Mod Pathol* 2007;20(Suppl 1):S94–112.

2 Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. *CA Cancer J Clin* 2010;60(5):277–300.

3 Bray F, Ren JS, Masuyer E, et al. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer* 2012 Jul 3.

4 Geer RJ, Brennan MF. Prognostic indicators for survival after resection of pancreatic adenocarcinoma. *Am J Surg* 1993;165(1):68–72; discussion 72–3.

5 Bakkevold KE, Arnesjo B, Dahl O, et al. Adjuvant combination chemotherapy (AMF) following radical resection of carcinoma of the pancreas and papilla of Vater – results of a controlled, prospective, randomised multicentre study. *Eur J Cancer* 1993;29A(5):698–703.

6 Cameron JL, Riall TS, Coleman J, et al. One thousand consecutive pancreaticoduodenectomies. *Ann Surg* 2006;244(1):10–15.

7 Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus Gemcitabine for metastatic pancreatic cancer. *New Engl J Med* 2011;364(19):1817–25.

8 Bilimoria KY, Bentrem DJ, Ko CY, et al. Validation of the 6th edition AJCC Pancreatic Cancer Staging System: report from the National Cancer Database. *Cancer* 2007;110(4):738–44.

9 Edge SB Byrd DR, Compton CC, et al. (eds.) (American Joint Committee on Cancer) (2010) *AJCC Cancer Staging Manual*, Springer, New York.

10 Mack TM, Yu MC, Hanisch R, et al. Pancreas cancer and smoking, beverage consumption, and past medical history. *J Natl Cancer Inst* 1986;76(1):49–60.

11 Doll R, Peto R. Mortality in relation to smoking: 20 years' observations on male British doctors. *BMJ* 1976;2(6051):1525–36.

12 Farrow DC, Davis S. Risk of pancreatic cancer in relation to medical history and the use of tobacco, alcohol and coffee. *Int J Cancer* 1990;45(5):816–20.

13 Lynch SM, Vrieling A, Lubin JH, et al. Cigarette smoking and pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium. *Am J Epidemiol* 2009;170(4):403–13.

14 Silverman DT, Dunn JA, Hoover RN, et al. Cigarette smoking and pancreas cancer: a case-control study based on direct interviews. *J Natl Cancer Inst* 1994;86(20):1510–6.

15 Fuchs CS, Colditz GA, Stampfer MJ, et al. A prospective study of cigarette smoking and the risk of pancreatic cancer. *Arch Intern Med* 1996;156(19):2255–60.

16 Lowenfels AB, Maisonneuve P, Whitcomb DC, et al. Cigarette smoking as a risk factor for pancreatic cancer in patients with hereditary pancreatitis. *JAMA* 2001;286(2):169–70.

17 Whitcomb DC, Gorry MC, Preston RA, et al. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet* 1996;14(2):141–5.

18 Lowenfels AB, Maisonneuve P, DiMagno EP, et al. Hereditary pancreatitis and the risk of pancreatic cancer. International Hereditary Pancreatitis Study Group. *J Natl Cancer Inst* 1997;89(6):442–6.
19 Rebours V, Boutron-Ruault MC, Schnee M, et al. Risk of pancreatic adenocarcinoma in patients with hereditary pancreatitis: a national exhaustive series. *Am J Gastroenterol* 2008;103(1):111–9.
20 Howes N, Lerch MM, Greenhalf W, et al. Clinical and genetic characteristics of hereditary pancreatitis in Europe. *Clin Gastroenterol Hepatol* 2004;2(3):252–61.
21 Whitcomb DC, Applebaum S, Martin SP. Hereditary pancreatitis and pancreatic carcinoma. *Ann N Y Acad Sci* 1999;880:201–9.
22 Lowenfels AB, Maisonneuve P, Whitcomb DC. Risk factors for cancer in hereditary pancreatitis. International Hereditary Pancreatitis Study Group. *Med Clin North Am* 2000;84(3):565–73.
23 Lowenfels AB, Maisonneuve P, Cavallini G, et al. Pancreatitis and the risk of pancreatic cancer. International Pancreatitis Study Group. *New Engl J Med* 1993;328(20):1433–7.
24 Malka D, Hammel P, Maire F, et al. Risk of pancreatic adenocarcinoma in chronic pancreatitis. *Gut* 2002;51(6):849–52.
25 Goldacre MJ, Wotton CJ, Yeates D, et al. Liver cirrhosis, other liver diseases, pancreatitis and subsequent cancer: record linkage study. *Eur J Gastroenterol Hepatol* 2008;20(5):384–92.
26 Kudo Y, Kamisawa T, Anjiki H, et al. Incidence of and risk factors for developing pancreatic cancer in patients with chronic pancreatitis. *Hepatogastroenterology* 2011;58(106):609–11.
27 Raimondi S, Lowenfels AB, Morselli-Labate AM, et al. Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection. *Best Pract Res Clin Gastroenterol* 2010;24(3):349–58.
28 Ekbom A, McLaughlin JK, Karlsson BM, et al. Pancreatitis and pancreatic cancer: a population-based study. *J Natl Cancer Inst* 1994;86(8):625–7.
29 Logsdon CD, Ji B. Ras activity in acinar cells links chronic pancreatitis and pancreatic cancer. *Clin Gastroenterol Hepatol* 2009;7(11 Suppl):S40–3.
30 Ji B, Tsou L, Wang H, et al. Ras activity levels control the development of pancreatic diseases. *Gastroenterology* 2009;137(3):1072–82, 82 e1–6.
31 Chari ST, Mohan V, Pitchumoni CS, et al. Risk of pancreatic carcinoma in tropical calcifying pancreatitis: an epidemiologic study. *Pancreas* 1994;9(1):62–6.
32 Giardiello FM, Brensinger JD, Tersmette AC, et al. Very high risk of cancer in familial Peutz–Jeghers syndrome. *Gastroenterology* 2000;119(6):1447–53.
33 Su GH, Hruban RH, Bansal RK, et al. Germline and somatic mutations of the STK11/LKB1 Peutz–Jeghers gene in pancreatic and biliary cancers. *Am J Pathol* 1999;154(6):1835–40.
34 Murphy KM, Brune KA, Griffin C, et al. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. *Cancer Res* 2002;62(13):3789–93.
35 Hahn SA, Greenhalf B, Ellis I, et al. BRCA2 germline mutations in familial pancreatic carcinoma. *J Natl Cancer Inst* 2003;95(3):214–21.
36 Hussussian CJ, Struewing JP, Goldstein AM, et al. Germline p16 mutations in familial melanoma. *Nat Genet* 1994;8(1):15–21.
37 Goldstein AM, Fraser MC, Struewing JP, et al. Increased risk of pancreatic cancer in melanoma-prone kindreds with p16INK4 mutations. *New Engl J Med* 1995;333(15):970–4.
38 Giardiello FM, Offerhaus GJ, Lee DH, et al. Increased risk of thyroid and pancreatic carcinoma in familial adenomatous polyposis. *Gut* 1993;34(10):1394–6.
39 Kastrinos F, Mukherjee B, Tayob N, et al. Risk of pancreatic cancer in families with Lynch syndrome. *JAMA* 2009;302(16):1790–5.
40 Lynch HT, Smyrk T, Kern SE, et al. Familial pancreatic cancer: a review. *Semin Oncol* 1996;23(2):251–75.
41 Michaud DS, Giovannucci E, Willett WC, et al. Physical activity, obesity, height, and the risk of pancreatic cancer. *JAMA* 2001;286(8):921–9.
42 Arslan AA, Helzlsouer KJ, Kooperberg C, et al. Anthropometric measures, body mass index, and pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium (PanScan). *Arch Intern Med* 2010;170(9):791–802.
43 O'Rorke MA, Cantwell MM, Cardwell CR, et al. Can physical activity modulate pancreatic cancer risk? A systematic review and meta-analysis. *Int J Cancer* 2010;126(12):2957–68.
44 Li D, Morris JS, Liu J, et al. Body mass index and risk, age of onset, and survival in patients with pancreatic cancer. *JAMA* 2009;301(24):2553–62.
45 Wideroff L, Gridley G, Mellemkjaer L, et al. Cancer incidence in a population-based cohort of patients hospitalized with diabetes mellitus in Denmark. *J Natl Cancer Inst* 1997;89(18):1360–5.
46 Stevens RJ, Roddam AW, Beral V. Pancreatic cancer in type 1 and young-onset diabetes: systematic review and meta-analysis. *Br J Cancer* 2007;96(3):507–9.
47 Jee SH, Ohrr H, Sull JW, et al. Fasting serum glucose level and cancer risk in Korean men and women. *JAMA* 2005;293(2):194–202.

48 Everhart J, Wright D. Diabetes mellitus as a risk factor for pancreatic cancer. A meta-analysis. *JAMA* 1995;273(20):1605–9.
49 Gullo L, Pezzilli R, Morselli-Labate AM. Diabetes and the risk of pancreatic cancer. *New Engl J Med* 1994;331(2):81–4.
50 Pannala R, Leirness JB, Bamlet WR, et al. Prevalence and clinical profile of pancreatic cancer-associated diabetes mellitus. *Gastroenterology* 2008;134(4):981–7.
51 Permert J, Ihse I, Jorfeldt L, et al. Pancreatic cancer is associated with impaired glucose metabolism. *Eur J Surg* 1993;159(2):101–7.
52 Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet* 2009;41(9):986–90.
53 Wolpin BM, Kraft P, Gross M, et al. Pancreatic cancer risk and ABO blood group alleles: results from the pancreatic cancer cohort consortium. *Cancer Res* 2010;70(3):1015–23.
54 Wolpin BM, Chan AT, Hartge P, et al. ABO blood group and the risk of pancreatic cancer. *J Natl Cancer Inst* 2009;101(6):424–31.
55 Annese V, Minervini M, Gabbrielli A, et al. ABO blood groups and cancer of the pancreas. *Int J Pancreatol* 1990;6(2):81–8.
56 Offerhaus GJ, Tersmette AC, Tersmette KW, et al. Gastric, pancreatic, and colorectal carcinogenesis following remote peptic ulcer surgery. Review of the literature with the emphasis on risk assessment and underlying mechanism. *Mod Pathol* 1988;1(5):352–6.
57 Ekbom A, Yuen J, Karlsson BM, et al. Risk of pancreatic and periampullar cancer following cholecystectomy: a population-based cohort study. *Dig Dis Sci* 1996;41(2):387–91.
58 Chow WH, Johansen C, Gridley G, et al. Gallstones, cholecystectomy and risk of cancers of the liver, biliary tract and pancreas. *Br J Cancer* 1999;79(3–4):640–4.
59 Stolzenberg-Solomon RZ, Blaser MJ, Limburg PJ, et al. *Helicobacter pylori* seropositivity as a risk factor for pancreatic cancer. *J Natl Cancer Inst* 2001;93(12):937–41.
60 Raderer M, Wrba F, Kornek G, et al. Association between *Helicobacter pylori* infection and pancreatic cancer. *Oncology* 1998;55(1):16–9.
61 Michaud DS, Giovannucci E, Willett WC, et al. Dietary meat, dairy products, fat, and cholesterol and pancreatic cancer risk in a prospective study. *Am J Epidemiol* 2003;157(12):1115–25.
62 Michaud DS, Skinner HG, Wu K, et al. Dietary patterns and pancreatic cancer risk in men and women. *J Natl Cancer Inst* 2005;97(7):518–24.
63 Nothlings U, Wilkens LR, Murphy SP, et al. Meat and fat intake as risk factors for pancreatic cancer: the multiethnic cohort study. *J Natl Cancer Inst* 2005;97(19):1458–65.
64 Silverman DT, Swanson CA, Gridley G, et al. Dietary and nutritional factors and pancreatic cancer: a case-control study based on direct interviews. *J Natl Cancer Inst* 1998;90(22):1710–9.
65 Rohrmann S, Linseisen J, Vrieling A, et al. Ethanol intake and the risk of pancreatic cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Cancer Causes Control* 2009;20(5):785–94.
66 Coughlin SS, Calle EE, Patel AV, et al. Predictors of pancreatic cancer mortality among a large cohort of United States adults. *Cancer Causes Control* 2000;11(10):915–23.
67 Hirayama T. Epidemiology of pancreatic cancer in Japan. *Jpn J Clin Oncol* 1989;19(3):208–15.
68 Michaud DS, Giovannucci E, Willett WC, et al. Coffee and alcohol consumption and the risk of pancreatic cancer in two prospective United States cohorts. *Cancer Epidemiol Biomarkers Prev* 2001;10(5):429–37.
69 Tsuchiya R, Noda T, Harada N, et al. Collective review of small carcinomas of the pancreas. *Ann Surg* 1986;203(1):77–81.
70 Brentnall TA, Bronner MP, Byrd DR, et al. Early diagnosis and treatment of pancreatic dysplasia in patients with a family history of pancreatic cancer. *Ann Intern Med* 1999;131(4):247–55.

Answers to multiple choice questions

1. E
2. C

At this time, there is no proven screening test for pancreatic cancer hence choices A and B are not correct. Although choices D and E are seemingly reasonable as part of general healthy living, in this instance the intervention that has the greatest impact in altering this person's risk for pancreatic cancer is smoking cessation. It is estimated that up to 30 % of cases of pancreatic cancer can be attributed to smoking.

3. Option C

This patient has "metastatic disease" hence option A, which is for local disease, is not valid. The patient's pain is well controlled (no requirement for celiac ganglia neurolysis) and additionally there is no data to suggest the use of erlotinib in this patient, hence options D and E are not valid. As she is an otherwise healthy patient with no other comorbidities, option C

offers her the best odds to prolong survival as compared to option B based on recent data, albeit with more side effects.

4. D

At this time, there is no proven screening test for pancreatic cancer, hence choices A, B, C, and E are invalid. The best option for asymptomatic relatives of patient with familial pancreatic cancer is to enroll in clinical registries.

5. A

This patient is an otherwise healthy woman who has no clear cause for her acute pancreatitis. Hence option A is a prudent choice as part of her follow-up care as pancreatitis can be the presenting symptom of pancreatic cancer especially in the elderly. A PET scan and rechecking amylase and lipase in 6 weeks will likely not detect the tumor or afford a chance for tissue sampling, and thus are suboptimal options.

29 Epidemiology of hepatitis B and C in the United States

Sumeet K. Asrani[1] *& W. Ray Kim*[2]

[1]Hepatology, Baylor University Medical Center, Dallas, TX, USA
[2]GI Epidemiology/Outcomes Unit, Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

Key points

- Both hepatitis B and C viruses (HBV and HCV) are transmitted parenterally via infected blood or body fluids and may be commonly transmitted by contaminated needles and unprotected sexual contacts. Perinatal exposure is also an important means of transmission, especially for hepatitis B in endemic populations.
- In the United States (USA), the incidence of new infections with HBV and HCV has been decreasing in the past two decades, largely due to safer needle-using practices and universal precaution in health care as well as exclusion of blood donors with infection. For hepatitis B, widespread vaccination programs have been effective in reducing its incidence in children.
- Despite these decreases in acute infections, the prevalence and burden of chronic HBV and HCV infection remain substantial in the USA. Population-based prevalence estimates indicate that there are around 3–4 million persons in the USA with chronic HBV and HCV infection.
- Globally, 600,000 and 350,000 deaths are due to HBV and HCV, respectively. In the USA, an estimated 3000 and 12,000 annual deaths are attributed to HBV and HCV, respectively.

Disease definitions and clinical diagnosis

Hepatitis B virus

Hepatitis B virus (HBV) is a DNA virus that belongs in the hepadna virus family. Infected hepatocytes produce at least three types of viral proteins which are utilized in the diagnosis of HBV infection (Table 29.1). The S protein constitutes the viral envelope and is detected as HBV surface antigen (HBsAg) in the serum. The C protein, a component of nucleocapsid of the virus, remains within hepatocytes and is not detectable in the serum. However, antibodies against this protein, namely, anti-HBc, are a marker of exposure to the virus. HBeAg consists of the C protein and pre-C protein. Presence of HBeAg connotes active replication of the virus. Patients who lack HBeAg usually have detectable antibodies against it in the serum (anti-HBe), which indicates either suppression of viral replication by the host immune system or presence of the so-called pre-core mutation, which allows active replication of the virus while not producing the pre-C protein. The most accurate marker of HBV replication, however, is the serum level of HBV DNA. Classically, serum levels $>10^5$ copies ml^{-1} has been understood to represent active viral replication, although more recent data indicate that liver damage occurs at lower levels.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 29.1 Diagnostic testing for HBV and HCV infection

Name	Marker
HBsAg	Active infection (acute or chronic)
Anti-HBs	Immunity to HBV infection
Anti-HBc (total)	Exposure to HBV
HBeAg	Evidence of active HBV replication
Anti-HBe	Low replication or precore mutant
HBV DNA	Quantifies amount of virus (replication)
Test	**Type of test**
HCV antibody (EIA)	Screening
HCV RNA (Qualitative)	Confirmatory test
HCV RNA (Quantitative)	Pre- and intra-treatment test to assess response to therapy
Genotype	Pre-treatment test to determine duration of treatment
RIBA	Confirmation of positive anti-HCV antibody (rarely used clinically)

Hepatitis C virus

Hepatitis C virus (HCV) is an RNA virus that belongs in the flavivirus family, along with dengue fever and yellow fever. The proteins generated by HCV may be structural (envelope and core) and nonstructural (polymerase, protease, etc.). The initial test in the detection of HCV infection utilizes anti-HCV antibodies directed against the core and nonstructural proteins. Currently used anti-HCV testing is highly sensitive and specific for these antibodies. Detection of HCV RNA is the hallmark of the infection with HCV. In population-based studies, 20–35 % of subjects who have anti-HCV do not have detectable HCV RNA in the serum, which indicates previous exposure to HCV and recovery thereof. These individuals test positive to RIBA (radioimmunoblot assays), as opposed to individuals in whom anti-HCV is false positive.

Transmission of hepatitis B and C

Both HBV and HCV are transmitted parenterally, that is, by exposure to blood, blood products and tissue. The incubation period of hepatitis B is 6–24 weeks (average 16 weeks) and that of HCV 3–12 weeks (average 7 weeks) [1,2].

Hepatitis B

HBV is transmitted by percutaneous and mucous membrane exposures to infectious body fluids, such as serum, semen, and saliva. Perinatal transmission is thought to be a major route by which HBV infection perpetuates in endemic countries. The risk of transmission in general correlates with the HBV DNA level in the maternal serum [3,4]. The risk is greatest for infants born to women who are HBeAg-positive with high levels of HBV DNA (often >100 million copies ml^{-1}); in those children, 70 to 90 % are HBsAg-positive at 6 months of age. The risk in infants born to mothers with negative HBeAg (and low levels of HBV DNA) ranges from 10 to 40 %. Fortunately, the risk of perinatal HBV transmission can be significantly reduced by passive and active immunizations. Although HBsAg has been found in breast milk, breast feeding by an HBsAg-positive mother has not been shown to pose an additional risk for the acquisition of HBV.

Children born to HBsAg-positive mothers who do not become infected during the perinatal period remain at risk of infection during early childhood [5]. Up to 40 % of infants born to HBeAg-negative mothers may become infected by 5 years of age. In this setting, "horizontal" transmission of HBV is known to occur during early childhood, in addition to the potential mother-to-child transmission. Although the exact mechanism by which this occurs is unknown, frequent interpersonal contacts of nonintact skin or mucous membranes with blood-containing secretions or saliva is likely the route of transmission. Because the concentration of virus in the blood is often extremely high in children and because HBV remains infectious on environmental surfaces for long periods of time (>1 week) under ambient conditions, indirect inoculation of HBV through inanimate objects may occur among children relatively efficiently.

Among adults, high-risk sexual activity is one of the most frequent routes of transmission for HBV [6]. Although homosexual men were one of the groups at highest risk for HBV infection, historically heterosexual transmission is the most common cause of acute HBV infection in adults. Factors associated with an increased risk of HBV infection among heterosexual

men and women include number of sexual partners, number of years of sexual activity, and history of other sexually transmitted diseases. Thus, transmission of HBV from persons with acute or chronic hepatitis B to their homo- or heterosexual partners is an important source of infection, because most persons with chronic HBV infection are not aware that they are infected.

Transmission of HBV via transfusion of blood and plasma-derived products has been all but eliminated in most countries through donor screening for HBsAg and viral inactivation procedures. However, transmission of HBV may continue to occur in other healthcare settings. For example, transmission of HBV among chronic hemodialysis patients may occur when appropriate isolation guidelines are not followed, which includes using dedicated equipment and staff in a separate room for patients with chronic HBV infection. In addition to contamination of instruments and equipment, direct person-to-person exposure may transmit HBV [7]. Finally, nonsexual interpersonal transmission of HBV can occur, such as long-term household (or institutional) contacts of chronically infected person(s) contact over a long period of time. The precise mechanisms of transmission are unknown, but it may mirror the spread of HBV among children as described above.

Hepatitis C

With regard to HCV, blood transfusion before 1992 and injection drug use have historically been the two most important risk factors in the United States. Presently, however, injection drug use is by far the most common route of transmission for HCV. In a recent report based on the National Health and Nutrition Examination Survey (NHANES), 58 % of participants aged 20–59 years who had used illicit drugs (excluding marijuana) were positive for anti-HCV (149 times more likely to have positive anti-HCV compared to those with drug-use history) [8]. Respondents with ≥20 lifetime sexual partners were five times more likely to be anti-HCV-positive compared to those with 0–1 partners. Other factors associated with positive anti-HCV included age at first sexual encounter, lower family income and education, a positive antibody to HSV-2, as well as a history of blood transfusion prior to 1992. Among persons aged 20–59 with HCV infection, 99 % had one of the following risk factors: (i) a history of illicit drug use (other than marijuana), (ii) transfusion prior to 1992, (iii) ≥20 lifetime sexual partners, or (iv) abnormal ALT [8].

Like HBV, HCV may be transmitted in the perinatal period from infected mother to the newborn. The risk of transmission is lower for HCV than HBV: less than 6 % of babies born to an infected mother have been reported to acquire the infection [9]. Co-infection with HIV increases the risk of perinatal HCV transmission. Limited data suggest that HCV is not transmitted from mother to baby by breastfeeding. Unfortunately, there is no known means to reduce the risk of transmission from mother to child.

Incidence

Hepatitis B

HBV and HCV are reportable infectious diseases in the USA and the Centers for Disease Control and Prevention (CDC) have put in place mechanisms to capture incident cases of HCV infection. These include passive surveillance programs such as the National Notifiable Disease Surveillance System and hepatitis-specific active surveillance programs such as the Sentinel Counties Study of Acute Viral Hepatitis.

Globally, there are approximately 4 million cases of acute hepatitis B [10]. In the USA, according to the CDC, the incidence of acute hepatitis B has steadily declined over the last two decades (Figure 29.1) [11]. The number of reported cases of acute hepatitis B decreased by 84.2 % from 21,277 cases (8.5 per 100,000) in 1990 to 3371 cases in 2009 (1.1 cases per 100,000). The incidence decreased across all age groups and the greatest decline was seen among persons aged 20–39 years. In 2009, the highest rate was observed among persons aged 30–39 years (2.3 cases per 100,000). The incidence of acute hepatitis B among men has been consistently higher than among women, though the gap has recently narrowed. In 2009, the incidence was 1.6 times higher in men as compared to women (1.4 and 0.8 cases per 100,000, respectively). The rates for acute hepatitis B decreased for all race/ethnicity groups between 1990 and 2009. In 2009, the incidence of acute hepatitis B was highest for non-Hispanic Blacks (1.7 cases per 100,000) and lowest for Asian Pacific Islander (API) and Hispanics (0.7 cases per 100,000 in both groups).

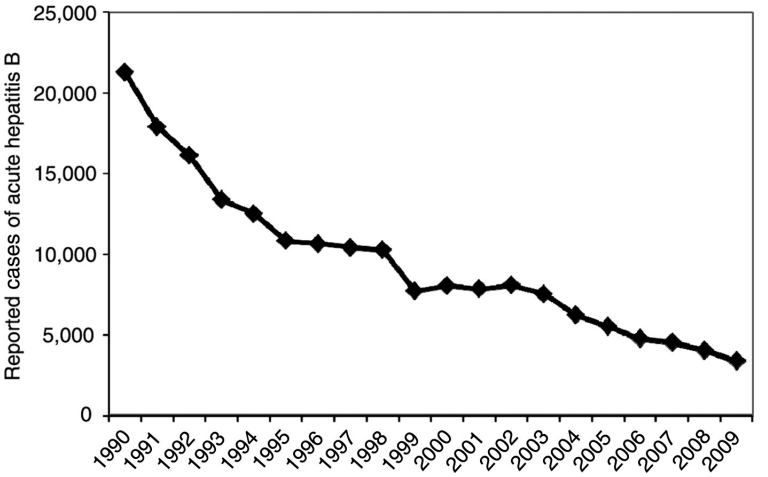


Figure 29.1 Reported incidence of acute hepatitis B. Source: Zanetti et al. 2008 [10].

The reduction in HBV incidence in the USA may be attributed to several measures implemented in 1991, which include universal infant vaccination, universal screening of pregnant women and post-exposure prophylaxis of infants born to infected mothers [1]. Between 1995 and 1999, the immunization strategy was expanded to include vaccination of all persons aged 0–18 years who have not been vaccinated previously. HBV vaccination increased from 73 % of persons aged 6–19 in 1999–2002 to 91 % between 2007 and 2008 [12]. The most common risk factors reported among adults with acute hepatitis B continue to be multiple sex partners, homosexual activity, and injection-drug use.

These trends are also seen globally. Immunization is provided globally for about 60 % of the world population (2006) as compared to 1 % in 1990. There has been a significant reduction in the incidence of acute hepatitis B infection, carrier rate in persons who are immunized, and a reduction in mortality related to hepatitis B. For example, in Italy the incidence of acute hepatitis B decreased from 11/100,000 in 1987 to 1.6/100,000 in 2006 [10].

Hepatitis C

The incidence of new HCV infection is very difficult to estimate accurately. According to the World Health Organization, approximately 3–4 million people are infected with HCV each year. This is because many patients with acute HCV infection are asymptomatic and thus do not present themselves for diagnosis. Underreporting by healthcare providers of diagnosed cases is also thought to be common. Furthermore, individuals at high risk of infection may not have ready access to health care, decreasing the likelihood of timely diagnosis of newly acquired HCV infection. Because of these limitations, enumerating reported cases of acute hepatitis C significantly underestimates the true incidence of hepatitis C infection [13].

Given these limitations, the number of reported cases of acute hepatitis C in the USA (Figure 29.2) decreased 87 % from 6010 in 1992 to 781 in 2009 (0.3 per 100,000) [11]. The incidence decreased for all groups. In 2009, the rates of acute hepatitis C

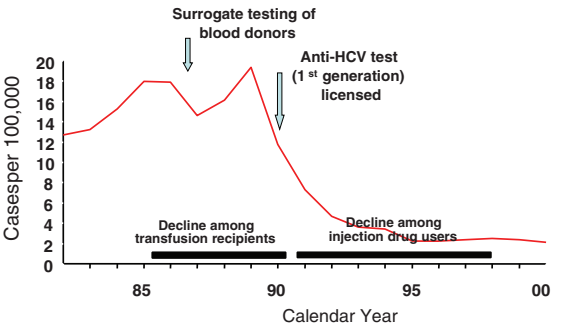


Figure 29.2 Estimated incidence of acute HCV infection in the United States. Source: Kim 2002 [13].

were highest among persons aged 20–29 years (0.7 per 100,000). The incidence decreased for both men and women. As compared to hepatitis B, the rates of acute hepatitis C were similar for men and women in 2009 (0.3 per 100,000). Similar to hepatitis B, the rates for acute hepatitis C decreased for all race/ethnicity groups from 1992–2009. In 2009, the incidence of acute hepatitis C was highest among American Indian/Alaska Natives, AI/AN (0.5 per 100,000) and lowest for APIs (0.04 per 100,000) [14].

The number of persons with transfusion-associated HCV infection decreased significantly since the introduction in 1985 of guidelines for selecting safer blood donors. It declined further with the institution of screening of blood donors for anti-HCV beginning in 1989 with the first-generation test and with the second-generation assay introduced in July 1992. Much of the recent decline in incidence can be accounted for by a decline in cases among injecting drug users which may be related to safer needle-using practices. Trends in other risk factors, including sexual, household and occupational exposures, have remained relatively stable over time.

Prevalence of HBV and HCV

On a global scale, HBV is vastly more common than HCV. More than 2 billion people in the world have been infected with HBV with active infection being present in over 350 million [15]. This compares to an estimated 170 million people currently infected with HCV [16]. The geographic distribution of HBV and HCV is not uniform. HBV is most common in the Far East and Southeast Asia, sub-Saharan Africa, the Amazon basin, and eastern Europe. HCV is more evenly distributed throughout the world than HBV and a significant number of people from North America and Europe have the infection [16].

The NHANES have been a valuable tool in estimating the prevalence of hepatitis B and C in the USA. The NHANES are a series of cross-sectional national surveys designed to provide representative prevalence estimates for a variety of health measures and conditions. Each survey is designed to be representative of the US civilian noninstitutionalized population. In studying the epidemiology of viral hepatitis, NHANES conducted in five periods have been used. The first was conducted between 1976 and 1980, the second between 1988 and 1994, the third between1999 and 2002, the fourth between 2003 and 2006, and the fifth between 2007 and 2008 [6].

Hepatitis B

The overall prevalence of hepatitis B core antibody (marker of exposure to the virus which does not distinguish between prior infection or chronic active infection) decreased from 5.4 % (1988–1994) to 3.6 % (2007–2008). The prevalence of HBV infection decreased from 1988–1994 to 2007–2008 for all ages. The decrease was significant for all age groups except for persons aged 40–59 years. In this age group, the prevalence of HBV infection has stayed around 6–7 % over the last two decades. From 1988–1994 to 2007–2008, prevalence of HBV infection decreased for both men (6.4 % to 4 %) and women (4.5 % to 3.2 %). Though a decrease was seen across all races/ethnicities, between 2007–2008 the prevalence was highest among non-Hispanic Blacks (9.1 %) as compared to Hispanic Americans (2.7 %) or non-Hispanic Whites(1.8 %). As expected, persons born outside the USA had a higher prevalence of hepatitis B core antibody as compared to US-born persons (10.1 % vs. 2.4 %).

The prevalence of chronic HBV infection was similar across the NHANES over the last two decades. Similar estimates of age-adjusted prevalence of HBsAg-positive individuals were observed in 1988–1994 (0.38 %) and during 1999–2006 (0.27 %) suggesting that there are 730,000 infected persons (95 % CI 550,000–940,000) [17]. The prevalence was higher among persons older than 50 years (0.5 %) and higher among men as compared to women (0.35 vs. 0.19 %). It was higher among persons classified as others (0.98 %) as compared to non-Hispanic Whites (0.09 %) and higher among foreign-born (0.89 %) as compared to US-born (0.16 %) individuals. In the setting of increased immunizations for children, the prevalence of HBV decreased among children in the USA but remained relatively unchanged among adults [17].

While the NHANES data are useful in the estimation of HBV prevalence in the USA in general, the surveys did not include statistically valid samples from populations in which HBV is most common, such as Asians, Pacific Islanders, and Alaskan Natives [6,18]. Thus, NHANES likely represent an underestimate of

the true prevalence of HBV in the USA. In Asian and western Pacific countries where HBV is endemic, estimated prevalence of chronic HBV infection ranges from 2.4–16 %. In a large systematic review, the prevalence of HBsAg among intravenous drug users was 5–10 % and upwards of 10 % in 10 countries suggesting that worldwide 6.4 million intravenous drug users are anti-HBc positive and 1.2 million have evidence of chronic infection (HBsAg positive) [19].

A recent survey assessed the prevalence of chronic HBV infection among Asian/Pacific Islander (A/PI) populations living in New York City [20]. Of 925 survey participants who reported not having been tested previously for HBV infection, 137 (14.8 %) were HBsAg-positive, whereas another 496 (53.6 %) had evidence of resolved HBV infection. The prevalence of chronic HBV infection was higher among males (19.7 %) compared to females (8.7 %) and among persons aged 20–39 years (23.2 %) compared to those aged >40 years (9.6 %). Prevalence of chronic HBV infection varied by country of birth, from 21.4 % among those born in China, to 4.6 % among those born in South Korea, to 4.3 % among those born in other Asian countries. Although this study was limited to New York City, screening programs in Atlanta, Chicago, New York City, Philadelphia, and California have reported similar prevalence of chronic HBV infection (10–15 %) among A/PI immigrants to the USA, pointing to a disproportionate burden of chronic HBV infection among A/PI and other immigrant populations.

Hepatitis C

The prevalence of HCV infection (antibodies to HCV) has been examined over serial NHANES. Overall, the prevalence decreased from 1.8 % (1988–1994) to 1.3 % (2007–2008), implying that an estimated 4 million people are anti-HCV positive. In 2007–2008, the prevalence of anti-HCV was higher among men (1.6 % vs. 1 % for women) and higher for non-Hispanic Blacks (2.6 %) as compared to non-Hispanic Whites (1.3 %) or Mexican Americans (1.1 %). Estimates for number of persons who have chronic HCV infection (HCV RNA positive) varies from 2.7–3.9 million people [8,21].

The comparison between the two estimates reveals that little change occurred in the prevalence of chronic HCV during the 1990s. While it lends support to the data indicating a low incidence of new HCV infection, it also indicates that advances in HCV therapy have not made a demonstrable impact in reducing the burden of chronic HCV infection at the population level.

According to the recent NHANES data, HCV prevalence has increased linearly with age with peak prevalence in the age group of 40 to 49 years. Within this age group, non-Hispanic Blacks had a higher prevalence at 9.4 % compared to non-Hispanic Whites at 3.8 % ($P < 0.001$). A birth cohort analysis indicated that the peak in age-specific prevalence moved from 30–39 years to 40 –49 years between the two NHANES data.

The limitation of NHANES data with regard to HCV is that some of the population groups with high HCV prevalence have been excluded. For example, in a study on homeless veterans, the prevalence of anti-HCV was as high as 41.7 %. Incarcerated persons also have a higher prevalence of HCV than the general population [22,23]. A recent study by Fox et al. reported that the prevalence of anti-HCV among incarcerated persons in California was 34.3 % [24]. In a systematic review of greater than 1000 sources, the prevalence of anti-HCV among intravenous drug users was 60–80 % and upwards of 80 % in 12 countries suggesting that 10 million intravenous drug users worldwide may be anti-HCV positive [19]. These data suggest that estimates of HCV prevalence based upon the NHANES data likely represent an underestimate of the true prevalence. A recent multicohort natural history model predicts that the prevalence of chronic hepatitis C-related cirrhosis and its complications will increase in the next two decades. It will mostly affect persons older than 60 years regardless of age at infection [25].

Natural history [26–30]

After acute infection with hepatitis B, a majority of infants (approximately 90 %) and a minority of adults become chronic carriers. Approximately 15–40 % of chronic carriers develop serious complications. The 5-year cumulative incidence of developing cirrhosis is 8–20 % without treatment; once cirrhosis is established the annual incidence of hepatocellular carcinoma is 2–8 %. A majority of patients do not clear acute hepatitis C (50–90 %). About 60–70 % of chronically infected

persons develop chronic liver disease, 5–20 % develop cirrhosis; the annual incidence of hepatocellular carcinoma is 1–5 %.

Mortality from HBV and HCV

Worldwide in 2008, there were 197,000 deaths due to viral hepatitis (B or C). In addition, there were 1.54 million deaths that were attributed to liver cirrhosis or liver cancer. However, the latter estimate does not specify the number of deaths resulting from viral hepatitis as compared to other causes of liver disease (http://www.who.int/evidence/bod). Hence, in totality it is assumed that 600,000 and 350,000 deaths are due to HBV and HCV, respectively [31]. In the USA, an estimated 3000 and 12,000 annual deaths are attributed to HBV and HCV, respectively [32].

Most mortality statistics in the USA are typically based on death certificate data [32]. Further, reported mortality estimates due to viral hepatitis (B or C) do not include deaths due to liver disease resulting from viral hepatitis deaths. Mortality from HBV-related liver disease has been estimated to have increased in the past two decades [33]. In 2007, there were 719 deaths (age-adjusted death rate 0.2 per 100,000) and 6571 (age-adjusted death rate 2.0 per 100,000) deaths attributed to hepatitis B and hepatitis C, respectively.

The age-adjusted death rate for HBV increased from 0.1 per 100,000 in 1978 to 0.3 in 2007. The death rate (per 100,000) was higher in men (0.4 for men, 0.1 for women) and in non-Whites (0.1 for Whites, 0.2 for Blacks and 0.9 for APIs). Rates were similar in non-Hispanic and Hispanic ethnicities (0.2). Although the increase in death rate over time was observed in all races and both genders through 1998 especially in men of other (non-White, non-Black) race, recent data suggests that the rates may have decreased. However, some of the decrease may reflect changes in coding practices (Figure 29.3) [34].

The age-adjusted death rate for HCV increased from 0.2 per 100,000 in 1978 (non-A non-B hepatitis) to 2.0 in 2007 (Figure 29.4) [34]. The death rate (per 100,000) was higher in men (2.7 for men, 1.3 for women) and in AI/AN (3.4), Blacks (2.6), as compared to Whites (2.0), and APIs (1.3). Rates were higher among Hispanic ethnicities (3.5 vs. 1.9). Projection studies have suggested that the burden of liver disease secondary to chronic HCV infection will continue to rise [35].

Conclusions

Hepatitis B and C viruses are both parenterally transmitted and, thus, share some common epidemiologic features. Both viruses may be transmitted by contaminated needles, unprotected sexual contacts, or perinatal exposure. Incidence of new infections with HBV and HCV has been largely decreasing in the USA and worldwide thanks to safer needle-using practices and universal precaution in health care as well as exclusion of blood donors with infection. In addition, vaccination programs in children have resulted in a profound decrease in acute HBV infection in adolescence and young adulthood. Despite these decreases in acute infections, the prevalence and burden of chronic HBV and HCV infection remain substantial. Chronic HBV is disproportionately high among Americans of Asian/Pacific Islander extractions. Chronic HCV infection is peculiarly prevalent among people born in the 1950s, especially among African and Mexican Americans and those who are homeless or incarcerated. In the USA as a whole, the burden of HBV and HCV (i.e. mortality) has been increasing in the recent past and focused epidemiologic attention is urgently necessary to screen, early diagnose and treat those with existing chronic infection as well as to continue prevention measures in children and adolescents.

Multiple choice questions

1 Which statement is true regarding the prevalence of hepatitis C?

A The prevalence of chronic hepatitis C-related cirrhosis will increase in the next two decades and primarily affect older individuals

B The prevalence of hepatitis C is highest in persons aged 20–29 years given intravenous drug use in this age group

C The prevalence of chronic HCV infection is highest among Asian Pacific Islander immigrants to the USA

D Worldwide, the prevalence of HCV is greater than that of HBV

E The National Health and Nutrition Examination Survey (NHANES) accurately represents the prevalence of chronic hepatitis C in the USA given that it is a nationally administered sample

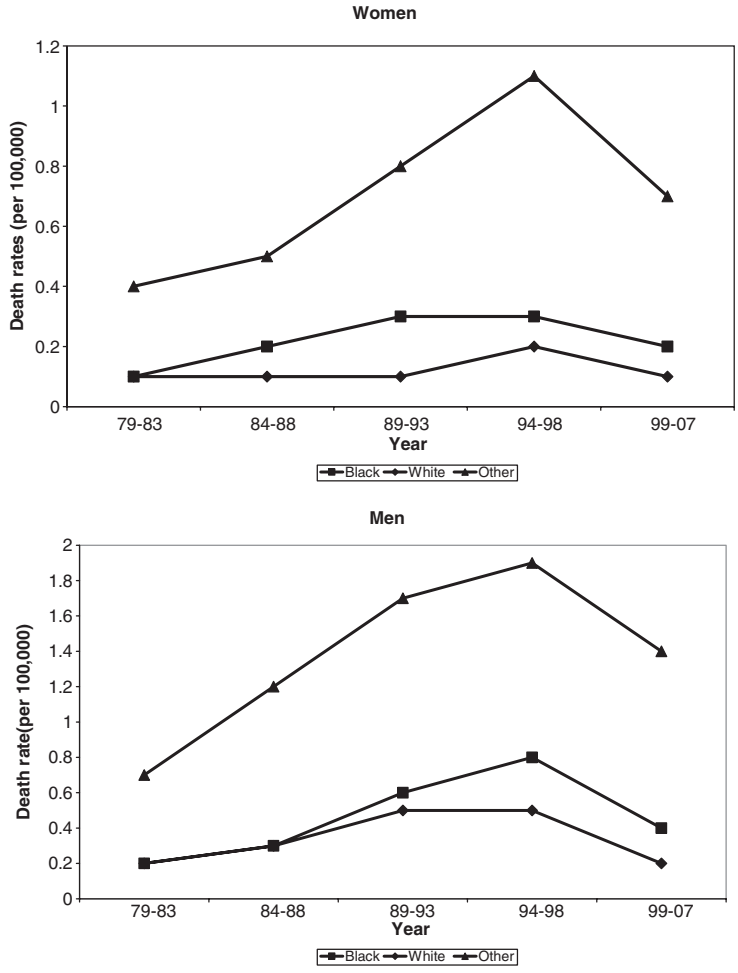


Figure 29.3 Age- and race-specific mortality from HBV-related disease in the United States. "Other" includes all decedents who did not belong in the white or black race category. Source: EASL-EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma 2012 [28].

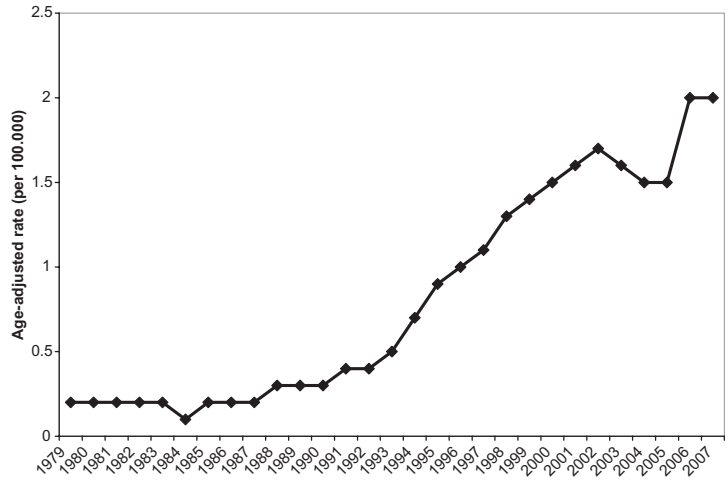


Figure 29.4 Mortality from HCV-related causes in the United States. Source: EASL-EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma 2012 [28].

2 The incidence of acute cases of hepatitis B has decreased in the USA for the following reasons except:

A Universal infant vaccination

B Universal screening of pregnant women

C Post-exposure prophylaxis of infants born to infected mothers

D Promotion of safer needle-using practices

E All of the above

3 Which statement is false regarding the mortality related to hepatitis B and C?

A Mortality related to hepatitis B has increased

B Mortality related to hepatitis C has increased

C The death rate is higher in men than women for both hepatitis B and C

D Rates are higher among Hispanic ethnicities with hepatitis B

E Mortality due to hepatitis B and hepatitis C is underestimated

4 Which statement is true regarding the natural history of hepatitis B and hepatitis C?

A Acute hepatitis B infection is often cleared in adults as compared to infants

B A majority of chronic carriers of hepatitis B develop serious complications of end-stage liver disease

C Among hepatitis B cirrhotics, the annual incidence of hepatocellular carcinoma is less than 1 %

D Acute hepatitis C develops into chronic hepatitis C in a minority of patients

E All of the above

5 Which statement is true regarding transmission of hepatitis B and hepatitis C?

A Among adults, high-risk sexual activity is a frequent route of transmission of HBV and HCV

B Injection drug use is the most common route of transmission for hepatitis C

C Transmission of HBV among chronic hemodialysis patients may occur when appropriate isolation guidelines are not followed

D Like HBV, HCV may be transmitted in the perinatal period from infected mother to the newborn

E All of the above

References

1 Alter MJ. Epidemiology and prevention of hepatitis B. *Semin Liver Dis* 2003;23:39–46.

2 Hoofnagle JH. Course and outcome of hepatitis C. *Hepatology* 2002;36:S21–9.

3 Shepard CW, Finelli L, Fiore AE, Bell BP. Epidemiology of hepatitis B and hepatitis B virus infection in United States children. *Pediatr Infect Dis J* 2005;24:755–60.

4 Xu ZY, Liu CB, Francis DP, et al. Prevention of perinatal acquisition of hepatitis B virus carriage using vaccine: preliminary report of a randomized, double-blind placebo-controlled and comparative trial. *Pediatrics* 1985;76:713–18.

5 Beasley RP, Hwang LY. Postnatal infectivity of hepatitis B surface antigen-carrier mothers. *J Infect Dis* 1983;147:185-90.

6 Centers for Disease Control and Prevention (CDC) (2004) NHANES 2001–2002 Public Data General Release File Documentation. CDC, Atlanta, GA.

7 Harpaz R, Von Seidlein L, Averhoff FM, et al. Transmission of hepatitis B virus to multiple patients from a surgeon without evidence of inadequate infection control. *New Engl J Med* 1996;334:549–54.

8 Armstrong GL, Wasley A, Simard EP, et al. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med* 2006;144:705–14.

9 Centers for Disease Control and Prevention (CDC) (1998) Recommendations for prevention and control of hepatitis C virus infection and HCV-related chronic disease. MMWR. CDC, Atlanta, GA, pp. 1–9.

10 Zanetti AR, Van Damme P, Shouval D. The global impact of vaccination against hepatitis B: a historical overview. *Vaccine* 2008;26:6266–73.

11 Centers for Disease Control and Prevention (CDC) (2009) Viral Hepatitis Surveillance – United States. CDC, Atlanta, GA.

12 McQuillan G, Kruszon-Moran D, Denniston M, Hirsch R. (2010) Viral Hepatitis. NCHS Data Brief. Centers for Disease Control and Prevention (CDC), Atlanta, GA, pp. 1–8.

13 Kim WR. The burden of hepatitis C in the United States. *Hepatology* 2002;36:S30–4.

14 Daniels D, Grytdal S, Wasley A. Surveillance for acute viral hepatitis – United States, 2007. *MMWR Surveill Summ* 2009;58:1–27.

15 World Health Organization (2006) Hepatitis B. WHO, Geneva.

16 Anonymous. Hepatitis C – global prevalence (update). *Wkly Epidemiol Rec* 2000;75:18–19.

17 Wasley A, Kruszon-Moran D, Kuhnert W, et al. The prevalence of hepatitis B virus infection in the United States in the era of vaccination. *J Infect Dis* 2010;202:192–201.

18 Coleman PJ, McQuillan GM, Moyer LA, et al. Incidence of hepatitis B virus infection in the United States, 1976–1994: estimates from the National Health and

Nutrition Examination Surveys. *J Infect Dis* 1998;178: 954–9.
19 Nelson PK, Mathers BM, Cowie B, et al. Global epidemiology of hepatitis B and hepatitis C in people who inject drugs: results of systematic reviews. *Lancet* 2011;378:571–83.
20 Centers for Disease Control and Prevention (CDC) (2006) Hepatitis Surveillance Report. CDC, Atlanta, GA, pp. 1–53.
21 Alter MJ, Kruszon-Moran D, Nainan OV, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *New Engl J Med* 1999;341:556–62.
22 Cheung RC, Hanson AK, Maganti K, et al. Viral hepatitis and other infectious diseases in a homeless population. *J Clin Gastroenterol* 2002;34:476–80.
23 Briggs ME, Baker C, Hall R, et al. Prevalence and risk factors for hepatitis C virus infection at an urban Veterans Administration medical center. *Hepatology* 2001;34:1200–5.
24 Fox RK, Currie SL, Evans J, et al. Hepatitis C virus infection among prisoners in the California state correctional system. *Clin Infect Dis* 2005;41:177–86.
25 Davis GL, Alter MJ, El-Serag H, et al. Aging of hepatitis C virus (HCV)-infected persons in the United States: a multiple cohort model of HCV prevalence and disease progression. *Gastroenterology* 2010;138:513–21, e1–6.
26 EASL clinical practice guidelines on the management of ascites, spontaneous bacterial peritonitis, and hepatorenal syndrome in cirrhosis. *J Hepatol* 2010;53:397–417.
27 2011 European Association for the Study of the Liver hepatitis C virus clinical practice guidelines. *Liver Int* 2012;32(Suppl 1):2–8.
28 EASL-EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* 2012;56:908–43.
29 Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology* 2011;53:1020–2.
30 Lok AS, McMahon BJ. Chronic hepatitis B: update 2009. *Hepatology* 2009;50:661–2.
31 Hatzakis A, Wait S, Bruix J, et al. The state of hepatitis B and C in Europe: report from the hepatitis B and C summit conference*. *J Viral Hepat* 2011;18(Suppl 1): 1–16.
32 Kim WR, Brown RS, Jr., Terrault NA, El-Serag H. Burden of liver disease in the United States: summary of a workshop. *Hepatology* 2002;36:227–42.
33 Kim WR, Ishitani MB, Dickson ER. Rising burden of hepatitis B in the United States: Should the 'other' virus be forgotten? *Hepatology* 2002;36:222A.
34 Centers for Disease Control and Prevention (CDC) WONDER [online database]. Underlying Cause of Death. Vol. 2011. CDC, Atlanta, GA.
35 Armstrong GL, Alter MJ, McQuillan GM, Margolis HS. The past incidence of hepatitis C virus infection: implications for the future burden of chronic liver disease in the United States. *Hepatology* 2000;31:777–82.

Answers to multiple choice questions

1. A
HCV prevalence has increased linearly with age. A recent multicohort natural history model predicts that the prevalence of chronic hepatitis C-related cirrhosis and its complications will increase in the next two decades. It will mostly affect persons older than 60 years regardless of age at infection.
2. E
All of the above factors have helped decrease the incident cases of hepatitis B.
3. D
Death rates (per 100,000) are higher among Hispanic ethnicities (3.5 vs. 1.9) for hepatitis C but similar in non-Hispanic and Hispanic ethnicities (0.2 vs. 0.2) with hepatitis B.
4. A
Persons who acquire hepatitis B viral infection during infancy due to vertical transmission are less likely to clear the virus and are at a higher risk of developing chronic hepatitis B.
5. E
All of the above

30 Epidemiology of alcoholic liver disease

Sumeet K. Asrani[1] *& William Sanchez*[2]
[1]Hepatology, Baylor University Medical Center, Dallas, TX, USA
[2]Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA

Key points

- Alcoholic liver disease (ALD) is a major cause of liver disease worldwide.
- The incidence and prevalence of ALD is likely underestimated.
- Risk factors that may affect development of significant ALD include pattern of alcohol consumption, gender, ethnicity, genetic factors, and coexisting liver disease.
- Worldwide in 2004, 3.8 % of all deaths were attributable to alcohol. Of these, 16.5 % were due to cirrhosis of the liver. In the United States, chronic liver disease and cirrhosis was the twelfth leading cause of death with a death rate of 9.1/100,000 in 2007. Of these, ALD was responsible for 4.5/100,000 deaths.
- Globally, alcohol was responsible for a loss of 4.6 % of all disability adjusted life-years. Of these, cirrhosis of the liver was responsible for 9.8 % of all alcohol attributable disability adjusted life-years.

Disease definition

According to the World Health Organization (WHO), about 2 billion people consume alcohol worldwide and upwards of 75 million are diagnosed with alcohol use disorders [1]. Alcohol-related disease results in approximately 2.5 million deaths each year [2]. Almost 4 % of all deaths worldwide are attributed to alcohol [2]. Liver disease related to significant alcohol consumption, alcoholic liver disease (ALD), is a major cause of liver disease worldwide [3]. Furthermore, alcohol use can compound the extent of liver injury when in coexistence with other factors (e.g. viral hepatitis).

Clinical manifestations

ALD comprises a spectrum of disorders ranging from asymptomatic liver test derangements, to severe acute hepatitis and end-stage chronic liver disease [4–6]. The diagnosis of ALD requires a strong index of suspicion; other competing or concomitant causes of liver disease (e.g. viral hepatitis) should be considered. Underreporting of alcohol use is common and recognition of alcohol dependence or abuse is often minimized by patients and thus, not adequately addressed by physicians [7]. The CAGE and AUDIT questionnaires (Table 30.1) may be helpful tools to screen for maladaptive alcohol use [8–10]. Mathematical models (Table 30.2) such as the Model for end-stage liver disease (MELD) score, Glasgow alcoholic hepatitis score, Lille model, ABIC score, and Maddrey discriminant factor may help identify patients at highest risk of mortality or complications from ALD [11–14].

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 30.1 Screening tools for alcohol dependence or abuse [8–10]

CAGE

1. Have you ever felt you needed to **cut** down on your drinking?
2. Have people **annoyed** you by criticizing your drinking?
3. Have you ever felt **guilty** about your drinking?
4. Have you ever felt you needed a drink first thing in the morning (**eye-opener**) to steady your nerves or to get rid of a hangover?

An answer or yes to 2 or more questions is clinically significant.

AUDIT (The Alcohol Use Disorders Identification Test)

	0	1	2	3	4
1. How often do you have a drink containing alcohol?	Never	Monthly or less	2–4 times a month	2–3 times a week	4 or more times a week
2. How many drinks containing alcohol do you have on a typical day when you are drinking?	1 or 2	3 or 4	5 or 6	7 to 9	10 or more
3. How often do you have six or more drinks on one occasion?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
4. How often during the last year have you found that you were not able to stop drinking once you had started?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
5. How often during the last year have you failed to do what was normally expected of you because of drinking?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
7. How often during the last year have you had a feeling of guilt or remorse after drinking?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
8. How often during the last year have you been unable to remember what happened the night before because of your drinking?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
9. Have you or someone else been injured because of your drinking?	No		Yes, but not in the last year		Yes, during the last year
10. Has a relative, friend, doctor, or other healthcare worker been concerned about your drinking or suggested you cut down?	No		Yes, but not in the last year		Yes, during the last year

A total ≥8 for men up to age 60, or ≥4 for women, adolescents, or men over age 60 is considered a positive screening test.

Incidence and prevalence

The incidence and prevalence of ALD is not well established in population-based studies, especially in the United States. There are striking differences depending upon geographic region, race, gender, ethnicity, and socioeconomic strata. Studies are further limited by referral bias (e.g. tertiary care center), composition of population under study (inpatient vs. outpatient) and disease definition (e.g. alcoholic hepatitis vs. cirrhosis). In some studies, a large proportion of patients have cirrhosis on presentation, possibly biasing the estimate towards patients with advanced fibrosis. Incomplete ascertainment of cases may also play a role. The prevalence of ALD may be higher, as noted by autopsy studies, given that underreporting of alcohol-related disease is common [15]. Further lack of standardized definitions (e.g. for alcoholic

Table 30.2 Scoring systems for prognosis in patients with alcoholic hepatitis

1. **M**addrey **D**iscriminant **F**unction (score ≥32 poor prognosis) [69]
MDF = 4.6 (Patient's PT – control PT) + total bilirubin (mg dL^{-1})
2. **M**odel for **E**ndstage **L**iver **D**isease score (score ≥18 poor prognosis) [12]
MELD score = 3.8 * $\log_e$ (bilirubin mg dL^{-1}) + 11.2 * $\log_e$ (INR) + 9.6 * $\log_e$ (creatinine mg dL^{-1}) + 6.4
3. Glasgow alcoholic hepatitis score (score >8 on day 1 or day 7 poor prognosis) [13]

	1	2	3
Age	<50	≥50	–
WCC	<15	≥15	–
Urea (mmol L^{-1})	<5	≥5	–
PT ratio	<1.5	1.5–2.0	≥2
Bilirubin (mg dL^{-1})	<7.3	7.3–14.6	≥14.6

4. **A**ge, serum **B**ilirubin, **I**NR, and serum **C**reatinine (ABIC) score (score >9 poor prognosis) [11]
ABIC score: (age × 0.1) + (serum bilirubin × 0.08) + (serum creatinine × 0.3) + (INR × 0.8).

hepatitis (AH)) also contribute. For example, in one population-based study, a substantial proportion of patients with AH were misdiagnosed in the community (60 %) suggesting that the true burden of AH is frequently underestimated [16]. Bearing these limitations in mind, Table 30.3 summarizes available data on the incidence and prevalence of ALD.

Risk factors

Several risk factors may affect development of significant ALD including pattern of alcohol consumption, gender, ethnicity, age, genetic factors, and coexisting causes of liver disease (e.g. viral hepatitis, obesity, metabolic liver disease, etc.) [3,6].

The quantity, frequency, and pattern of alcohol consumption are all significant determinants of ALD [4–6,17]. It is unclear whether cirrhosis is related to a cumulative dose of alcohol over the lifetime or occurs after reaching a threshold dose of alcohol consumption [4–6,17,18]. The relationship between the quantity of alcohol ingestion and development of ALD may not be linear [19,20].

Frequency

The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) analyzed cross-sectional drinking patterns in the US population between 2001 and 2002; 65 % were current drinkers (≥1 drink in the past year). Among current drinkers, 62 % were light drinkers (<3 drinks per week), 21 % were moderate drinkers, and 16 % were heavy drinkers (men >2 daily drinks and women >1 daily drink) [21]. Alcohol consumption varied by gender and ethnicity with higher rates observed in men and American Indians/Alaska Natives (AI/AN) [21]. The limitations of the survey, however, include nonresponder bias and underreporting of personal alcohol consumption. Similar results were reported by the Behavioral Risk Factor Surveillance System (BRFSS), an annual cross-sectional telephone survey [22].

Abuse and dependence

Globally in 2004, the 1-year prevalence of alcohol use disorders (e.g. alcohol dependence and abuse) among persons aged 15–64 was 3.6 % (worldwide range 0.3–10.9 %) with a higher rate among men (6.3 % vs. 0.9 %) [2,3]. The highest rates were observed in the eastern European region (10.9 %), lowest in the eastern Mediterranean region (0.3 %), and intermediate in the Americas (5.2 %).

In the United States in 2001–2002, the cross-sectional prevalence of alcohol dependence and abuse was 3.8 % and 4.7 %, cumulatively affecting 17.6 million Americans. Once again, rates were higher among men and higher in AI/AN [21,23]. However, other longitudinal studies have shown a lifetime prevalence of alcohol abuse as high as 30 % [24].

Table 30.3 Incidence and prevalence of alcoholic liver disease

Study	Population characteristics	Primary result	Comments
INCIDENCE			
Becker et al. ***Copenhagen Heart Study*** [19]	Danish national hospital discharge register (1976–1978)	Incidence of alcoholic cirrhosis 0.2 % per year in men and 0.03 % per year in women	Self-reported alcohol use Survey response = 72 % 12-year follow-up
Asrani et al. [16]	Olmsted County, MN, USA (1992–2007)	Incidence of alcoholic hepatitis increased from 10.3/100,000 to 17.1/100,000	Predominant white population, vulnerable populations not included
Bell et al. [70]	All gastroenterology practice in three representative counties in Connecticut, Oregon and northern California (1999–2001)	Incidence of chronic liver disease (including ALD): 63.9/100,000 (range 53.7–74.3); 45–54 years (127.8/100,000); men as compared to women (77.7/100,000 vs. 50.7/100,000)	Referral bias, exclusion of inpatient practices, and low alcohol consumption rates. A large proportion had cirrhosis on presentation.
PREVALENCE			
Bellentani et al. ***Italian Dionysos Study*** [30]	Entire population of two small Italian towns	Prevalence for cirrhosis = 289/100,000, chronic hepatitis or fatty liver = 3368/100,000 and excessive drinking = 10,221/100,000	13.5 % of heavy drinkers developed cirrhosis
Nationwide Inpatients Sample (NIS) [71]	NIS, the largest all payer inpatient database, USA (1988–2004)	Alcoholic cirrhosis (per 100,000 persons) increased by 52 % from 11.9 (95 % CI 11.0–12.8) to 18.1 (95 % CI 16.8–19.4). The rates per 100,000 persons of alcoholic hepatitis diagnosis changed minimally, from 4.9/100,000 (95 % CI 4.4–5.3) to 4.2/100,000 (95 % CI 3.9–4.5). Prevalence of alcoholic cirrhosis (per 100,000) was 16.9 among Hispanics versus 11.1 among Whites.	Incomplete case ascertainment, decompensated liver disease may be overrepresented, outpatient practices were excluded and the unique number of patients were not analyzed.
National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) [21]	Longitudinal survey, USA (2001–2002)	Prevalence of cirrhosis of the liver was 2.15/1000. Higher rates were seen among persons aged 45–64 (3.68/1000) and among men as compared to women (3.17/1000 vs. 1.22/1000). Overall, the rate was highest in former drinkers (4.6/1000) followed by current drinkers (1.69/1000) and lifetime abstainers (1.46/1000). It was higher for those starting drinking at age less than 14 (8.53/1000) as compared to those starting after age 21 years (1.35/1000). Among those with an alcohol use disorder the rate was 5.39/1000 and among those consuming hard liquor the rate was 3.88/1000. Among current drinkers, the rate was higher in heavy drinkers, 4.64/1000 as compared to moderate drinkers, 1.05/1000, and light drinkers, 1.1/1000.	Survey response = 81 % Self-reported prevalence of chronic liver disease, rather than ALD, recall bias
Fischer et al. [72]	Two medical centers predominantly serving AI/AN, Alaska (2003–2004)	1496 (4.9 %) had chronic liver disease. Overall, 41.5 % (621/1496) had alcohol-related liver disease alone and 9.1 % (136/1496) had chronic hepatitis C (HCV) and alcohol-related liver disease.	Referral bias

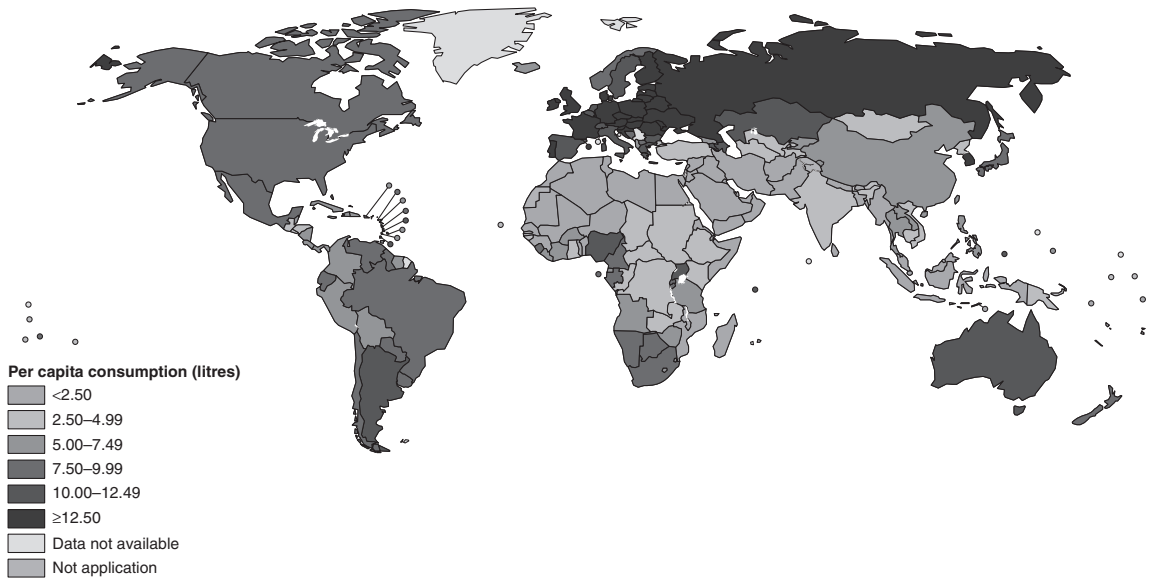


Figure 30.1 Total adult per capita consumption of pure alcohol (liters) in 2005 [2].
Source: Reproduced from *Global Status Report on Alcohol and Health*, World Health Organization, 2011.

Alcohol type/quantity

The association between alcohol consumption of beer and spirits and cirrhosis- related mortality trends has been proposed [25,26]. In a survey of greater than 30,000 persons in Denmark, drinking beer or spirits was more likely to be associated with liver disease than drinking wine [27].

Globally, the amount of alcohol consumption had remained steady since the 1980s but has increased in recent years. Worldwide in 2005, the mean adult consumption per capita was 6.1 liters (1.6 gallons) of pure alcohol [2]. Overall, 45 % of total recorded alcohol is consumed in the form of spirits, 36 % in the form of beer, and 8.6 % in the form of wine [2]. Figure 30.1 shows the global consumption patterns in 2005. In the United States (Figure 30.2), though alcohol consumption decreased from a high of 2.8 gallons of ethanol in 1980 to 2.1 gallons in 1997, it has recently increased

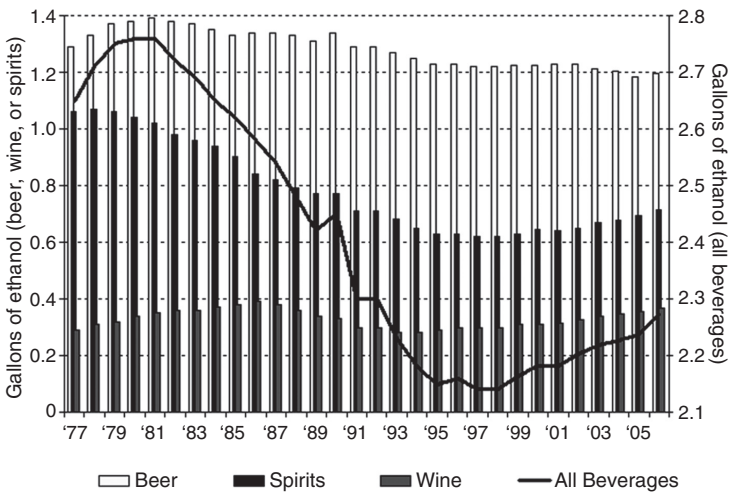


Figure 30.2 Apparent per capita alcohol consumption in the United States, 1977–2006 [28].
Source: Lakins et al. 2008 [28].

[28]. The per capita alcohol consumption of ethanol in 2007 was 2.3 gallons (8.7 liters). The majority of consumption was in the form of beer (1.2 gallons or 4.6 liters), followed by spirits (0.7 gallon or 2.8 liters), and wine (0.4 gallon or 1.4 liters) [28]. In the Italian Dionysos Study, the relative odds ratio (OR) for chronic liver disease increases with ethanol consumption with an OR 0.8 (0.7–0.8) for consumption of less than 30 grams of ethanol per day, OR 2.4 (95 % CI 2.3–2.4) for consumption of 31–60 g of ethanol per day, and OR 5.1 (4.8–5.4) for consumption of greater than 120 g per day. The risk of chronic liver disease was related to a total lifetime alcohol intake of more than 100 kg, or a daily intake of greater than 30 g [29,30]. However, only 13.5 % of persons with significant alcohol consumption developed cirrhosis.

Pattern

The pattern of alcohol consumption may also play a role, especially binge drinking [31]. The severity of alcohol use disorders is the same for persons with binge drinking (5 drinks in one setting over 2 hours) once a month as it is for those that consume 1 drink per day [32]. According to NESARC and BRFSS data, the frequency of binge drinking has increased [21,22,33]. Unfortunately, the rates of progression to ALD by pattern of alcohol consumption are not well defined.

Age

Besides pattern of drinking, age at initiation of drinking is associated with future maladaptive alcohol-related behaviors [31,34]. According to the National Institute on Alcohol Abuse and Alcoholism (NIAAA) [35], the mean age at initiation of drinking in 2005 was 14.2 years, up from 13.8 years in 1991. The prevalence of drinking in the past 30 days among 12–20 year olds was 28.3 % in 2005 (33.4 % in 1991). Males had a higher average frequency, quantity, and volume of consumption in the past 30 days than females. Alcohol use was highest in non-Hispanic whites (32.6 %) followed by Hispanics (25.0 %), AI/AN (22.7 %), African Americans (AA) (19.1 %), and Asians (14.7 %). Amongst youth, 21.1 to 34.2 % reported starting drinking at age less than 12 years. Between 2005 and 2007, youth drank 5.8 days in the past 30 days with 4.8 drinks on drinking days with an average of 34.2 monthly drinks [35]. The prevalence of binge drinking was 18.6–25.5 % [35].

Hepatitis C

The interaction between alcohol use and hepatitis C has been well described [36]. In a population-based cohort study, the overall incidence rate of end-stage liver disease was 3.1 per 1,000 person-years. In this select population, the incidence rate was 1.7, 2.7, and 6.4/1,000 person-years in chronic hepatitis C (HCV) persons consuming <90 g (1 drink =12–13.6 g), 90–260 g, and >260 g of ethanol per week [36]. Similar results were seen in the Italian Dionysos Study [37].

Obesity

Emerging research highlights the interaction between alcohol use and obesity [38–40]. For example, the combination of obesity and an alcohol consumption of more than 150 g of alcohol per week was associated with a fivefold increased risk of cirrhosis as compared to obese women who drank less than 70 g of alcohol a week [40]. In the United Kingdom, the incidence of liver cirrhosis among nondrinking women was 0.8/1000 in nonobese (BMI <25) and 1/1000 in obese (BMI >30) women. Among persons consuming more than 150 g, the incidence of liver cirrhosis was 2.7/1000 in the nonobese and 5/1000 in obese women. The interaction between obesity and alcohol use in the risk of developing significant liver disease may be synergistic rather than additive. Similar findings were observed in a population-based study of persons with alcoholic hepatitis [38].

Genetic factors

Recently, genome-wide association studies have shown that genetic polymorphism (rs738409 C>G) in the *PNPLA3*/adiponutrin gene may be associated with an increased risk of ALD and cirrhosis in Mestizo Mexicans and European Caucasians with excessive alcohol intake [41–43]. In multivariable analysis of samples from a single center study, rs738409 was the strongest independent factor associated with risk of cirrhosis (OR 2.08; 95 % CI 1.15–3.77). However, similar association with the allele in question has also been observed with nonalcoholic fatty liver disease-related cirrhosis [44]. Other targets are being explored [45,46]. Hence, the genetic epidemiology of ALD remains to be fully elucidated.

Natural history and mortality

Worldwide in 2004, 3.8 % (n = 2,255,000) of all deaths were attributable to alcohol. Of these, 16.5 % (n = 373,000) were due to cirrhosis of the liver [3]. Though more deaths due to cirrhosis occurred in men (297,000 vs. 76,000), the percentage of alcohol-attributable deaths due to cirrhosis was higher in women (17.1 % vs. 14.6 %) [3]. Globally between 2000 and 2002, alcohol-related liver cirrhosis mortality was a significant cause of death ranging from 2.8/100,000 men in Singapore to 68.3/100,000 men in Hungary. In the European Union, the mortality was 33.8/100,000 in men and 5.9/100,000 in women between 2000–2002 [47].

According to the National Center for Health Statistics (NCHS), in 2007 the age-adjusted death rate from any cause was 760.2 per 100,000 (2,423,712 deaths) in the United States. Chronic liver disease and cirrhosis was the twelfth leading cause of death with a death rate of 9.1/100,000 (29,165 deaths). Of these, ALD was responsible for 4.5/100,000 deaths (14,406 deaths) [48]. In 2007, the ALD-related death rate (per 100,000) was higher in men (6.8 vs. 2.3) and highest among AI/AN (17.1) followed by Whites (4.7), Blacks (3.4), and Asians (1.1). Mortality (per 100,000) was greatest among persons aged 55–64 (13.2) followed by ages 45–54 (11.7) and ages 65–74 (10.1). In 2007, alcohol-related cirrhosis death rates (per 100,000) were 12.6 in Hispanic white males, 6.5 in non-Hispanic white males, and 5.6 in non-Hispanic black males. Among females, alcohol-related cirrhosis deaths were highest in non-Hispanic Whites (2.5) followed by Hispanic Whites (2.3), and non-Hispanic Blacks (1.8) [48].

Globally, variable trends in cirrhosis-related mortality have been observed. For example, total recorded alcohol consumption in Britain doubled between 1960 and 2002 and correlated with 104 % increase in cirrhosis mortality in men in Scotland and 69 % increase in England and Wales. This was in contrast to an apparent decline in cirrhosis mortality in other parts of Europe [49]. However, actual cirrhosis related to alcohol was not reported.

In the United States from 1970 to 2007, the age-adjusted death rate from liver cirrhosis (all cause) declined by 48.9 % from 17.8/100,000 to 9.1/100,000 [48]. The age-adjusted death rate from ALD-related cirrhosis decreased by 28.6 % (6.3 deaths/100,000 population in 1970 to 4.5 deaths/100,000 in 2007). From 1970 to 2005, rates of cirrhosis (all cause) and alcohol-related liver cirrhosis were higher for males as compared to females. Disparities in deaths in AA and Whites attributed to ALD decreased during this time [50]. Rates of alcohol-related liver cirrhosis for white males, white females, AA males, and AA females declined by 16.4, 22.5, 67.6, and 77.2 percent from 1970 to 2007, respectively [48].

The amount of alcohol consumption matters with higher rates of mortality observed in those with increased alcohol consumption [51,52]. Higher rates are seen in persons with alcoholic hepatitis and cirrhosis as compared to cirrhotics or alcoholics alone [53]. Further, higher mortality rates are observed in certain minority groups (e.g. AI/AN) [54] and those co-infected with HCV and concomitant alcohol use [50, 55], high-risk populations (e.g. prisoners) [56], and by certain socioeconomic factors (e.g. urban residents, unemployed, and nonprofessional employees) [57,58].

Disability, quality of life, and healthcare seeking

Hospitalizations

In 2006, the rate of alcohol-related discharges per 10,000 population was 72.4/10,000 (1.7 million discharges). Discharges due to chronic liver disease and cirrhosis increased from 15.2/10,000 in 1989 to 16.1 in 1999 to 21.3 in 2006. Specifically, discharges due to alcoholic cirrhosis increased from 7.4/10,000 in 1989 to 8.5 in 1999 to 10/10,000 in 2006 [59].

Transplantation

Between 1998 and 2007, 15–16 % of the candidates on the waiting list were listed due to ALD and 6–8 % of candidates were listed for both hepatitis C and alcohol [60]. Between 1998 and 2007, the percentage of deceased donor transplant recipients with ALD ranged from 11–13 % and for both alcohol and hepatitis C ranged from 5–8 %. In 2007, 5 % of living donor transplant recipients were transplanted for ALD and 2 % were transplanted for both hepatitis C and alcohol.

Costs

The costs attributed to alcohol abuse including healthcare expenditure, alcohol services, medical

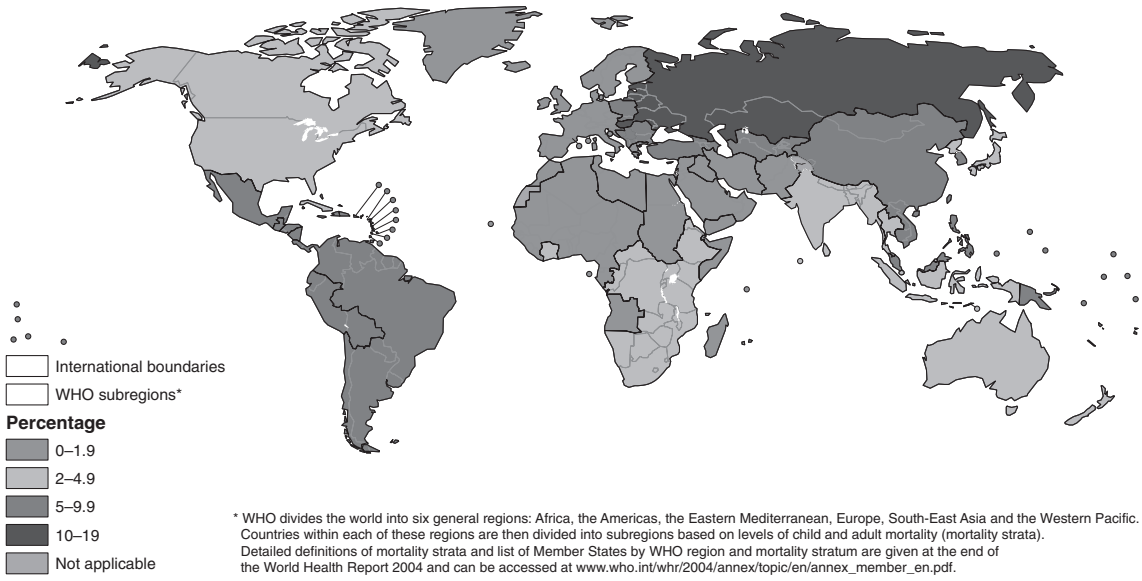


Figure 30.3 Alcohol attributable disability associated life years, 2004 [2].
Source: Reproduced from *Global Status Report on Alcohol and Health*, World Health Organization, 2011.

consequences, loss of productivity due to premature death or morbidity, and other impacts on society increased from $148.0 billion in 1992 to $184.6 billion in 1998 in the United States. Adjusted to 2007 US dollars, the global cost was $234.9 billion with a cost of $837 per head. All countries spent more than 1 % of their GDP, with the highest in the United States (2.7 %) and in South Korea (3.3 %) [2]. Healthcare cost was 12.7 % of total cost [3].

Years of potential life lost

The actual burden of disease was even higher. Disability-adjusted life years is a composite measure of years lost due to premature deaths and years spent in unhealthy states [61]. Globally, alcohol was responsible for a loss of 4.6 % (n = 70,910,000) of all disability-adjusted life years (Figure 30.3). Of these, cirrhosis of the liver was responsible for 9.8 % of all alcohol-attributable disability-adjusted life years [3].

Areas for further study

Despite its global pervading influence, ALD is understudied [50,62]. There are limited population-based cohort studies that determine the incidence, prevalence, and natural history of ALD, especially in the United States. The lack of clinical research attention may reflect the fact that ALD affects less affluent and less influential populations of our society [62]. Further, the stigma associated with alcohol abuse and preconceived notions regarding individuals who have ensuing liver disease may preclude objective research. Many studies that report the "natural history" of ALD tend to be single-center studies that follow a highly selected group of individuals seeking attention for their disease, mostly at a late decompensated stage. The characteristics of this group of patients may be systematically different from the representative patient with ALD in the general population [63].

The impact of interventions aimed at decreasing alcohol use among young adults needs to be assessed. Finally, the contribution of genetic predictors of propensity to develop significant ALD needs to be further described.

Conclusion

The burden of liver disease attributed to alcohol is significant and likely underestimated. For example,

most of the survey data on the prevalence of chronic liver disease attributed to alcohol is self-reported and subject to nonresponder bias and recall bias. There is heterogeneity in reporting on alcohol-related deaths and inconsistent definitions of alcohol consumption [64]. Finally, autopsy data would probably provide a more reliable indicator of deaths due to cirrhosis [65]. Recently, it has been shown that if a more comprehensive definition of liver-related mortality is used, in contrast to the declining trend in deaths, there has been in fact little change in overall liver mortality since 1970 [66–68]. Hence, the pervasive effect of alcohol as a significant cause of liver-related morbidity and mortality requires continued attention. Despite the interest paid to other important causes of liver disease (viral hepatitis or nonalcoholic fatty liver disease), the interest in ALD needs to be revived.

Multiple choice questions

1 Which of the following is *NOT* a tool used to predict the severity of alcoholic liver disease?
- **A** MELD score
- **B** Glasgow score
- **C** AUDIT questionnaire
- **D** Lille model
- **E** ABIC score

2 Which region of the world has the highest prevalence of alcohol use disorders?
- **A** North America
- **B** South America
- **C** Western Europe
- **D** Eastern Europe
- **E** Southeastern Asia

3 Which of the following has been associated with an increased risk in alcoholic liver disease?
- **A** Average daily alcohol consumption of >30 g
- **B** Younger age at the onset of alcohol use
- **C** Obesity (BMI $\geq$30 kg m^{-2})
- **D** Hepatitis C infection
- **E** Polymorphism in the *PNPLA3*/adiponutrin gene

4 Which of the following groups has the highest death rate due to alcoholic liver disease in the United States?
- **A** Women
- **B** Non-Hispanic Whites
- **C** Persons aged over 65 years
- **D** American Indian/Alaskan Native
- **E** Asians

References

1 World Health Organization (WHO) (2004) *Global Status Report on Alcohol 2004*, WHO, Geneva.
2 WHO (2011) *Global Status Report on Alcohol and Health*, WHO, Geneva.
3 Rehm J, Mathers C, Popova S, et al. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet* 2009;373(9682):2223–33.
4 O'Shea RS, Dasarathy S, McCullough AJ. Alcoholic liver disease. *Am J Gastroenterol* 2010;105(1):14–32; quiz 3.
5 McCullough AJ, O'Shea RS, Dasarathy S. Diagnosis and management of alcoholic liver disease. *J Dig Dis* 2011;12(4):257–62.
6 O'Shea RS, Dasarathy S, McCullough AJ. Alcoholic liver disease. *Hepatology* 2010;51(1):307–28.
7 McQuade WH, Levy SM, Yanek LR, et al. Detecting symptoms of alcohol abuse in primary care settings. *Arch Fam Med* 2000;9(9):814–21.
8 O'Brien CP. The CAGE questionnaire for detection of alcoholism: a remarkably useful but simple tool. *JAMA* 2008;300(17):2054–6.
9 Ewing JA. Detecting alcoholism. The CAGE questionnaire. *JAMA* 1984;252(14):1905–7.
10 Saunders JB, Aasland OG, Babor TF, et al. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption – II. *Addiction* 1993;88(6):791–804.
11 Dominguez M, Rincon D, Abraldes JG, et al. A new scoring system for prognostic stratification of patients with alcoholic hepatitis. *Am J Gastroenterol* 2008;103(11):2747–56.
12 Dunn W, Jamil LH, Brown LS, et al. MELD accurately predicts mortality in patients with alcoholic hepatitis. *Hepatology* 2005;41(2):353–8.
13 Forrest EH, Evans CD, Stewart S, et al. Analysis of factors predictive of mortality in alcoholic hepatitis and derivation and validation of the Glasgow alcoholic hepatitis score. *Gut* 2005;54(8):1174–9.
14 Louvet A, Naveau S, Abdelnour M, et al. The Lille model: a new tool for therapeutic strategy in patients with severe alcoholic hepatitis treated with steroids. *Hepatology* 2007;45(6):1348–54.
15 Haberman PW, Weinbaum DF. Liver cirrhosis with and without mention of alcohol as cause of death. *Br J Addict* 1990;85(2):217–22.
16 Asrani SK, Benson JT, Kim WR. The epidemiology of alcoholic hepatitis: a population-based study in the US. *Gastroenterology* 2010;138(5):S-805.

17 Zakhari S, Li TK. Determinants of alcohol use and abuse: Impact of quantity and frequency patterns on liver disease. *Hepatology* 2007;46(6):2032–9.
18 Tsukamoto H. Conceptual importance of identifying alcoholic liver disease as a lifestyle disease. *J Gastroenterol* 2007;42(8):603–9.
19 Becker U, Deis A, Sorensen TI, et al. Prediction of risk of liver disease by alcohol intake, sex, and age: a prospective population study. *Hepatology* 1996;23(5):1025–9.
20 Kamper-Jørgensen M, Grønbaek M, Tolstrup J, Becker U. Alcohol and cirrhosis: dose–response or threshold effect? *J Hepatol* 2004;41(1):25–30.
21 Chen CM, Yi H, Falk DE, et al. (2006) Alcohol use and alcohol use disorders in the United States: Main findings from the 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). National Institutes of Health, Bethesda, MD, pp. 1–247.
22 Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System Survey Data, 1995–2008. Available at: http://www.cdc.gov/brfss/ (accessed May 2013).
23 Grant BF, Dawson DA, Stinson FS, et al. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. *Drug Alcohol Depend* 2004;74(3):223–34.
24 Vaillant GE. A 60-year follow-up of alcoholic men. *Addiction* 2003;98(8):1043–51.
25 Ramstedt M. Alcohol consumption and liver cirrhosis mortality with and without mention of alcohol – the case of Canada. *Addiction* 2003;98(9):1267–76.
26 Roizen R, Kerr WC, Fillmore KM. Cirrhosis mortality and per capita consumption of distilled spirits, United States, 1949–94: trend analysis. *BMJ* 1999;319(7211):666–70.
27 Becker U, Grønbaek M, Johansen D, Sorensen TI. Lower risk for alcohol-induced cirrhosis in wine drinkers. *Hepatology* 2002;35(4):868–75.
28 Lakins NE, LaVallee RA, Williams GD, Yi H. (2008) Surveillance Report No. 85: Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends, 1977–2006. National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institutes of Health, Bethesda, MD.
29 Bellentani S, Saccoccio G, Costa G, et al. Drinking habits as cofactors of risk for alcohol induced liver damage. The Dionysos Study Group. *Gut* 1997;41(6):845–50.
30 Bellentani S, Tiribelli C, Saccoccio G, et al. Prevalence of chronic liver disease in the general population of northern Italy: The Dionysos Study. *Hepatology* 1994;20(6):1442–9.
31 Mathurin P, Deltenre P. Effect of binge drinking on the liver: an alarming public health issue? *Gut* 2009;58(5):613–7.
32 Li TK. Quantifying the risk for alcohol-use and alcohol-attributable health disorders: present findings and future research needs. *J Gastroenterol Hepatol* 2008;23(Suppl 1):S2–8.
33 Naimi TS, Brewer RD, Mokdad A, et al. Binge drinking among US adults. *JAMA* 2003;289(1):70–5.
34 McCarty CA, Ebel BE, Garrison MM, et al. Continuity of binge and harmful drinking from late adolescence to early adulthood. *Pediatrics* 2004;114(3):714–9.
35 Newes-Adeyi G, Chen CM, Williams GD, Faden VB. (2007) Surveillance Report No. 81: Trends in underage drinking in the United States, 1991–2005. National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institutes of Health, Bethesda, MD.
36 Thomas DL, Astemborski J, Rai RM, et al. The natural history of hepatitis C virus infection: host, viral, and environmental factors. *JAMA* 2000;284(4):450–6.
37 Bedogni G, Miglioli L, Masutti F, et al. Natural course of chronic HCV and HBV infection and role of alcohol in the general population: The Dionysos Study. *Am J Gastroenterol* 2008;103(9):2248–53.
38 Asrani SK, Larson JJ, St. Sauver JL, Kim WR. Impact of obesity on alcoholic hepatitis: A population-based study. *Gastroenterology* 2011;140(5):S-914.
39 Naveau S. Body mass index and risk of liver cirrhosis in middle-aged UK women: prospective study. *Gastroenterol Clin Biol* 2010;34(8–9):429–30.
40 Liu B, Balkwill A, Reeves G, Beral V. Body mass index and risk of liver cirrhosis in middle-aged UK women: prospective study. *BMJ* 2010;340:c912.
41 Trepo E, Gustot T, Degre D, et al. Common polymorphism in the *PNPLA3*/adiponutrin gene confers higher risk of cirrhosis and liver damage in alcoholic liver disease. *J Hepatol* 2011;55(4):906–12.
42 Stickel F, Buch S, Lau K, et al. Genetic variation in the *PNPLA3* gene is associated with alcoholic liver injury in Caucasians. *Hepatology* 2011;53(1):86–95.
43 Tian C, Stokowski RP, Kershenobich D, et al. Variant in *PNPLA3* is associated with alcoholic liver disease. *Nat Genet* 2010;42(1):21–3.
44 Rotman Y, Koh C, Zmuda JM, et al. The association of genetic variability in patatin-like phospholipase domain-containing protein 3 (PNPLA3) with histological severity of nonalcoholic fatty liver disease. *Hepatology* 2010;52(3):894–903.
45 McClain C, Barve S, Joshi-Barve S, et al. Dysregulated cytokine metabolism, altered hepatic methionine metabolism and proteasome dysfunction in alcoholic liver disease. *Alcohol Clin Exp Res* 2005;29(11 Suppl):180S–8S. [Review]
46 Monzoni A, Masutti F, Saccoccio G, et al. Genetic determinants of ethanol-induced liver damage. *Mol Med* 2001;7(4):255–62.

47 Bosetti C, Levi F, Lucchini F, et al. Worldwide mortality from cirrhosis: an update to 2002. *J Hepatol* 2007;46(5):827–39.
48 Yoon YH, Yi HY (2008) Surveillance Report No. 83: Liver cirrhosis mortality in the United States, 1970–2005. National Institute on Alcohol Abuse and Alcoholism, Division of Epidemiology and Prevention Research, Alcohol Epidemiologic Data System.
49 Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. *Lancet* 2006;367(9504):52–6.
50 Paula H, Asrani SK, Boetticher NC, et al. Alcoholic liver disease-related mortality in the United States: 1980–2003. *Am J Gastroenterol* 2010;105(8):1782–7.
51 Potter JD. Hazards and benefits of alcohol. *New Engl J Med* 1997;337(24):1763–4.
52 Thun MJ, Peto R, Lopez AD, et al. Alcohol consumption and mortality among middle-aged and elderly U.S. adults. *New Engl J Med* 1997;337(24):1705–14.
53 Asrani SK, Larson JJ, Benson J, et al. Survival of patients with alcoholic hepatitis: a population-based study. *Hepatology* 2010;52(S1):1A–1352A.
54 Vong S, Bell BP. Chronic liver disease mortality in the United States, 1990–1998. *Hepatology* 2004;39(2):476–83.
55 Chen CM, Yoon YH, Yi HY, Lucas DL. Alcohol and hepatitis C mortality among males and females in the United States: a life table analysis. *Alcohol Clin Exp Res* 2007;31(2):285–92.
56 Harzke AJ, Baillargeon J, Paar DP, et al. Chronic liver disease mortality among male prison inmates in Texas, 1989–2003. *Am J Gastroenterol* 2009;104(6):1412–19.
57 Singh GK, Hoyert DL. Social epidemiology of chronic liver disease and cirrhosis mortality in the United States, 1935–1997: trends and differentials by ethnicity, socioeconomic status, and alcohol consumption. *Hum Biol* 2000;72(5):801–20.
58 Harford TC, Brooks SD. Cirrhosis mortality and occupation. *J Stud Alcohol* 1992;53(5):463–8.
59 Chen C, Yi H. (2008) Surveillance Report No. 83. Liver cirrhosis mortality in the United States, 1970–2005. National Institute on Alcohol Abuse and Alcoholism, Division of Epidemiology and Prevention Research, Alcohol Epidemiologic Data System.
60 Berg CL, Steffick DE, Edwards EB, et al. Liver and intestine transplantation in the United States 1998–2007. *Am J Transplant* 2009;9(4 Pt 2):907–31.
61 McKenna MT, Michaud CM, Murray CJ, Marks JS. Assessing the burden of disease in the United States using disability-adjusted life years. *Am J Prev Med* 2005;28(5):415–23.
62 Shah VH. Alcoholic liver disease: the buzz may be gone, but the hangover remains. *Hepatology* 2010;51(5):1483–4.
63 Greenberg RS, Daniels SR, Flanders WD, et al. *Medical Epidemiology*, 4th ed., Lange Medical Books/McGraw-Hill, New York, 2005.
64 Harris D, Brunt P. Prognosis of alcoholic liver disease – 100 years on and the need for international standards and guidelines. *Alcohol Alcohol* 1995;30(5):591–600.
65 Mandayam S, Jamal MM, Morgan TR. Epidemiology of alcoholic liver disease. *Semin Liver Dis* 2004;24(3):217–32.
66 Kim WR, Brown RS, Jr., Terrault NA, El-Serag H. Burden of liver disease in the United States: summary of a workshop. *Hepatology* 2002;36(1):227–42.
67 Manos MM, Leyden WA, Murphy RC, et al. Limitations of conventionally derived chronic liver disease mortality rates: Results of a comprehensive assessment. *Hepatology* 2008;47(4):1150–7.
68 McGinnis JM, Foege WH. Actual causes of death in the United States. *JAMA* 1993;270(18):2207–12.
69 Maddrey WC, Boitnott JK, Bedine MS, et al. Corticosteroid therapy of alcoholic hepatitis. *Gastroenterology* 1978;75(2):193–9.
70 Bell BP, Manos MM, Zaman A, et al. The epidemiology of newly diagnosed chronic liver disease in gastroenterology practices in the United States: results from population-based surveillance. *Am J Gastroenterol* 2008;103(11):2727–36; quiz 37.
71 Yang AL, Vadhavkar S, Singh G, Omary MB. Epidemiology of alcohol-related liver and pancreatic disease in the United States. *Arch Intern Med* 2008;168(6):649–56.
72 Fischer GE, Bialek SP, Homan CE, et al. Chronic liver disease among Alaska-Native people, 2003–2004. *Am J Gastroenterol* 2009;104(2):363–70.

Answers to multiple choice questions

1. C

The AUDIT questionnaire is a tool used to screen for alcohol abuse disorders. The remaining choices are mathematical models used to predict the severity of alcoholic hepatitis.

2. D

The one-year prevalence of alcohol use disorders (alcohol abuse and dependence) among persons aged 15–64 was 3.6 % worldwide. The highest rate (10.9 %) was observed in eastern Europe.

3. B
Younger age at the onset of alcohol use as well as a binge pattern of alcohol use are associated with increased frequency of alcohol use disorders. The other choices have been associated with an increased risk of alcoholic liver disease.

4. D
Death rates due to alcoholic liver disease are highest among American Indian/Alaskan Natives. The death rate is higher among men than women and highest among persons aged 55–64 years of age.

31 Epidemiology of cirrhosis and hepatocellular carcinoma

Joe West[1] *& Guruprasad P. Aithal*[2]
[1]Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK
[2]NIHR Nottingham Digestive Diseases Biomedical Research Unit, Nottingham University Hospitals NHS Trust and University of Nottingham, Nottingham, UK

Key points

- Alcohol is an important cause of cirrhosis in the Western Hemisphere and Russia but worldwide hepatitis B viral (HBV) infection is the most common cause of chronic liver disease.
- The occurrence of cirrhosis worldwide effectively mirrors the prevalence of both HBV and alcohol consumption and the trends in these risk factors.
- Relative risk of death in people with cirrhosis ranges from a hazard ratio of 5 in compensated disease to 10 in decompensated disease compared to the general population. The 1- and 5-year survival following a diagnosis is comparable to that of some cancers.
- Cirrhosis underlies over 75 % of hepatocellular carcinoma (HCC) and about three-quarters of cases are attributable to chronic viral hepatitis, which is highly prevalent in low- and middle-income countries.
- The most effective global measures to prevent both cirrhosis occurrence and liver cell cancer are probably universal vaccination against HBV of newborn babies and reduction of harmful alcohol consumption.

Cirrhosis

Disease definitions

The term "cirrhosis" was first used by Laennec around the turn of the eighteenth century to describe the histologic appearances in the liver he observed [1]. Yet cirrhosis is far more than a histologic entity; it is the final common clinical pathway for most, if not all, chronic liver diseases. Despite marked structural and architectural changes associated with its development, cirrhosis does not necessarily result in symptoms that are recognizable by those affected with it. Many live with cirrhosis unhindered in the functions of their normal lives and this is known as "compensated" disease. Only when the consequences of architectural changes lead to complications and functional impairment of the liver do clinical problems become apparent ("decompensated" disease). Aside from nonspecific symptoms such as fatigue, cirrhosis manifests primarily as the sequelae of portal hypertension with gastroesophageal variceal formation, ascites, and encephalopathy heralding a period of decompensation. Almost all etiologies underlying cirrhosis are also risk factors for the development of primary liver cancers. Treatment of chronic liver disease is aimed at reducing the progression of fibrosis and hence, development of architectural changes. Once

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

decompensated, management of cirrhosis is focused on reducing the harmful effects of portal hypertension but are essentially palliative. Without liver transplantation, which is only feasible in a minority of patients, mortality among people with decompensated cirrhosis and primary liver cancer is high and both account for a substantial global burden of disease.

Occurrence

In contrast to cancer and infectious diseases, there are not routinely collected data on the occurrence of cirrhosis of the liver worldwide. For most of the twentieth century epidemiologists have had to rely upon mortality statistics to estimate the occurrence of this disease. However, it must be borne in mind that while many people with cirrhosis when diagnosed have decompensated disease and therefore a short life expectancy, a fairly large proportion with compensated disease may survive many years following their diagnosis, so mortality statistics will not be reliable in estimating incidence. Fortunately, recent developments in the field of linked electronic healthcare databases and the publication of incidence rates from inception cohorts have given some approximation of the contemporary occurrence of cirrhosis and alongside this some picture of trends over time. However, estimating incidence through these methods also has limitations. Given that cirrhosis is a chronic disease with a long sojourn time any reported measure of incidence will only be an approximation of the truth. In reality such a measure represents the number of diagnosed new cases per unit time which of course can be hugely altered depending upon definition of disease, healthcare setting, access to resources, and ability to make the diagnosis.

Incidence

Cirrhosis

Few population-based studies exist describing the incidence of cirrhosis prior to 1970. However, since then, data either prospectively collected or retrospectively analyzed have given estimates of incidence (or diagnosis rate) from Iceland, Denmark, Sweden, and the United Kingdom. Saunders et al. [2] gave an indication of the size of the problem around Birmingham, England when they reported a rise in incidence rates from about 5 per 100,000 population in 1959 to a peak of 15 per 100,000 in 1974. They also showed the rise was predominately among people with cirrhosis related to alcohol use (Figure 31.1).

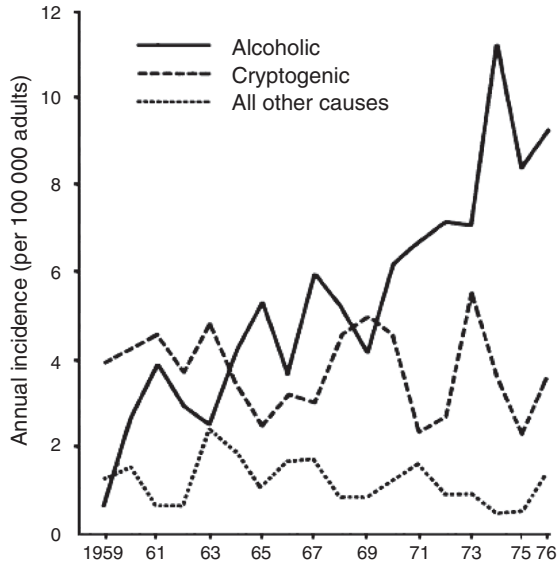


Figure 31.1 Annual incidence of alcoholic, cryptogenic, and other types of cirrhosis in West Birmingham 1959–1976. Source: Saunders et al. 1981 [2]. Reproduced with permission of BMJ Publishing Group Ltd.

Another report using electronic primary care records from the United Kingdom showed a continuing rise through the 1990s in both alcohol and nonalcohol-related cirrhosis in both males and females with an overall increase from 12 per 100,000 to 17 per 100,000 across a 10-year period [3]. Similar incidence rates of alcoholic cirrhosis for men, women and overall have been reported from Denmark [4] for the same periods although an increasing trend was not observed between 1994 and 2005. Incidence was considered to be decreasing reflecting mortality rates across Europe [5]. By contrast, reports from Iceland indicate a far lower overall rate of diagnosis (approximately 3 per 100,000) [6].

More recently still, a report from the Million Women Study [7], which is a cohort study that recruited women (average age 55 years) between 1996 and 2001 in England and Scotland, estimated the rate of first hospitalization or death due to cirrhosis over a mean follow-up of 6 years. A remarkably high rate reported of about 120 per 100,000 person-years is

probably related to the use of very broad categories of ICD codes to define cirrhosis. Overall observations from these studies show that men have about a twofold greater incidence rate than women and that the age of diagnosis is slightly lower among males than females.

Portal hypertension, esophageal bleeding, and ascites

An alternative method of measuring incidence of any disease is to focus instead upon the occurrence of clearly defined serious complications. This has some advantages in that the definition of cirrhosis itself in observational studies differs markedly compared to bleeding esophageal varices. In most countries an event such as bleeding esophageal varices should result in a hospital admission. There may still be some variation in the ascertainment of admissions for esophageal varices, ascites, and hepatic encephalopathy depending on varying clinical practices as well as changes in management that may have altered occurrence over time [8]. Nonetheless, these indices are useful adjuncts to the overall cirrhosis incidence figures.

Data from the United States from 1988–2002 using their National Inpatient Sample showed no clear trend in hospitalization for bleeding esophageal varices from 1994 onwards [9] and the reported absolute rate during this period was 11 per 100,000. In Sweden, the rate of hospitalization for esophageal varices was 5 per 100,000 population. However, this included both bleeding and nonbleeding varices [10]. In both these studies, the ratio of men to women was about 2:1. More recent data from England shows no change between 1999 and 2007 with a rate of bleeding esophageal varices of 3 per 100,000 [11]. Unfortunately, few population-based studies are available that report the hospital admission rate related to ascites, encephalopathy, or spontaneous bacterial peritonitis.

Prevalence

Prevalence of cirrhosis is key to understanding healthcare utilization due to the burden of morbidity. People with cirrhosis do regularly seek or need both primary and secondary health care. Recent database studies from the United Kingdom and Denmark have estimated the prevalence of cirrhosis. In the UK the estimated point prevalence in 2001 was 87 per 100,000 in men and 66 per 100,000 in women. When applied to the population demographics this approximated to 30,000 people living with the disease in the United Kingdom at the time [3]. In Denmark the figures for alcoholic cirrhosis during the period 2001–2005 were 132 per 100,000 in men and 70 per 100,000 in women [4].

Risk factors

Alcohol is an important cause of cirrhosis in the Western Hemisphere and Russia but worldwide hepatitis B viral (HBV) infection is the most common cause of chronic liver disease and it is particularly highly prevalent in Southeast Asia and sub-Saharan Africa. Chronic hepatitis C continues to add to a large healthcare burden worldwide with the incidence rates varying across the world. In most epidemiologic studies, about 30 % of people diagnosed with cirrhosis do not have an obvious etiology.

Alcohol

Across the United States and Europe, approximately 50–60 % of cirrhosis has alcohol as a central component cause. This is perhaps unsurprising given the alcohol consumption per capita of these areas as shown in the map of the world below (Figure 31.2). Equally, it is unsurprising that rates of cirrhosis have been observed to be rising in the United Kingdom as alcohol consumption has been rising since 1948.

Viral hepatitis

Prevalence of HBV infection ranges from over 10 % in Asia to under 0.5 % in the United States and northern Europe. In untreated individuals with predominantly HBeAg positive HBV infection, the incidence of cirrhosis ranges from 2 to 5.4 per 100 person-years with a 5-year cumulative incidence of cirrhosis of 8–20 % [13]. Older age at presentation (or time of infection) and persistent viral replication are predictors for development of cirrhosis as well as mortality. The presence of any other independent hepatotoxic factors such as alcohol ingestion or HCV co-infection can contribute to progression to cirrhosis. About 180 million people worldwide are thought to be infected by hepatitis C virus [14] and 3–20 % of untreated patients develop cirrhosis [15]. Risk factors that are considered to increase the risk of development of cirrhosis include older age at time of infection, male sex, coinfection with human immunodeficiency virus

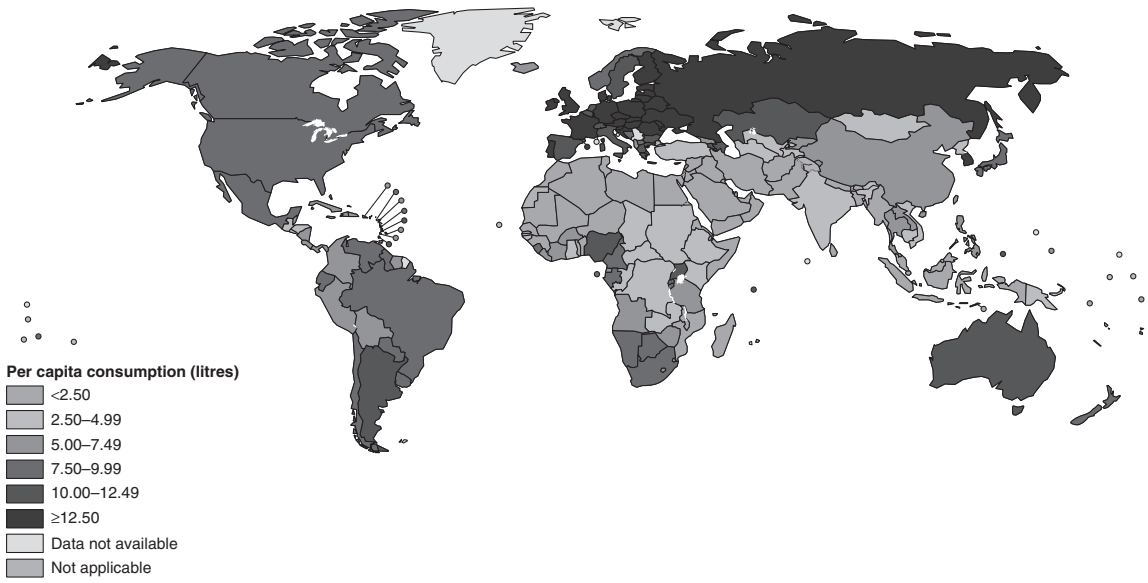


Figure 31.2 World alcohol consumption map. Source: World Health Organization 2011 [12]. Reproduced with permission of the World Health Organization.

(HIV), or hepatitis B virus (HBV), and comorbid conditions such as immunosuppression, insulin resistance, nonalcoholic steatohepatitis, and schistosomiasis [16].

Other risk factors

Several other chronic liver diseases such as hemochromatosis, alpha-1 antitrypsin deficiency, Wilson's disease, auto-immune hepatitis, primary biliary cirrhosis, primary sclerosing cholangitis, cystic fibrosis, and hepatotoxic drugs are among the causes of cirrhosis. When etiology underlying cirrhosis is not obvious, it is termed "cryptogenic". As around 75 % of patients with cryptogenic cirrhosis have a history of type 2 diabetes or obesity, it is considered that much of what is classified as "cryptogenic" cirrhosis could be the end stage of nonalcoholic fatty liver disease (NAFLD) [17].

Natural history

Although cirrhosis may have developed many years prior to its identification, once diagnosed, cirrhosis follows a typical course that includes a mostly asymptomatic period (compensated disease) followed by a symptomatic period (decompensated disease) [18]. The latter phase is essentially characterized by the presence of complications related to portal hypertension. The rate at which compensated disease becomes decompensated disease is less easy to quantify primarily due to the variation in definitions, regularity of surveillance programs for esophageal varices and recording of this process routinely. Recent use of large healthcare databases has allowed contemporary estimation of the rate of decompensation among people newly diagnosed with cirrhosis. Figure 31.3 shows that once diagnosed there is a high rate of early decompensation (defined as the occurrence of ascites or gastrointestinal bleeding in this study), which is likely to be simply part of the workup and assessment of individuals initially. However, these data also show that if individuals remain compensated after one year of follow-up, they then have a rate of decompensation of about 6 % per year.

Data reported from Denmark have recently detailed the progression of disease in alcoholic cirrhosis in a similar fashion and show high probabilities of progression to the specific consequences of portal hypertension. For example, among 114 patients with no complications at diagnosis after 1 year, 68 % were alive and complication-free, 15 % were alive but had developed complications, 10 % had died without developing complications, and 7 % had died after developing

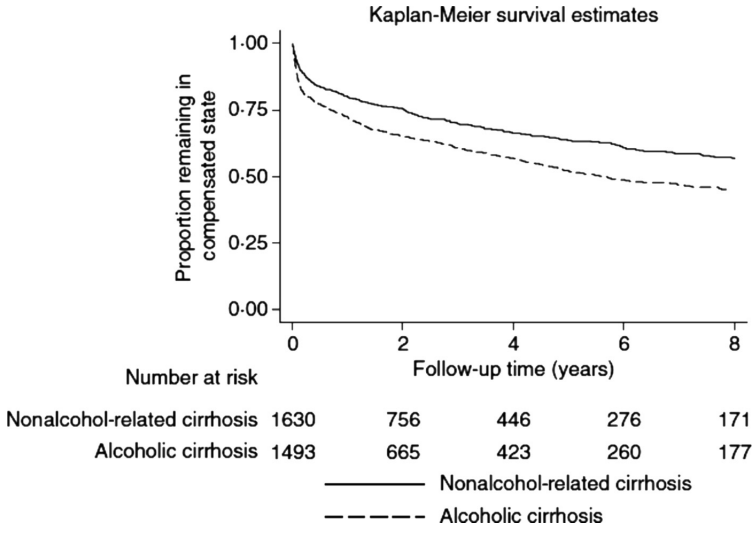


Figure 31.3 Rate of decompensation among people with compensated alcoholic and nonalcoholic-related cirrhosis in the UK. Source: Fleming et al. 2008 [3]. Reproduced with permission of Elsevier.

complications. After 5 years, the corresponding proportions were 28 %, 13 %, 22 %, and 35 %.

Depending on the population studied and at what time point in the disease course, the proportion of people with cirrhosis and gastroesophageal varices varies considerably. Estimates vary from approximately 30 to 50 % of patients [19,20]. Patients without varices develop them at a rate of between 4 and 8 % per year [19–21] while variceal hemorrhage reportedly occurs at a yearly rate of 5 to 15 % [22].

Survival

Overall

While progression of disease has been described in various ways and in many studies [23], estimates of the long-term survival of people with cirrhosis of the liver following diagnosis that have been generated from population-based studies are few and far between. In 1981, Sanders et al. described the survival of 512 people admitted to hospital with cirrhosis in the West Midlands region of the United Kingdom between 1959 and 1976 [2]. They observed high mortality rates in the first year following admission and 5-year survival rates of 36 %, 14 %, and 14 % for groups with alcoholic, cryptogenic, and post-hepatitic cirrhosis, respectively. More recently, case fatality rates for people admitted to hospital with liver cirrhosis have been described for the period 1968–1999 in the Oxford region of the United Kingdom [24]. In the first month following admission, the case fatality rate was 15 % and at 1 year 33 % with no real change over the period of the study. In Denmark, among approximately 15,000 patients with cirrhosis identified through hospital registry data between 1995 and 2006 [25], poor 1- and 10-year survival figures were observed of 65 % and 21 %, respectively. Broadly these studies show concordance and disappointingly little change in the reported survival over time.

Compensated versus decompensated disease

While overall survival has not changed much over the course of the last 40 years there is some evidence that it has when cirrhosis is divided into those with compensated and decompensated disease. However, this is dependent on the definitions of this categorization. For example, D'Amico and colleagues in 1986 reported the 6-year survival of 54 % and 21 % in patients with compensated and decompensated cirrhosis respectively from a small population in Sicily [26]. These stark differences were also apparent in the West Midlands in the period 1959 to 1976 but for the compensated group the survival rates appear to be somewhat improved. Data from the General Practice Research Database in the United Kingdom from 1992 to 2001 indicate that for those patients with cirrhosis defined as compensated at diagnosis, the 1- and 5-year survival rates are 87 % and 66 %, respectively whereas for decompensated disease, the rates are 75 % and 45 %. Similarly, in people hospitalized with

esophageal varices in Sweden between 1969 and 2002 [10] there also appears to have been some improvement in survival over time. These figures suggest that medical intervention whether as prevention or therapeutic has had a positive impact.

Comparisons to the general population

Given these poor survival figures, unsurprisingly, cirrhosis of the liver decreases a patient's life expectancy in comparison to the general population. Indeed the survival figures are comparable to some common cancers such as those of the colorectum and breast. How much extra risk people with cirrhosis are at compared to the general population has been estimated in terms of the relative increase in risk of death and ranges from a standardized mortality ratio of 12 (12-fold increase) [24] at 1 year following hospitalization to a hazard ratio of 5 in compensated disease and 10 in decompensated disease for those identified in primary care in the United Kingdom. This increased rate of death appears to be independent of other comorbidities [25].

Mortality

The great advantage of mortality statistics is that in some parts of the world recording of fact, date and cause of death has been routine for many years. It is therefore possible to describe causes of death in, for example, the United States and the United Kingdom for most of the twentieth century. During this period there have been numerous changes in the coding of diseases, in particular chronic liver disease and cirrhosis, so interpretation of the figures presented must take into consideration that definitions used varied considerably. In 1967 Terris [27] summarized the available mortality data from 1900 onwards for several countries and showed that from approximately the end of World War II (1945) mortality rates had begun to rise. In England, Wales, and Scotland these rates continued to rise inexorably particularly in those aged 15 to 45 for the rest of the twentieth century, which is in direct contrast to the rest of Europe [5,28]. From these mortality figures the overall rate of death from cirrhosis in England and Wales was seen to rise from 3.4 per 100,000 in men in 1957–1961 to 14.1 per 100,000 in 1997–2001 (i.e. a fourfold increase) [28]. There was a similarly rapid increase among women. However, the latter study used a broad definition of cirrhosis that included ICD-10 coding for alcoholic liver disease (K70) and chronic hepatitis (K73) as a cause of death, which may have inflated their estimates.

Hepatocellular carcinoma

Disease definition

Hepatocellular carcinoma (HCC) is the third leading cancer-related cause of death and the seventh most common form of cancer worldwide. Cirrhosis underlies over 75 % of HCC and the relative risk of liver cell cancer varies according to the underlying etiology of the cirrhosis. About three-quarters of HCC are attributed to chronic HBV and HCV infections and global variations in incidence rates of this cancer closely reflects the prevalence of HBV and HCV infections. Alcoholic liver disease is the second most common risk factor for HCC. With increasing prevalence of metabolic syndrome and diabetes mellitus, nonalcoholic steatohepatitis (NASH) is likely to contribute to the increasing proportion of cases of HCC. In areas endemic for HBV, viral transmission occurs at an early age, and affected individuals develop HCC in mid-adulthood, their most productive years of life, accounting for a substantial burden on the health services. For most patients, liver cell cancer is a terminal complication of cirrhosis without any access to potentially curative interventions. Probably the most effective measure to prevent HCC is universal vaccination of newborn babies; this has been shown to result in dramatic reduction in HBV infection as well as reduce the incidence of HCC.

Incidence

While cancer occurrence is recorded in many countries worldwide, there is substantial variation in coverage and quality of these data [29] and the coding used to define the disease of interest is not consistent. Most studies describing the occurrence of and trends in primary liver cancer use ICD-10 code C22 and all its subsidiary codes to allow international comparisons. However, this grouping includes intrahepatic biliary cancer and unspecified liver cancers. Using this definition, liver cancer is the fifth most common cancer in men worldwide (523,000 cases, 7.9 % of the total) and the seventh in women (226,000 cases, 6.5 % of the total). Most of the burden is in developing countries,

where almost 85 % of the cases occur. This cancer particularly affects men with an overall sex ratio male-to-female of 2:4. The regions of high incidence are eastern and southeastern Asia, middle and western Africa, but also Melanesia and Micronesia/Polynesia (particularly in men). Low rates are estimated in developed regions, with the exception of southern Europe where the incidence in men (Age Standardized Rate 10.5 per 100,000) is far higher than in other developed regions [29].

Risk factors

Cirrhosis

Over 75 % of HCC develops in chronic liver disease with advanced fibrosis or cirrhosis [30]. All causes of cirrhosis can be considered risk factors for HCC although the relative risks may vary according to etiology [31]. There are few good studies of the occurrence rates of HCC in people with cirrhosis and some are detailed in Table 31.1. It would appear that HCV infection is associated with the highest HCC incidence in people with cirrhosis, with a 5-year cumulative incidence of 30 % in Japan (17 % in Western countries). In HBV-related cirrhosis, the 5-year cumulative HCC risk is 15 % in high endemic areas and 10 % in Western countries. Co-infection with HCV and HBV increase the risk of HCC two- to sixfold relative to each infection on its own. Incidence of HCC is lower in alcoholic cirrhosis (5-year cumulative incidence of 8 %), but excess alcohol is thought to increase the risk of HCC by two- to fourfold in those with chronic viral hepatitis. Traditionally, HCC has been considered a rare complication of cirrhosis secondary to primary biliary cirrhosis and autoimmune hepatitis. For example, in patients with primary biliary cirrhosis the absolute excess risk per year compared to the general population is only 0.2 % [32] and HCC is estimated to develop at an annual rate of about 1 % in cirrhosis secondary to autoimmune hepatitis [33].

HBV infection

The development of HCC is one of the main causes of death in people with HBV infection. Chronically infected subjects have a 100 times increased risk of HCC compared with noncarriers [34]. Several studies have indicated that male sex, age over 45 years, longer duration of viral infection, positive HBeAg, co-infection with HCV or hepatitis delta virus infection, having a first-degree relative with HCC, consumption of aflatoxin in diet, and the presence of cirrhosis are independent factors for HCC in HBV-infected patients [35–37].

Obesity

In a large prospective cohort carried out in the United States of more than 900,000 adults, the heaviest men and women (with a BMI $\geq$40.0), had a higher rate of death due to cancer of the liver when compared with men and women of normal weight [38]. The relative risk of death from liver cancer in people with highest BMI compared to normal was about 1.7 in women and 4.5 in men. In men, the relative risk for liver cancer was highest of all the cancers. Potential biologic explanations for these observations include an increased level of sex hormones, insulin and insulin-like growth factor 1, which are associated with obesity. Obesity also appears to interact synergistically with alcohol and smoking to increase the risk of HCC [39].

Table 31.1 Incidence of hepatocellular carcinoma in cirrhosis

	Number of patients (number of studies)	Geographic location	Incidence (per 1000 person-years) [95 % CI]
Compensated cirrhosis	1284 (13)	Europe, USA	37 [32,42]
Compensated cirrhosis	626 (7)	Japan	71 [62,80]
HBV compensated cirrhosis	401 (6)	Europe	22 [16,28]
Alcoholic cirrhosis (no HBV/HCV)	174 (3)	Europe	17 [12,22]

Source: Adapted from Fattovich 2004 [31].

Diabetes

Diabetes is associated with a two- to threefold increase in the risk of HCC independent of HCV, HBV, alcoholic liver disease, and hemochromatosis [40]. Diabetes, as part of the insulin resistance syndrome, has been implicated as a risk factor for NAFLD, including its most severe form NASH. NASH has been identified as a suspected cause of both "cryptogenic" cirrhosis and HCC. In addition, significant synergy exists between heavy alcohol consumption and chronic hepatitis virus infection and diabetes mellitus in relation to HCC development suggesting common pathways of carcinogenesis [41].

Alcohol

Alcoholic liver disease is the second most common risk factor for HCC in the United States, and alcoholic cirrhosis is a major contributor to HCC in the United Kingdom. In women, even low to moderate alcohol consumption can increase the risk of liver cancer [42]. In addition, alcohol is thought to act synergistically with other risk factors to magnify the risk of HCC.

Aflatoxin

Aflatoxins are naturally occurring mycotoxins that are produced by many species of the fungus Aspergillus. They can colonize grain before harvest or during storage. Aflatoxins frequently contaminate food in sub-Saharan Africa and eastern Asia. Crops frequently affected include cereals, oilseeds, spices, and tree nuts. Aflatoxins are metabolized by the liver to a reactive epoxide that causes mutations in the p53 gene and attenuates its tumor suppressor function [43]. Aflatoxin can also act synergistically with HBV in the pathogenesis of HCC [44].

Survival

Survival following a diagnosis of liver cancer is poor and worldwide is probably only 5 % at 1 year. However, over the period 1971–1999 in England and Wales 1-year survival has increased from around 5 % to nearly 20 % for both men and women and survival at 5 years for the same time period has increased from almost zero to about 5 %. These improving figures most likely reflect earlier diagnosis and therefore earlier stage disease rather than radical improvements in the therapeutic options for treating this cancer. That said, access to therapeutic interventions such as transplantation or ablation is likely to vary considerably around the world.

Mortality

Worldwide there were an estimated 694,000 deaths from liver cancer in 2008 (477,000 in men, 217,000 in women), and because of its high fatality (overall ratio of mortality to incidence of 0.93), liver cancer is the third most common cause of death from cancer worldwide. Therefore the geographical distribution of the mortality rates is similar to that observed for incidence.

Trends in the incidence and mortality of HCC have been described for the United States as well as England and Wales. In the United Kingdom, the incidence has increased in both males and females between 1975 and 2007 with a greater relative increase among men with similar trends observed in the United States [45, 46]. These trends are presumably due to the linked increase in cirrhosis described earlier in this chapter, or to related risk factors. In the United Kingdom, HCC represents the most common malignancy of the liver, gallbladder, and biliary system. However, in contrast to the worldwide figure of approximately 7 % of all cancers, it represents only 1 % of all cancers diagnosed in the UK.

Age

Across the world the incidence of this cancer increases with age up to about age 80 where it then begins to dip. In the United Kingdom, the greatest age-specific increase in incidence over time has been seen in those aged between 70 and 80 years.

Sex

Overall men have far higher absolute rates of primary liver cancer than women. For example, when cancer data for England and Wales were restricted to those with histology available and analyzed by type of liver cancer a threefold increase in hepatocellular carcinoma was seen among men compared with only a steady increase for women over the period 1971–2001 [45].

Ethnicity

From the available Surveillance, Epidemiology and End Results (SEER) data from the United States it

Table 31.2 Incidence rates of HCC by ethnicity and sex in the United States

	Incidence rates by race	
Race/ethnicity	Male	Female
All races	11.2 per 100,000 men	3.9 per 100,000 women
White	9.6 per 100,000 men	3.3 per 100,000 women
Black	15.1 per 100,000 men	4.6 per 100,000 women
Asian/Pacific Islander	22.1 per 100,000 men	8.4 per 100,000 women
American Indian/Alaska Native	17.4 per 100,000 men	7.8 per 100,000 women
Hispanic	15.8 per 100,000 men	6.2 per 100,000 women

Source: Howlader et al. (eds). *SEER Cancer Statistics Review, 1975–2008*, National Cancer Institute, Bethesda, MD; http://seer.cancer.gov/csr/1975_2008/, based on November 2010 SEER data submission, posted to the SEER website, 2011.

appears that rates of liver cancer (C22) do vary considerably by ethnicity in the USA for men and women alike (Table 31.2), and this presumably reflects the distribution of environmental risk factors for the disease given the lack of a specific genotype recognized to cause this cancer.

Prevention

Probably the most effective measure from a worldwide perspective to prevent both cirrhosis occurrence and liver cell cancer is universal vaccination against HBV of newborn babies; this has been shown to result in a dramatic reduction in HBV infection as well as reducing the incidence of HCC [47]. Clearly, among high alcohol-consuming populations, efforts to curb harmful alcohol intake are also likely to have a large impact if successful. Among people at high risk of HCC such as those infected with HBV, surveillance with a combination of regular alpha-feto protein test and ultrasonography to detect early lesions may have the potential to reduce mortality from HCC [48].

Conclusions

Cirrhosis

Chronic liver disease and cirrhosis are among the 10 leading causes of deaths in the United States and in the United Kingdom. Mortality from cirrhosis has increased markedly over the last few decades, in striking contrast to falling mortality from heart, kidney, and respiratory diseases, strokes, and cancers. Chronic viral hepatitis, alcohol abuse, and nonalcoholic fatty liver disease account for the vast majority of cases of cirrhosis worldwide. Cirrhosis is associated with poor survival with a relative increase in mortality rate of 5 for "compensated" and 10 for "decompensated" state compared to the general population. Recent data on survival for patients with evidence of portal hypertension suggest that medical interventions whether preventative or therapeutic have had some positive impact.

HCC

HCC is the third leading cancer-related cause of death accounting for a substantial burden on the healthcare system. About three-quarters of HCC are attributed to chronic viral hepatitis which is highly prevalent in low- and middle-income countries. Alcoholic liver disease probably accounts for a substantial proportion of HCC in high-income countries even though the incidence is relatively low among those with alcoholic cirrhosis compared to some other etiologies.

Multiple choice questions

1 Cirrhosis associated with marked structural and architectural changes in the liver is the final common clinical pathway for most, if not all, chronic liver diseases.

With regard to the etiology of cirrhosis:

A Aflotoxin acts synergistically with chronic hepatitis C leading to cirrhosis

B Excess alcohol consumption globally is the commonest component cause of cirrhosis

C Hepatitis A interacts with HBV infection increasing the risk of developing cirrhosis
D If untreated HCV infection leads to cirrhosis in the majority of patients
E Nonalcoholic fatty liver disease underlies the majority of "cryptogenic" cirrhosis

2 Several chronic liver diseases are associated with the development of progressive fibrosis which over a period of years leads to cirrhosis. Lack of liver-specific symptoms despite marked structural and architectural changes within the liver makes it difficult to identify the point at which cirrhosis has been established.

With regard to the natural history of cirrhosis:
A Extra risk of mortality associated with cirrhosis has been demonstrated only when cirrhosis reaches a "decompensated" phase
B Majority of the complications of cirrhosis are related to the loss of its synthetic function leading to decompensation
C Presence or absence of symptoms determines the classification of cirrhosis into "compensated" and "decompensated" phase of cirrhosis
D Rate of decompensation is higher in alcoholic cirrhosis when compared with other etiology
E Rate of development of portal hypertension and decompensation increases progressively every year following the diagnosis of cirrhosis

3 Persistent inflammation in chronic liver diseases markedly increases the risk of primary liver cell cancer. Hepatocellular carcinoma (HCC) is the third leading cancer-related cause of death and the seventh most common form of cancer worldwide.

When considering the risk factors to HCC:
A Diabetes increases the risk of HCC by interacting with obesity
B In patients with cirrhosis, HBV infection has the highest cumulative incidence of HCC
C Majority of HCC develop in the liver with advanced fibrosis or cirrhosis
D Risk of HCC is higher with alcoholic cirrhosis when compared with cirrhosis from other etiology
E Women are at higher risk of developing HCC compared to men

4 Recent developments in the field of linked electronic healthcare databases and the publication of incidence rates from inception cohorts have given some approximation of the contemporary occurrence of cirrhosis and its burden on health services.

With regard to the time trends in cirrhosis:
A Although hospitalization due to ascites has remained constant, incidence of spontaneous bacterial peritonitis has increased
B Hospitalization for bleeding varices has fallen in the United States over a period of two decades
C Incidence of cirrhosis has increased over a period of two decades in the United Kingdom
D Incidence rate of cirrhosis in women has risen more steeply changing the male-to-female ratio
E Mortality from cirrhosis has generally increased in the majority of countries in Europe where data is available

5 The development of hepatocellular carcinoma (HCC) is one of the main causes of death in people with HBV infection. Chronically infected subjects have a 100 times increased risk of HCC compared with noncarriers.

The following is associated with an increased risk of HCC in HBV-infected subjects:
A Co-infection with hepatitis E virus
B Degree of inflammation on histology
C Family history of HBV infection
D Infection acquired after 45 years of age
E Male sex

References

1 Duffin JM. Why does cirrhosis belong to Laennec? *CMAJ* 1987;137(5):393–6.
2 Saunders JB, et al. A 20-year prospective study of cirrhosis. *BMJ (Clin Res Ed)* 1981;282(6260):263–6.
3 Fleming KM, et al. Incidence and prevalence of cirrhosis in the United Kingdom, 1992–2001: a general population-based study. *J Hepatol* 2008;49(5):732–8.
4 Jepsen P, Vilstrup H, Sorensen HT. Alcoholic cirrhosis in Denmark – population-based incidence, prevalence, and hospitalization rates between 1988 and 2005: a descriptive cohort study. *BMC Gastroenterol* 2008;8:3.
5 Ramstedt M. Per capita alcohol consumption and liver cirrhosis mortality in 14 European countries. *Addiction* 2001;96(Suppl 1):S19–33.
6 Gunnarsdottir SA. Liver cirrhosis in Iceland and Sweden: incidence, aetiology and outcomes. *Scand J Gastroenterol* 2009;44(8):984–93.
7 Liu B. Body mass index and risk of liver cirrhosis in middle-aged UK women: prospective study. *BMJ* 2010;340:c912.

8 Garcia-Tsao G, Bosch J. Management of varices and variceal hemorrhage in cirrhosis. *New Engl J Med* 2010;362(9):823–32.
9 Jamal MM. Declining hospitalization rate of esophageal variceal bleeding in the United States. *Clin Gastroenterol Hepatol* 2008;6(6):689–95; quiz 605.
10 Stokkeland K. Improved prognosis for patients hospitalized with esophageal varices in Sweden 1969–2002. *Hepatology* 2006;43(3):500–5.
11 Crooks CJ, West J, Card T. Is variceal bleeding increasing within the UK population? *Gut* 2011;60(Suppl 1):A9–10.
12 World Health Organization (WHO) (2011) Global Status Report on Alcohol and Health, WHO, Geneva, pp. 1–57, 273–86.
13 Fattovich G. Natural history of hepatitis B. *J Hepatol* 2003;39(Suppl 1):S50–8.
14 Global surveillance and control of hepatitis C. Report of a WHO Consultation organized in collaboration with the Viral Hepatitis Prevention Board, Antwerp, Belgium. *J Viral Hepat* 1999;6(1):35–47.
15 Seeff LB, Hoofnagle JH. National Institutes of Health Consensus Development Conference: management of hepatitis C: 2002. *Hepatology* 2002;36(5 Suppl 1): S1–2.
16 Chen SL, Morgan TR. The natural history of hepatitis C virus (HCV) infection. *Int J Med Sci* 2006;3(2):47–52.
17 Caldwell SH. Cryptogenic cirrhosis: clinical characterization and risk factors for underlying disease. *Hepatology* 1999;29(3):664–9.
18 Gines P. Compensated cirrhosis: natural history and prognostic factors. *Hepatology* 1987;7(1):122–8.
19 Fleming KM. The rate of decompensation and clinical progression of disease in people with cirrhosis: a cohort study. *Aliment Pharmacol Ther* 2010;32(11–12): 1343–50.
20 Merli M. Incidence and natural history of small esophageal varices in cirrhotic patients. *J Hepatol* 2003; 38(3):266–72.
21 Groszmann RJ. Beta-blockers to prevent gastroesophageal varices in patients with cirrhosis. *New Engl J Med* 2005;353(21):2254–61.
22 Prediction of the first variceal hemorrhage in patients with cirrhosis of the liver and esophageal varices. A prospective multicenter study. *New Engl J Med* 1988; 319(15):983–9.
23 D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol* 2006;44(1):217–31.
24 Roberts SE, Goldacre MJ, Yeates D. Trends in mortality after hospital admission for liver cirrhosis in an English population from 1968 to 1999. *Gut* 2005;54(11):1615–21.
25 Jepsen P. Comorbidity and survival of Danish cirrhosis patients: a nationwide population-based cohort study. *Hepatology* 2008;48(1):214–20.
26 de Franchis R. Evolving consensus in portal hypertension. Report of the Baveno IV consensus workshop on methodology of diagnosis and therapy in portal hypertension. *J Hepatol* 2005;43(1):167–76.
27 Terris M. Epidemiology of cirrhosis of the liver: National Mortality Data. *Am J Public Health* 1967;57(12):2076–88.
28 Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. *Lancet* 2006;367(9504):52–6.
29 Ferlay J. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127(12): 2893–917.
30 Bralet MP. Hepatocellular carcinoma occurring in nonfibrotic liver: epidemiologic and histopathologic analysis of 80 French cases. *Hepatology* 2000;32(2): 200–4.
31 Fattovich G. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology* 2004;127(5 Suppl 1):S35–50.
32 Jackson H. Influence of ursodeoxycholic acid on the mortality and malignancy associated with primary biliary cirrhosis: a population-based cohort study. *Hepatology* 2007;46(4):1131–7.
33 Yeoman AD. Evaluation of risk factors in the development of hepatocellular carcinoma in autoimmune hepatitis: Implications for follow-up and screening. *Hepatology* 2008;48(3):863–70.
34 Hsu YS. Long-term outcome after spontaneous HBeAg seroconversion in patients with chronic hepatitis B. *Hepatology* 2002;35(6):1522–7.
35 McMahon BJ. Hepatitis B-related sequelae. Prospective study in 1400 hepatitis B surface antigen-positive Alaska native carriers. *Arch Intern Med* 1990;150(5):1051–4.
36 Beasley RP. Hepatitis B virus. The major etiology of hepatocellular carcinoma. *Cancer* 1988;61(10):1942–56.
37 Yang HI. Hepatitis B e antigen and the risk of hepatocellular carcinoma. *New Engl J Med* 2002;347(3):168–74.
38 Calle EE. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *New Engl J Med* 2003;348(17):1625–38.
39 Marrero JA. Alcohol, tobacco and obesity are synergistic risk factors for hepatocellular carcinoma. *J Hepatol* 2005;42(2):218–24.
40 Davila JA. Diabetes increases the risk of hepatocellular carcinoma in the United States: a population-based case control study. *Gut* 2005;54(4):533–9.
41 Hassan MM. Risk factors for hepatocellular carcinoma: synergism of alcohol with viral hepatitis and diabetes mellitus. *Hepatology* 2002;36(5):1206–13.

42 Allen NE. Moderate alcohol intake and cancer incidence in women. *J Natl Cancer Inst* 2009;101(5):296–305.
43 Yang JD, Roberts LR. Hepatocellular carcinoma: A global view. *Nat Rev Gastroenterol Hepatol* 2010;7(8): 448–58.
44 Qian GS. A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol Biomarkers Prev* 1994;3(1):3–10.
45 West J. Trends in the incidence of primary liver and biliary tract cancers in England and Wales 1971–2001. *Br J Cancer* 2006;94(11):1751–8.
46 Bosch FX. Primary liver cancer: worldwide incidence and trends. *Gastroenterology* 2004;127(5 Suppl 1):S5–16.
47 Chang MH. Decreased incidence of hepatocellular carcinoma in hepatitis B vaccinees: a 20-year follow-up study. *J Natl Cancer Inst* 2009;101(19):1348–55.
48 Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol* 2004;130(7):417–22.

Answers to multiple choice questions

1. E

A False. Aflatoxin, a natural mycotoxin contaminating the food acts synergistically with HBV in the pathogenesis of HCC.

B False. Across the USA and Europe approximately 50–60 % of cirrhosis has alcohol as the central component cause; but globally, chronic viral hepatitis is the most common etiology underlying cirrhosis.

C False. Interaction between HCV and HBV increases the risk of cirrhosis. HAV is an acute infection and does not cause chronic liver disease.

D False. Between 3 % and 20 % of untreated HCV patients develop cirrhosis.

E *True.* As 74 % of patients with cryptogenic cirrhosis have history of type 2 diabetes or obesity, it is presumed that much of that is classified as "cryptogenic" cirrhosis is an end stage of nonalcoholic fatty liver disease.

2. D

A False. In patients with cirrhosis identified in primary care in the UK, there is a fivefold increase in risk of death in compensated disease and 10-fold risk in decompensated disease.

B False. The majority of complications of cirrhosis are related to the development and consequences of portal hypertension.

C False. Cirrhosis manifests primarily as the sequelae of portal hypertension and appearance of portal hypertension with gastroesophageal variceal formation and ascites even if asymptomatic heralds the period of decompensation.

D *True.* Rate of decompensation is higher in alcoholic cirrhosis when compared with other etiology.

E False. Evidence of portal hypertension in the form of varices is present in approximately 50 % of patients with cirrhosis at the time of diagnosis. If individuals remain compensated after one year of follow-up they then have a rate of decompensation of about 6 % per year.

3. C

A False. Diabetes is associated with a two- to threefold increase in the risk of HCC independent of HCV, HBV, alcoholic liver disease, and hemochromatosis. In addition, significant synergy exists between heavy alcohol consumption and chronic hepatitis virus infection and diabetes mellitus in relation to HCC development suggesting common pathways of carcinogenesis.

B False. Hepatitis C virus (HCV) infection is associated with the highest HCC incidence in people with cirrhosis, with a 5-year cumulative incidence of 30 % in Japan (17 % in Western countries). In hepatitis B virus (HBV)-related cirrhosis, the 5-year cumulative HCC risk is 15 % in high endemic areas and 10 % in Western countries.

C *True.* Over 75 % of HCC develop in chronic liver disease with advanced fibrosis or cirrhosis.

D False. Incidence of HCC incidence is lower in alcoholic cirrhosis (5-year cumulative incidence of 8 %) compared with cirrhosis due to chronic viral hepatitis.

E False. Overall men have far higher absolute rates of primary liver cancer than women. In HBV infection and obesity risk of HCC is higher among men.

4. C

A False. There are not sufficient data to assess the time trends related to ascites, encephalopathy, or spontaneous bacterial peritonitis.

B False. Data from the United States from 1988 to 2002 using their National Inpatient Sample showed no clear trend in hospitalization for bleeding esophageal varices from 1994 onwards.

C *True.* Incidence of cirrhosis has increased over a period of two decades in the United Kingdom.

D False. Overall observations from these studies show that men have about a twofold greater incident rate

than women which suggests that the age of diagnosis is slightly lower among males than females.

E False. Incidence of cirrhosis has been considered to be decreasing reflecting mortality rates across Europe.

5. E

A False. Co-infection with HCV or hepatitis delta virus infection is an independent risk factor in HBV-infected individuals.

B False. Presence of cirrhosis increases the risk of HCC.

C False. Having a first-degree relative with HCC is an independent risk factor.

D False. Age over 45 years, longer duration of viral infection

E *True.* Male sex is an independent risk factor for HBV-related HCC.

32 Epidemiology of nonalcoholic fatty liver disease (NAFLD)

Guruprasad P. Aithal[1], Kshaunish Das[2], & Abhijit Chowdhury[3]

[1]NIHR Nottingham Digestive Diseases Biomedical Research Unit, Nottingham University Hospitals NHS Trust and University of Nottingham, Nottingham, UK
[2]Division of Gastroenterology, School of Digestive and Liver Diseases, IPGME & R, Kolkata, India
[3]Division of Hepatology, School of Digestive and Liver Diseases, IPGME & R, Kolkata, India

Key points

- Nonalcoholic fatty liver disease (NAFLD) encompasses a spectrum of pathology ranging from simple steatosis, steatohepatitis, with a potential to progress to cirrhosis.
- Prevalence of NAFLD has risen throughout the world with urbanization, economic prosperity, and increasing obesity.
- Incidence and severity of NAFLD correlate with the components of metabolic syndrome.
- Nonalcoholic fatty liver disease is strongly associated with risk of incident type 2 diabetes.
- There is evidence to suggest that NAFLD is associated with increased risk of atherosclerosis and cardiovascular events.

Introduction

Nonalcoholic fatty liver disease (NAFLD) is a term that encompasses a spectrum of liver involvement that, defined pathologically, ranges from isolated steatosis to steatohepatitis (nonalcoholic steatohepatitis, NASH) which progresses indolently to cryptogenic cirrhosis (CC) [1–3]. Although obesity associated with fatty liver has been known for a long time [4], the histologic similarity of NASH with certain "pathognomonic" features of alcoholic steatohepatitis delayed recognition of this condition until around three decades ago [5]. This is responsible for the fact that NAFLD is diagnosed by exclusion in an individual consuming no or a negligible amount of alcohol [6]. Although hepatic steatosis and steatohepatitis can occur in numerous metabolic, nutritional, and toxin-induced liver injuries, this chapter will focus on primary NAFLD, which is currently considered to be the hepatic manifestation of metabolic syndrome (MetS), a constellation of metabolic abnormalities including glucose intolerance, obesity (especially, central obesity), dyslipidemia, and hypertension [7,8]. These abnormalities often congregate in individuals, more frequently than by chance alone, and lead to increased risk of cardiovascular disease (CVD) and mortality [8]. The burgeoning epidemic of obesity and MetS in both developed and developing countries [9,10] is making NAFLD a prime concern of hepatologists worldwide.

Disease definitions

NAFLD is a diffuse parenchymal liver disease defined by both clinical (nonalcoholic) and histopathologic (steatosis, steatohepatitis, cirrhosis) criteria after serologic exclusion of other etiologies of liver diseases (e.g. viral, autoimmune, hemochromatosis, etc.) and secondary causes of hepatic steatosis (Table 32.1) [11,12]. Diagnosis of NAFLD requires a person to drink <20 g day^{-1} (about 3 drinks per day) [12].

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Table 32.1 Secondary causes of fatty liver disease

Nutritional	Protein-calorie malnutrition (PCM)†; Starvation†; Total parenteral nutrition (TPN) †; Rapid weight loss†; Gastrointestinal surgery for obesity†
Drugs	Glucocorticoids†; Synthetic estrogens†; Calcium-channel blockers†; Tamoxifen†; Methotrexate†; Amiodarone¶; Perhexiline maleate¶; Aspirin‡; Tetracycline‡; Valproate‡; Cocaine‡; Zidovudine†; Diadanosine‡; Fialuridine‡
Metabolic/Genetic	Lipodystrophy†; Dysbetalipoproteinemia†; Weber–Christian disease†; Wolman's disease¶; Acute fatty liver of pregnancy (AFLP) ‡
Toxins	*Amanita phylloides* mushroom†; Phosphorus‡; Petrochemicals†‡; *Bacillus cereus* toxins‡
Infections	Human immunodeficiency virus infection†; Hepatitis C (predominantly genotype 3) †; Small bowel diverticulosis with bacterial overgrowth†

†Cause macrovesicular steatosis (due to imbalance in the hepatic synthesis and export of lipids).
‡Cause microvesicular steatosis (due to defects in mitochondrial beta-oxidation).
¶Cause hepatic phospholipidosis (due to accumulation of phospholipids in lysosomes).

Steatosis, typically macrovesicular, defined by >5 % fat accumulation in the liver by weight and estimated in practice either as the percentage of fat-laden hepatocytes on light microscopy or by magnetic resonance spectroscopy (H^1-MRS), is the sine qua non of NAFLD in noncirrhotic livers [12–14]. Steatosis is necessary, but not sufficient, for the diagnosis of NASH. In adults, NASH is characterized by a combination of steatosis, hepatocyte ballooning, lobular inflammation, and perisinusoidal or "chicken-wire" fibrosis in a predominant zone 3 distribution [13]. In cryptogenic cirrhosis, all lesions of steatosis and NASH may be absent [13].

The cut-off value of nonpathologic steatosis of <5 % appears relatively arbitrary, based on chemical analysis of livers at autopsy and later supported by H^1-MRS quantification of the 95th percentile of intrahepatic triglyceride content (IHTG) in a multiethnic adult US population having no risk of hepatic steatosis [12–14]. However, another H^1-MRS study from the United States, found the 95th percentile of normal IHTG to be 3 % in lean, nondiabetic, Caucasians [15]. Moreover, obese nondiabetic individuals have an abnormal metabolic profile, characterized by impaired hepatic and systemic insulin-sensitivity and abnormal lipid kinetics, at even <5 % IHTG levels [16,17].

Clinical diagnosis

A liver biopsy is currently the "gold standard" for diagnosis of the full spectrum of NAFLD, despite its limitations concerning sampling variability, quantity of tissue obtained, surgical or nonsurgical specimen, experience of hepatopathologist, high interobserver variability in interpretations of lesions even among experts, and disagreement about minimal criteria for diagnosis of NASH among hepatopathologists [13]. Although numerous attempts have been made to develop noninvasive tests to predict NAFLD with fibrosis [18], biopsy remains the only reliable way to grade and stage NAFLD, including identification of precirrhotic fibrosis or remodeling of liver parenchyma which have a poorer prognosis [11–13]. Moreover, biopsy is the only way to exclude NASH as a cause of unexplained elevated liver enzymes in clinical hepatology practice [19].

Because of the cost, access, and potential complications associated with liver biopsy, it is not feasible to use it as a screening tool for NAFLD in the general population for epidemiologic studies [20,21]. Here clinicians have relied upon indirect, often less sensitive/specific, biochemical markers (e.g. liver enzymes like alanine aminotransferase [ALT]) or radiologic tests (e.g. ultrasound) to obtain prevalence and incidence data about NAFLD [21]. Serum ALT estimation is inexpensive, easily available, and a widely used screening test for a variety of liver diseases, and if ALT is used for the diagnosis of NAFLD, exclusion of other chronic liver disease by serology is important. Equally, NAFLD can be present with "normal" ALT [19,22,23]. In addition, the normal range of ALT values is controversial and the test is subjected to numerous physiologic, intraindividual and interlaboratory variability [24,25]. Although serum ALT has a statistically significant correlation with intrahepatic

Table 32.2 Comparison of radiological modalities for diagnosing hepatic steatosis

	US	CT (Unenhanced/ contrast-enhanced)	MRI	^{1}H-MRS
Readily available?	Yes	Yes	No	No
Analytic method difficulty	Simple	Simple	Complex	Complex
Entire liver assessed?	Yes	Yes	Yes	No
Qualitative assessment	Useful	Useful	Better than CT	Not useful
Quantitative assessment	Inaccurate	Useful for steatosis >30 %	Accurate for steatosis >8–10 % depending on protocol used	Very accurate
Differentiates fibrosis and iron overload	No	No	Yes	Yes
Radiation exposure?	No	Yes	No	No
Sensitivity				
>5 % steatosis +ve on biopsy	73.3 (62.2–82.1)	46.1 (22.2–71.8)	82.0 (63.7–92.2)	88.5 (76.6–94.7)
>10–20 % steatosis +ve on biopsy	90.5 (79.3–96.0)	57.0 (51.5–62.3)	90.0 (73.2–96.7)	82.6 (61.8–93.3)
>30 % steatosis +ve on biopsy	85.7 (78.4–90.8)	72.0 (59.7–81.7)	97.4 (83.5–99.6)	72.7 (41.4–91.0)
Specificity				
>5 % steatosis +ve on biopsy	84.4 (76.2–90.1)	93.5 (86.2–97.7)	89.9 (81.0–94.9)	92.0 (80.5–97.0)
>10–20 % steatosis +ve on biopsy	69.6 (60.0–77.7)	88.1 (81.1–92.7)	95.3 (83.2–98.8)	94.3 (79.8–98.6)
>30 % steatosis +ve on biopsy	85.2 (76.9–90.9)	94.6 (88.1–97.7)	76.1 (49.6–91.2)	95.7 (84.5–98.9)
Relative cost	Inexpensive	Moderately expensive	Expensive	Expensive
Factors determining accuracy	Operator, equipment, transducer, scanning parameters, hepatic/ renal parenchymal disease	Scanning parameters (kVP, mAs), acute liver injury, other infiltrative diseases of liver (e.g. iron, amyloid, etc.)	Imaging parameters, method of analysis	Acquisition parameters, method of analysis, location of regions of interest

fat content (IHTG) as assessed by the highly sensitive H^1-MRS, it alone is a poor surrogate marker of hepatic steatosis, as it only explains 15–19 % of variability of liver fat content, when controlled for age, sex, and BMI [26,27]. Rather, other components of MetS, like waist circumference, explain ~50 % of variability of liver fat content [26].

Radiologic tests such as ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI) are widely used to diagnose hepatic steatosis in clinical practice (Table 32.2) [28,29]. For epidemiologic studies, US is frequently used as it is simple, relatively inexpensive, portable, and has good accuracy to detect steatosis, especially if present in >30 % of hepatocytes, but suffers from poor interobserver and intraobserver agreement [30]. In addition, only MRI and H^1-MRS are capable of identifying fibrosis and NASH to some extent, but both have limited applicability as they are nonportable, expensive, and resource-intensive. To overcome these

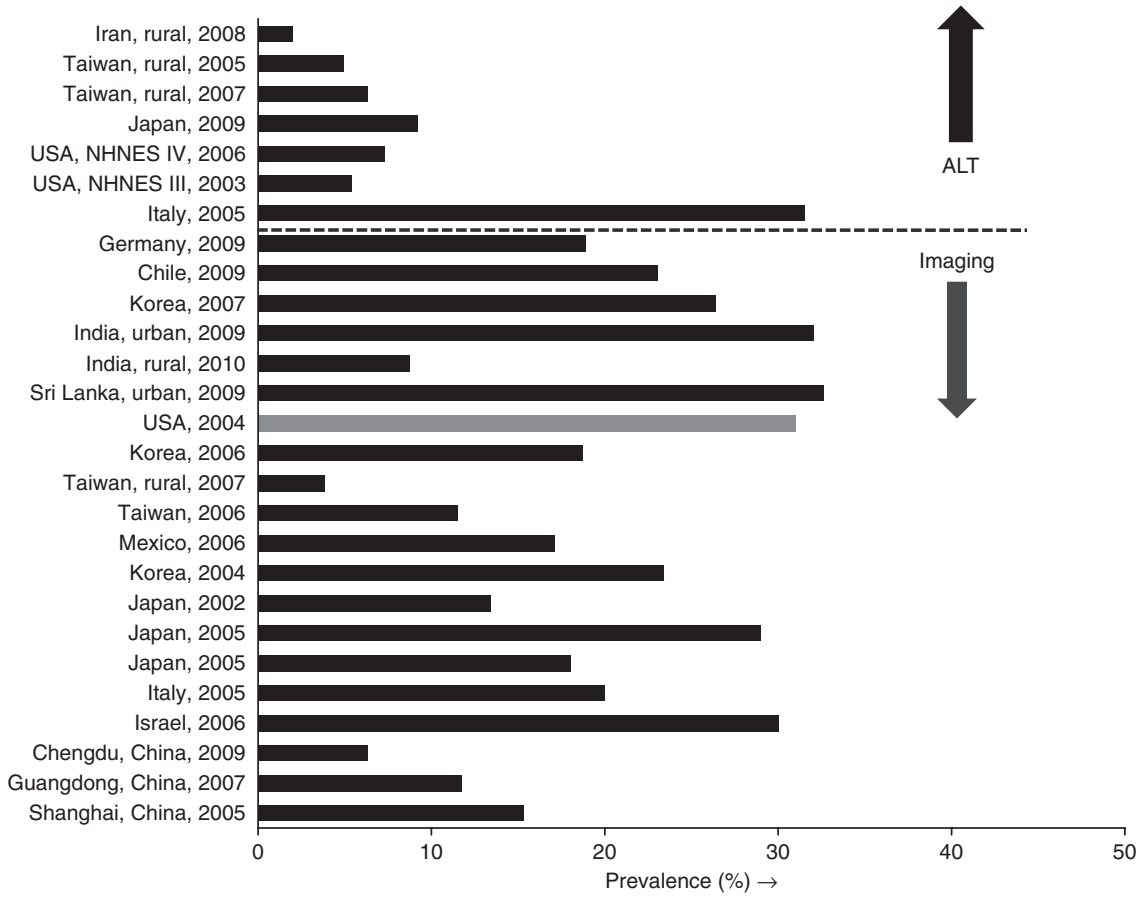


Figure 32.1 Prevalence of NAFLD in the general population of different countries based on ultrasonography (except USA which used H^1-MRS) and ALT levels. The studies from Korea and Japan are based on people attending health check-ups.

limitations, a number of noninvasive biomarkers, often combined to form a scoring system, and innovative technological application systems (e.g. elastography) are being investigated to predict hepatic steatosis and identify histologically advanced disease [18]. While "Fatty Liver Index" (FLI) has been used for detection of hepatic steatosis [31], transient elastography (TE), "NAFLD fibrosis score" (NFS) and "BARD score" have been proposed as methods for the detection of advanced fibrosis [18].

Prevalence

Considering the above limitations in performing epidemiologic studies of NAFLD, the prevalence of NAFLD using the most sensitive proton-MRS (H^1-MRS) is 33.6 % in a multiethnic urban Dallas cohort in the United States [32]. However, most studies have used US to detect steatosis, with prevalence ranging from 8.7 % in rural India [33] to 32.6 % in urban Sri Lanka [34]. Overall, the prevalence of NAFLD on imaging is 6–32 % in Asia (lower in rural compared to urban areas), 18–33 % in Europe–North America, 30 % in the Middle-East (Israel), and 23 % in South America (Chile) (Figure 32.1). The reason for this varied prevalence can be attributed to the varying prevalence of obesity in the respective background population (Figure 32.2). When ALT is used for screening, the prevalence rates are usually low (Figure 32.1). However, the true prevalence of the full spectrum of NAFLD, which requires histology, is

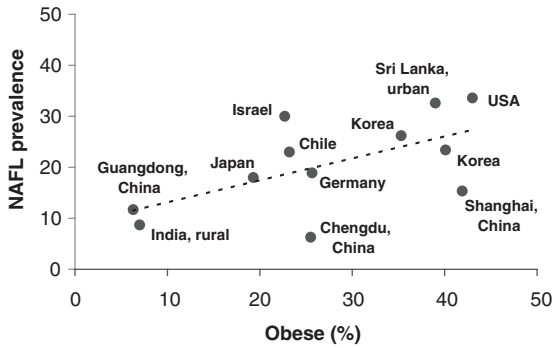


Figure 32.2 The prevalence of fatty liver (by ultrasound, except USA which used H^1-MRS) in the general population shows a linear relationship with respect to the prevalence of obesity (defined as BMI >25 kg m^{-2} and >30 kg m^{-2} in Asians and non-Asians, respectively) in the background population.

unknown. In one study, where almost all subjects with steatosis on US and persistently elevated ALT (prevalence 2.3 %) were subjected to liver biopsy, NASH and cryptogenic cirrhosis were present in 31 % and 11 % respectively [33]. A similar biopsy study from primary care in the United States (with a prevalence of steatosis of 46 %), revealed a 29.9 % prevalence of histologic NASH and 5.8 % prevalence of NASH with advanced fibrosis, in those with steatosis on US [35]. Moreover, in healthy prospective liver donors, the prevalence of NAFLD varies from 15.9 to 51.4 % and that of NASH from 1.9 to 15.5 % [36–39].

Incidence

True population-based estimates of the incidence of NAFLD are not available. The incidence of NAFLD, by US or ALT, has been studied in predominantly Asian populations in cohorts of employees, often exclusively males, undergoing regular health check-ups (Table 32.3). Incidence rates have ranged from 10 % to 14.6 % (31 to 71.4 per 1000 patient-years) over a follow-up ranging from 1.1 to 5 years. The baseline risk factors for development of NAFLD in these studies included male sex, having ALT in the higher quintiles of the reference range, obesity (both general and visceral), metabolic syndrome and/or insulin resistance. Interestingly, weight gain and persistent or rising insulin resistance over the follow-up period were strong independent predictors of development of NAFLD [40–47]. Chronologically, weight gain was followed by low HDL and increased ALT, ultimately culminating in glucose intolerance [40]. However, incidence data from non-Asians and/or mixed-gender groups are lacking.

Importantly, regression of NAFLD can also occur in 16 % on US and in 45 % by ALT after 1 year [40,48]. Weight reduction was the most important predictor of regression of NAFLD on US [40] or persistent normalization of elevated ALT [48].

Risk factors

These include both modifiable as well as nonmodifiable risk factors that are associated with NAFLD (Table 32.4).

Age and sex

Although NAFLD is found in all ages, population-based studies have demonstrated an increasing prevalence with rising age and a male preponderance in Asians and Caucasians [32,33,42,49,50]. Male sex is also an independent risk factor of new-onset NAFLD [40,42]. However, postmenopausal women have similar prevalence rates as men of the same age but higher than that of premenopausal women [49,50].

Ethnicity

Ethnicity is also a significant risk factor of NAFLD as demonstrated by the Dallas Heart Study [32]. Hispanics have a higher prevalence of hepatic steatosis (45 %) when compared with Caucasians (33 %) and African Americans (24 %). Interestingly, male sex was not associated with a higher prevalence of steatosis in Hispanics and African Americans. Hispanics also had a higher median IHTG content (4.6 %) compared with Caucasians (3.6 %) and African Americans (3.3 %). African Americans show lower histologic severity of lesions and probably have a lower prevalence of NASH and cryptogenic cirrhosis [51–54]. Others have demonstrated that among patients with newly diagnosed chronic liver disease due to presumed NAFLD, Hispanics and Asians are overrepresented and the Asian patients have a lower BMI and male preponderance [54]. In addition, increased prevalence of insulin resistance in Asian Americans is

Table 32.3 Incidence of NAFLD

	Country	Diagnosis	Male (%)	Mean age (yr; SD)	Mean BMI ($kg\ m^{-2}$; SD)	Duration of follow-up (years)	Incidence	Risk factors
Suzuki, 2005[40]	Japan ($n = 529$)	ALT	73.2	35 (± 8)	22 (± 3)	5	31/1000 patient-years (13.4 %)	Age < 40y: Male; ↓ Age; ↑ BMI; Hypertension; ↓ HDL; Age >40y; Glucose intolerance
Lee, 2001[41]	Korea ($n = 6846$)	ALT	100	38.3	22.6	4	14.1 %	↑ BMI at baseline; ↑ Weight gain on follow-up
Hamaguchi, 2005[42]	Japan ($n = 3147$)	US	53.8	47.6 (± 8.8)	22.6 (± 3.0)	1.1	10 %	Male sex; Weight gain on follow-up; Metabolic syndrome (ATP III) at baseline
Chang, 2009[43,44]	Korea ($n = 4246$)	US	100	36.7 (± 4.8)	22.6 (± 2.4)	4	74.1/1000 patient-years (14.6 %)	↑ ALT (even within reference range); ↑ Baseline BMI; Weight gain
Rhee, 2011[45,46]	Korea ($n = 4954$)	US	50.5	40.0 (± 5.9)	22.4 (± 2.5)	5	13.0 %	↑ Baseline and persistently ↑ Fasting insulin level; ↑ Baseline BMI; Uric acid; Male sex; Weight gain
Xu, 2010[47]	China ($n = 6890$)	US	65.2	44.4 (± 12.7)	22.4 (± 2.7)	3	11.8 %	↑ Baseline BMI; Waist circumference; Diastolic BP; ALT (within normal range); Uric acid; Triglyceride; Creatinine ↓ baseline HDL

Table 32.4 Risk factors for NAFLD

Nonmodifiable	Modifiable
Age	Obesity
	(BMI $\geq$25 kg m^{-2} in Asians and $\geq$30 kg m^{-2} in non-Asians)
Sex	Metabolic syndrome (MetS): International Diabetes Federation definition, 2005
	Central obesity: Waist circumference (ethnicity specific; e.g. Europids: Men $\geq$94 cm; Women $\geq$80 cm South Asians: Men $\geq$90 cm; Women $\geq$80 cm)
	Plus any two:
	Raised triglycerides: >150 mg dL^{-1} (1.7 mmol L^{-1}) *or* Specific treatment for this.
	Reduced HDL cholesterol: <40 mg dL^{-1} (1.03 mmol L^{-1}) in men; <50 mg dL^{-1} (1.29 mmol L^{-1}) in women *or* Specific treatment for this.
	Raised blood pressure: systolic $\geq$130 mm Hg; diastolic $\geq$85 mm Hg; *or* On anti-hypertensive Rx.
	Raised fasting plasma glucose: Fasting plasma glucose $\geq$100 mg dL^{-1} (5.6 mmol L^{-1}) *or* Previously diagnosed type 2 diabetes.
Ethnicity	Hyperuricemia
Genetic	Diet
	Physical activity

associated with a twofold greater IHTG content than Caucasians [15].

Genetic susceptibility

A genome-wide association analysis in 3383 African American, Caucasian, and Hispanic participants of the Dallas Heart Study identified a single-nucleotide polymorphism, G allele, encoding isoleucine substitution for methionine (I148M) in a gene designated as patatin-like phospholipase A3 (*PNPLA3*), also called adiponutrin-3 [55]. Hepatic fat content as measured by MRS was twofold higher in *PNPLA3*-148M homozygotes than in noncarriers and the association remained highly significant after adjusting for BMI and diabetes status. Overall, variation in *PNPLA3* contributed to ethnic and inter-individual differences in hepatic fat content and susceptibility to NAFLD. This *PNPLA3* polymorphism is also associated with steatosis severity and advanced histologic changes in pediatric NAFLD [56]. The adiponutrins seem to play cooperative roles in both lipolysis and its opposite process of triglyceride synthesis [57,58].

Variants in the gene coding apolipoprotein C3 (*APOC3*) are associated with hypertriglyceridemia [59] as well as NAFLD in Asian and non-Asian men [60]. While variants of *APOC3* and *PNPLA3* account for 11 % and 6.5 % of the variance in the risk of NAFLD respectively, the combination of both accounts for 13 % of the risk suggesting that these genetic factors may have an interactive effect [60].

Metabolic syndrome

Although when initially described, NAFLD was regarded as a disease occurring in "*obese, middle-aged women with asymptomatic hepatomegaly who are diabetic or hyperlipidemic*" [61], it is now recognized that NAFLD can occur in all ages and both sexes as well as in the nonobese (defined by BMI) [33,62–64]. Nonetheless, obesity, not alcohol, is the commonest cause of fatty liver in the community, with the prevalence of NAFLD rising. However, alcohol is a separate risk factor and with increasing BMI, the highest prevalence of NAFLD is observed in obese drinkers [65–67]. Weight gain is one of the most important risk factors in the development of NAFLD [41–46]. Still, not all obese individuals have hepatic steatosis [68] and not all morbidly obese persons have histologically severe NAFLD [69]. Rather, cross-sectional studies have revealed that the risk and severity of NAFLD correlates more strongly with central obesity (defined by sex and ethnicity-specific waist circumference) than BMI [70,71]. A systematic review of clinical and epidemiologic studies also found an independent association between central obesity and hepatic steatosis; the association reflects a strong link with visceral (intra-abdominal) rather than

subcutaneous abdominal fat [72]. This relationship has been found to hold true across diverse ethnicities [73,74].

Moreover, the prevalence and severity of NAFLD increases in parallel to progressive degrees of abnormal glucose tolerance [66,75]. About 40–70 % of type 2 diabetics have NAFLD and they have 80 % greater liver fat than age, sex, and weight-matched nondiabetic subjects on the H^1-MRS [69–76]. Diabetics with NAFLD are more insulin-resistant and have a dysmetabolic phenotype with increased visceral adiposity [77,78]. In nonobese and nondiabetic subjects, visceral obesity, dyslipidemia, and insulin resistance were found to be independent predictors of NAFLD [64]. Biopsy-proven NAFLD is associated with insulin resistance even in nonobese and nondiabetic subjects, with impaired insulin actions on both glucose and lipid metabolism [79,80]. Increasing hepatic steatosis in turn is associated with increasing degrees of hepatic insulin resistance [16,77].

Visceral obesity and insulin resistance are regarded to be the pathophysiologic bedrock of the constellation of anthropometric and metabolic variables grouped under the term of "metabolic syndrome" (MetS) [81–84]. It has also been demonstrated in several hospital and community-based studies that the prevalence, incidence, and severity of NAFLD, strongly correlate with MetS, independent of obesity [7,66,67,85–90]. Moreover, NAFLD is a more sensitive marker of MetS than its individual components alone [91]. Individuals with MetS, independent of age, gender, and BMI, have increased hepatic steatosis and both hepatic and visceral/intra-abdominal steatosis strongly correlate with the cardinal components of syndrome, including hepatic insulin resistance, in these individuals [26,27].

Thus, hepatic steatosis appears to be a sensitive marker of visceral adiposity and strongly correlates with hepatic and peripheral insulin resistance, independent of age, sex, and BMI. Hepatic steatosis results in decreased insulin-mediated suppression of hepatic gluconeogenesis and/or hepatic glucose output with unabated lipolysis from peripheral adipose tissues and increased serum free fatty acids (FFA). The increased FFA flux into the liver promotes further steatosis despite increased hepatic lipid oxidation and efflux by VLDL-synthesis. Although it is increasingly clear that peripheral IR, hepatic steatosis, and hepatic IR are linked in a self-perpetuating vicious cycle, which of these is the primary defect has not currently been established [92].

In addition, hyperuricemia and gout have been regarded as components of metabolic syndrome since the 1920s [8]. Cross-sectional studies have shown that hyperuricemia is an independent risk factor for NAFLD after correcting for obesity and MetS [93]. Moreover, the incidence of future NAFLD is higher with increasing levels of serum uric acid at baseline [46,47].

Diet and physical activity

Economic prosperity correlates with increasing prevalence of NAFLD [33,94]. Industrialization throughout the world has resulted in a shift in dietary and physical activity pattern. Globally, our diet has become more energy-dense and sweet, with higher consumption of meat, partially hydrogenated fat, and sugar-sweetened beverages (SSBs) with lower intake of fiber [95]. Simultaneously, activity patterns at work, home, leisure, and travel are shifting towards lower energy expenditure [95,96]. These changes are the driving force of the global obesity pandemic [9,10]. A population-based study of NAFLD from Israel, revealed a higher intake of meat and sweetened drinks and lower intake of food containing ω-3 fatty acids (n-3 polyunsaturated fatty acids (PUFAs)) in NAFLD subjects versus those without NAFLD on US [97]. A higher intake of meat and sweet drinks was associated with an increased risk of NAFLD after adjustment for age, gender, BMI, and total calorie intake [97]. A population-based study from Japan revealed that higher intake of n-3 PUFAs was associated with a lower prevalence of US-diagnosed NAFLD, but the effect was specific to men and not women [98]. Higher intake of sweet drinks was again independently associated with more severe fibrosis in NAFLD, after controlling for age, gender, BMI, and total calorie intake [99]. In an Italian study, nondiabetic, nonobese, and normolipidemic NASH patients again reported lower PUFA intake which was associated with increased insulin resistance and postprandial hypertriglyceridemia [100]. Moreover, the same Israeli study demonstrated that the risk of NAFLD is lower in those who had a higher leisure-time physical activity [101]. Among the dietary risk factors for NAFLD, alcohol consumption is regarded as exclusionary for diagnosis. However, recent data is emerging that modest alcohol consumption, especially

wine, may be associated with a reduced prevalence of NAFLD, when diagnosed by elevated ALT [102,103].

Natural history and mortality

The important issues in the natural history of NAFLD are the rate of progression from simple steatosis to more advanced NASH with fibrosis and/or cirrhosis, and the clinical outcome of subjects among the varied histologic spectrum of NAFLD.

Natural history studies based on liver histology are subject to publication and referral biases as these depend on liver biopsies performed in specialist centers. Cross-sectional studies have shown age (especially >50 years), BMI (>28 to 32 kg m^{-2}), insulin resistance/diabetes mellitus and abnormal aminotransferases are associated with histologically advanced stages of NAFLD [69,104]. A systematic review of longitudinal studies with paired biopsy samples [104], showed that in those cohorts ($n = 221$) consisting of predominantly female (63.8 %), obese (median BMI 31.8 kg m^{-2}), insulin-resistant (79 % frankly diabetic) people, and in the fifth decade of life (mean age, 47.4 years) on a mean follow-up period of 5.3 (SD: 4.2) years, 37.6 % progressed to a higher fibrosis stage, 41.6 % had no change, and 20.8 % showed improvement in fibrosis. One third (31.7 %) of the cohort had stage 3 fibrosis and/or cirrhosis in the final biopsy. Only older age and presence of any inflammation on index biopsy, not obesity or diabetes or metabolic syndrome, were independent predictors of histologic progression, including cirrhosis [104]. Those with inflammation (either lobular or periportal) on index biopsy had 2.5 times higher risk of progression and progressed more rapidly (median 4.2 vs. 13.4 years) compared with those without inflammation (i.e. steatosis with fibrosis only). Severity of steatosis decreased over follow-up in these cohorts [104].

Despite paired biopsy studies showing that about two-thirds of patients with NAFLD have nonprogressive disease, population and community-based cohort studies of NAFLD (defined by biochemical, US or histologic criteria) having a longer follow-up (median: 7.3 to 24 years), have shown that the pooled overall mortality in NAFLD is 57 % higher than that of the general population (OR 1.57; 95 % CI 1.18–2.10), with the burden of this excess mortality being borne by those with histologic NASH or predominantly concentrated in those aged 45–54 years (having adjusted standardized mortality ratios [SMR] of 4.40 and 8.15, for all-cause and cardiovascular mortality, respectively) [105,106]. In these cohorts of individuals with NAFLD, however, liver disease is usually the third leading cause of death (6–19 % of all deaths) after ischemic heart disease (25–30 % of all deaths) and extrahepatic malignancies (24–28 % of all deaths)[105,107–109].

When the cohorts of biopsy-proven NAFLD are analyzed, the survival of those with steatosis alone is similar to that of the general population [105], and they have lower overall mortality compared to those with NASH (Figure 32.3) [110]. During a median follow-up period ranging 7.6–24 years, NASH patients have a significantly higher risk of developing cirrhosis and increased liver-related mortality (11–17.5 %) as compared to those with steatosis (1.7–2.7 %) alone (Figure 32.3) [105,110]. In addition, the more advanced the stage of fibrosis in NASH, the higher the liver-related mortality [105].

NAFLD as a risk factor for cardiovascular disease

Obesity and MetS are primary risk factors for cardiovascular (CVS) mortality, the leading cause of deaths globally [111]. However, not all obese people are metabolically unhealthy and conversely not all normal-weight people are metabolically healthy [112, 113]. Interestingly, obese and metabolically healthy individuals, characterized by having remarkably high insulin sensitivity, no hypertension, and normal lipid, inflammation, and hormonal profiles (low triglycerides and C-reactive protein concentrations and high HDL cholesterol and adiponectin concentrations) [112], have a lower hepatic steatosis and visceral adiposity compared with other obese individuals [114] indicating that NAFLD is a critical component of MetS. Hence, it is unsurprising that studies have investigated the contribution of NAFLD as an independent risk factor associated with the development of type 2 diabetes as well as cardiovascular disease.

NAFLD is associated with endothelial dysfunction [115] and increased carotid intima-medial thickness (CIMT), markers of early atherosclerosis, in most, but not all, cross-sectional studies [116–118]. A meta-analysis revealed that individuals with US-diagnosed hepatic steatosis have a 13 % higher prevalence of CIMT than those without steatosis [117]. In

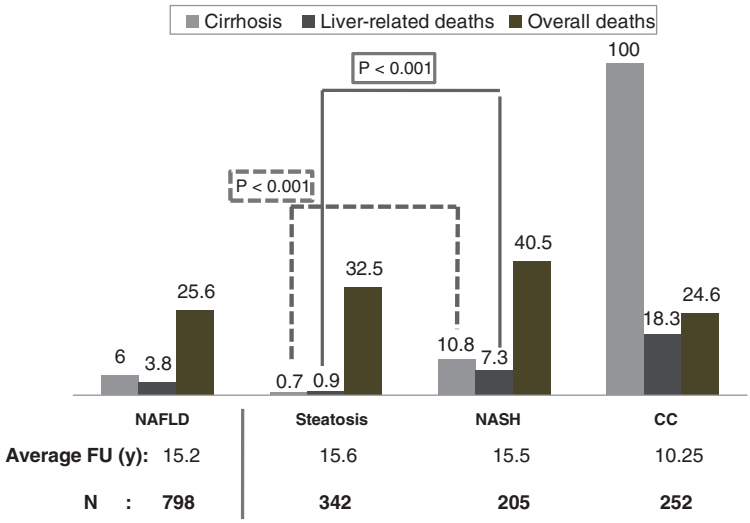


Figure 32.3 Prognosis of NAFLD: community-based studies where NAFLD was confirmed by imaging and/or liver biopsy and with >5 years of follow-up. The size of the cohorts/subgroups and period of follow-up are presented in the horizontal axis; the vertical axis shows percentage of patients with a particular endpoint.

prospective population-based cohort studies an increase of 1 uL^{-1} of natural logged serum γ-glutamyltransferase (GGT), one of the liver enzymes, is associated with a 20 % increase in the risk of coronary heart disease (CHD), a 54 % increase in the risk of stroke, and a 34 % increase in the risk of combined CHD/stroke over a follow-up duration ranging from 3.3–15.7 years [119]. In a recent meta-analysis, GGT, even within normal laboratory range, was associated with future CVS events and mortality, even after adjusting for MetS [114]. Interestingly, US-diagnosed hepatic steatosis improves the predictive power of GGT for incident CVS events, even after adjusting for baseline cardiometabolic factors [120]. Moreover, the pooled odds ratio, in eight community-based cohorts of US- or histologically diagnosed NAFLD, for incident and fatal CVS events was 2.05 (95 % CI 1.81–2.31) and 2.16 (95 % CI 1.88–2.49), respectively [105]. Consistent with this, CVS mortality is higher in those with cryptogenic cirrhosis when compared with HCV-related cirrhosis [121]. US-detected hepatic steatosis is an independent risk factor for CVS disease even in type 1 diabetics, a group of individuals who classically do not have insulin resistance [122].

However, when ALT is used as a biomarker of NAFLD, no such increased risk of CVS events (fatal or nonfatal) was discernible with increasing ALT [105,119,123]. Although it can be argued that patients with NAFLD can be considered as targets for lifestyle modifications to reduce CVS risks [124], the question that is as yet unanswered is whether adding NAFLD to the already standard risk scores, for example Framingham Risk Score, improves the prediction of future CVS events [125].

NAFLD as a risk factor for type 2 diabetes

Fatty liver (diagnosed on US) leads to a two- to fourfold increased risk of incident diabetes even after adjusting for confounding variables [126–128]. However, the risk may be greater in those who have impaired fasting glucose at baseline [126]. On the other hand, in contrast to dichotomy between the effect of GGT and ALT for incident CVS outcomes, both of these liver enzymes are strongly associated with risk of incident type 2 diabetes in multiple, prospective, population-based studies [105,128]. A meta-analysis of all these studies revealed, after adjusting for confounders, a 1 uL^{-1} increase in natural logged ALT and GGT was associated with an 85 % and 92 %, respectively, increase in diabetes risk [128]. The multiple-adjusted risk of incident diabetes between the highest and lowest quintiles of ALT and GGT were 192 % and 271 %, respectively [105]. Although GGT appears to be a more sensitive biomarker, ALT is more liver-specific. ALT shows a strong association with insulin resistance but less so with hepatic steatosis, in both nondiabetic and diabetic men and women [26,27,76,129]. On the other hand GGT, but not ALT, strongly correlated with IR in nondiabetic men but not nondiabetic women [130]. Overall, the presence of NAFLD assessed

biochemically or by imaging, carries an increased risk of future diabetes.

Conclusions

Prevalence of NAFLD has risen throughout the world with urbanization, economic prosperity, and increasing obesity; it is now the commonest chronic liver disease in Western countries. Pathology of NAFLD ranges from isolated steatosis to NASH with a potential to progress to cirrhosis and its complication, hepatocellular carcinoma. Although a minority develops symptomatic liver disease with consequent mortality, the high prevalence of NAFLD accounts for the large burden on health services. There is a strong association between NAFLD and individual components of MetS and hepatic steatosis is a critical step in the evolution of insulin resistance. Therefore, NAFLD is a risk factor for the development of type 2 diabetes; it is associated with an increased risk of CVS events and related mortality. The importance of NAFLD is due to its impact upon liver-related outcomes as much as on all-cause mortality.

Multiple choice questions

1 Which is the most specific radiologic noninvasive modality to detect hepatic steatosis?

A Ultrasound
B Computed tomography (CT)
C MRI
D H^1-MRS
E Transient elastography

2 The histologic variable that is the highest risk factor for development of cirrhosis in an index biopsy from an individual with suspected NAFLD is?

A Steatosis alone
B Steatosis with inflammation
C Steatosis with fibrosis only
D Fibrosis alone
E Portal inflammation

3 Which of the liver enzymes is the most sensitive predictor of future development of diabetes and/or cardiovascular mortality?

A AST
B ALT
C GGT
D ALP
E AST:ALT ratio

References

1 de Alwis NMW, Day CP. Non-alcoholic fatty liver disease: The mist gradually clears. *J Hepatol* 2008;48:S104–12.
2 Bugianesi E, Leone N, Vanni E, et al. Expanding the natural history of nonalcoholic steatohepatitis: from cryptogenic cirrhosis to hepatocellular carcinoma. *Gastroenterology* 2002;123:134–40.
3 Clark JM, Diehl AM. Nonalcoholic fatty liver disease: an under-recognized cause of cryptogenic cirrhosis. *JAMA* 2003;289:3000–4.
4 Ludwig J, McGill DB, Lindor KD. Metabolic liver disease. Review: nonalcoholic steatohepatitis. *J Gastroenterol Hepatol* 1997;12:398–403.
5 Ludwig J, Viggiano TR, McGill DB, Oh BJ. Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. *Mayo Clin Proc* 1980;55:434–8.
6 Powell EE, Cooksley WG, Hanson R, et al. The natural history of nonalcoholic steatohepatitis: a follow-up study of forty-two patients for up to 21 years. *Hepatology* 1990;11:74–80.
7 Marchesini G, Brizi M, Bianchi G, et al. Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. *Diabetes* 2001;50:1844–50.
8 Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. *Lancet* 2005;365:1415–28.
9 Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999–2008. *JAMA* 2010;303:235–41.
10 Misra A, Khurana L. Obesity and the metabolic syndrome in developing countries. *J Clin Endocrinol Metab* 2008;93:S9–S30.
11 Angulo P. Nonalcoholic fatty liver disease. *New Engl J Med* 2002;346:11221–31.
12 Neuschwander-Teri BA, Caldwell SH. Nonalcoholic steatohepatitis: summary of an AASLD single topic conference. *Hepatology* 2003;37:1202–19.
13 Brunt EM. Histopathology of non-alcoholic fatty liver disease. *Clin Liver Dis* 2009;13:533–44.
14 Szczepaniak LS, Nurenberg P, Leonard D, et al. Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population. *Am J Physiol Endocrinol Metab* 2005;288:E462–8.
15 Petersen KF, Dufour S, Feng J, et al. Increased prevalence of insulin resistance and non-alcoholic fatty liver disease in Asian-Indian men. *Proc Natl Acad Sci* 2006;103:18273–7.
16 Korenblat KM, Fabbrini E, Mohammed BS, Klein S. Liver, muscle, and adipose tissue insulin action is directly related to intrahepatic triglyceride content

in obese subjects. *Gastroenterology* 2008;134:1369–75.
17 Fabbrini E, Mohammed BS, Magkos F, et al. Alterations in adipose tissue and hepatic lipid kinetics in obese men and women with nonalcoholic fatty liver disease. *Gastroenterology* 2008;134:424–31.
18 Dowman JK, Tomlinson JW, Newsome PN. Systematic review: the diagnosis and staging of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis. *Aliment Pharmacol Ther* 2011;33:525–40.
19 de Lédinghen V, Ratziu V, Causse X, et al. Diagnostic and predictive factors of significant liver fibrosis and minimal lesions in patients with unexplained elevated transaminases. A prospective multicenter study. *J Hepatol* 2006;45:592–9.
20 Bravo AA, Seth SG, Chopra S. Liver biopsy. *New Engl J Med* 2001;344:495–500.
21 Clark JM, Diehl AM. Defining nonalcoholic fatty liver disease: implications for epidemiologic studies. *Gastroenterology* 2003;124:498–500.
22 Mofrad P, Contos MJ, Haque M, et al. Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values. *Hepatology* 2003;37:1286–92.
23 Cave M, Appana S, Patel M, et al. Polychlorinated biphenyls, lead, and mercury are associated with liver disease in American adults: NHANES 2003–2004. *Environ Health Perspect* 2010;118:1735–42.
24 Kaplan MM. Alanine aminotransferase levels: what's normal? *Ann Intern Med* 2002;137:49–51.
25 Das K. "Normal" alanine aminotransferase and Christopher Boorse. *Hepatology* 2010;52:1173.
26 Kotronen A, Yki-Järvinen H, Sevastianova K, et al. Comparison of the relative contributions of intra-abdominal and liver fat to components of the metabolic syndrome. *Obesity* 2011;19:23–8.
27 Kotronen A, Westerbacka J, Bergholm R, et al. Liver fat in the metabolic syndrome. *J Clin Endocrinol Metab* 2007;92:3490–7.
28 Ma X, Holalkere NS, Kambadakone R, et al. Imaging-based quantification of hepatic fat: methods and clinical applications. *RadioGraphics* 2009;29:1253–80.
29 Bohte AE, vanWerven JR, Bipat S, Stoker J. The diagnostic accuracy of US, CT, MRI, and ^{1}H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: a meta-analysis. *Eur Radiol* 2011;21:87–97.
30 Strauss S, Gavisch E, Gottlieb P, Katsnelson L. Interobserver and intraobserver variability in the sonographic assessment of fatty liver. *Am J Radiol* 2007;189:W320–3.
31 Bedogni G, Bellentani S, Miglioli L, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol*;2006(6):33.
32 Browning JD, Szczepanaik LS, Dobbins R, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 2004;40:1387–95.
33 Das K, Das K, Mukherjee PS, et al. Nonobese population in a developing country has a high prevalence of nonalcoholic fatty liver and significant liver disease. *Hepatology* 2010;51:1593–1602.
34 Dassanayeke AS, Kasturirante A, Rajindrajith S, et al. Prevalence and risk factors for non-alcoholic fatty liver disease among adults in an urban Sri Lankan population. *J Gastroenterol Hepatol* 2009;24:1284–8.
35 Williams CD, Stengel J, Asike MI, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. *Gastroenterol* 2011;140:124–31.
36 Lee JY, Kim KM, Lee SG, et al. Prevalence and risk factors of non-alcoholic fatty liver disease in potential living liver donors in Korea: a review of 589 consecutive liver biopsies in a single center. *J Hepatol* 2007;47:239–44.
37 Ryan CK, Johnson LA, Germin BI, Marcos A. One hundred consecutive hepatic biopsies in the workup of living donors for right lobe liver transplantation. *Liver Transpl* 2002;8:1114–22.
38 Savas N, Coskun M, Bilezikci B, et al. Value of an individual liver biopsy in the preoperative evaluation of apparently healthy potential liver donors. *Liver Transpl* 2008;14:541–6.
39 Minervini MI, Ruppert K, Fontes P, et al. Liver biopsy findings from healthy potential living liver donors: reasons for disqualification, silent diseases and correlation with liver injury tests. *J Hepatol* 2009;50:501–10.
40 Suzuki A, Angulo P, Lymp J, et al. Chronological development of elevated aminotransferases in a nonalcoholic population. *Hepatology* 2005;41:64–71.
41 Lee DH, Ha MH, Christiani DC. Body weight, alcohol consumption and liver enzyme activity: a 4-year follow-up study. *Int J Epidemiol* 2001;30:766–70.
42 Hamaguchi M, Kojima T, Takeda N, et al. The metabolic syndrome as a predictor of non-alcoholic fatty liver disease. *Ann Intern Med* 2005;143:722–8.
43 Chang Y, Ryu S, Sung E, Jang Y. Higher concentrations of alanine aminotransferase within the reference interval predict nonalcoholic fatty liver disease. *Clin Chem* 2007;53:686–92.
44 Chang Y, Ryu S, Sung E, et al. Weight gain within the normal weight range predicts ultrasonographically

detected fatty liver in healthy Korean men. *Gut* 2009; 58:1419–25.
45 Rhee EJ, Lee WY, Cho YK, et al. Hyperinsulinemia and the development of nonalcoholic fatty liver disease in nondiabetic adults. *Am J Med* 2011;124:69–76.
46 Lee JW, Cho YK, Ryan MC, et al. Serum uric acid as a predictor for the development of nonalcoholic fatty liver disease in apparently healthy subjects: a 5-year retrospective cohort study. *Gut Liver* 2010;4:378–83.
47 Xu C, Yu C, Xu L, et al. High serum uric acid increases the risk for nonalcoholic fatty liver disease: a prospective observational study. *PLoS One* 2010; doi:10.1371/journal.pone.0011578
48 Suzuki A, Lindor K, St. Sauver J, et al. Effect of changes on body weight and lifestyle in non-alcoholic fatty liver disease. *J Hepatol* 2005;43:1060–66.
49 Fan JG, Zhu J, Li XJ, et al. Prevalence of and risk factors for fatty liver in a general population of Shanghai, China. *J Hepatol* 2005;43:508–14.
50 Zhou YJ, Li YY, Nie YQ, et al. Prevalence of fatty liver disease and its risk factors in the population of South China. *World J Gastroenterol* 2007;47:6419–24.
51 Kallwitz ER, Guzman G, TenCate V, et al. The histologic spectrum of liver disease in African-American, Non-Hispanic White, and Hispanic obesity surgery patients. *Am J Gastroenterol* 2009;104:64–9.
52 Mohanty SR, Mohanty SR, Troy TN, et al. Influence of ethnicity on histological differences in non-alcoholic fatty liver disease. *J Hepatology* 2009;50:797–804.
53 Browning JD, Kumar KS, Saboorian MH, Thiele DL. Ethnic differences in the prevalence of cryptogenic cirrhosis. *Am J Gastroenterol* 2004;99:292–8.
54 Weston SR, Leyden W, Murphy R, et al. Racial and ethnic distribution of nonalcoholic fatty liver in persons with newly diagnosed chronic liver disease. *Hepatology* 2005;41:372–9.
55 Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers susceptibility to non-alcoholic fatty liver disease. *Nat Genet* 2008;40:1461–5.
56 Luca Valenti, Anna Alisi, Enrico Galmozzi, et al. I148M Patatin-like phospholipase domain-containing 3 gene variant and severity of pediatric non-alcoholic fatty liver disease. *Hepatology* 2010;52:1274–80.
57 Farrell GC. PNPLeAse get the fats right: does lipogenesis or lipolysis cause NASH? *Hepatology* 2010;52:818–21.
58 Browning JD, Cohen JC, Hobbs HH. Patatin-like phospholipase domain-containing 3 and the pathogenesis and progression of pediatric nonalcoholic fatty liver disease. *Hepatology* 2010;52:1189–92.
59 Guettier JM, Georgopoulos A, Tsai MY, et al. Polymorphisms in the fatty acid-binding protein 2 and apolipoprotein C-III genes are associated with the metabolic syndrome and dyslipidemia in a South Indian population. *J Clin Endocrinol Metab* 2005;90: 1705–11.
60 Petersen KF, Dufour S, Hariri A, et al. Apolipoprotein C3 gene variants in nonalcoholic fatty liver disease. *New Engl J Med* 2010;362:1082–9.
61 Sheth SG, Gordon FD, Chopra S. Nonalcoholic steatohepatitis. *Ann Intern Med* 1997;126:137–45.
62 Bacon BR, Farahvash MJ, Janney CG, Neuschwander-Teri BA. Nonalcoholic steatohepatitis: an expanded clinical entity. *Gastroenterol* 1994;107:1103–9.
63 Balridge AD, Perez-Atayde AR, Graeme-Cook F, et al. Idiopathic steatohepatitis in childhood: a multicenter retrospective study. *J Pediatr* 1995;127:700–4.
64 Kim HJ, Kim HJ, Lee KE, et al. Metabolic significance of nonalcoholic fatty liver disease in nonobese, nondiabetic adults. *Arch Intern Med* 2004;164:2169–75.
65 Bellentani S, Saccoccio G, Masutti F, et al. Prevalence of and risk factors for hepatic steatosis in Northern Italy. *Ann Intern Med* 2000;132:112–7.
66 Fan J-G, Saibara T, Chitturi S, et al. and the Asia-Pacific Working Party for NAFLD. What are the risk factors and settings for non-alcoholic fatty liver disease in Asia-Pacific? *J Gastroenterol Hepatol* 2007;22:794–800.
67 Okanoue T, Umemura A, Yasui K, Itoh Y. Nonalcoholic fatty liver disease and non-alcoholic steatohepatitis in Japan. *J Gastroenterol Hepatol* 2011;26(S1): 153–62.
68 Vega GL, Chandalia M, Szczepaniak LS, Grundy SM. Metabolic correlates of nonalcoholic fatty liver in women and men. *Hepatol* 2007;46:716–22.
69 Argo CK, Caldwell SH. Epidemiology and natural history of non-alcoholic steatohepatitis. *Clin Liv Dis* 2009;13:511–31.
70 Hamaguchi M, Kojima T, Itoh Y, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. *Am J Gastroenterol* 2007;102:2708–15.
71 van der Poorten D, Milner K-L, Hui J, et al. Visceral fat: a key mediator of steatohepatitis in metabolic liver disease. *Hepatol* 2008;48:449–57.
72 Jakobsen MU, Berentzen T, Sørensen TIA, Overvad K. Abdominal obesity and fatty liver. *Epidemiol Rev* 2007;29:77–87.
73 Park BJ, Kim YJ, Kim DH, et al. Visceral adipose tissue area is an independent risk factor for hepatic steatosis. *J Gastroenterol Hepatol* 2008;23:900–7.
74 Guerrero R, Vega GL, Grundy SM, Browning JD. Ethnic differences in hepatic steatosis: an insulin resistance paradox? *Hepatol* 2009;49:791–801.
75 Mohan V, Farooq S, Deepa M, et al. Prevalence of non-alcoholic fatty liver disease in urban south Indians

in relation to different grades of glucose intolerance and metabolic syndrome. *Diab Res Clin Pract* 2009;84: 84–91.
76 Kotronen A, Juurinen L, Hakkarainen A, et al. Liver fat is increased in type 2 diabetic patients and underestimated by serum alanine aminotransferase compared with equally obese nondiabetic subjects. *Diab Care* 2008;31:165–9.
77 Kelley DE, McKolanis TM, Hegazi RAF, et al. Fatty liver in type 2 diabetes mellitus: relation to regional adiposity, fatty acids, and insulin resistance. *Am J Physiol Endocrinol Metab* 2003;285:E906–16.
78 Gastaldelli A, Cusi K, Pettiti M, et al. Relationship between hepatic/visceral fat and hepatic insulin resistance in nondiabetic and type 2 diabetic subjects. *Gastroenterol* 2007;133:496–506.
79 Sanyal AJ, Campbell-Sargent C, Mirshahi F, et al. Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities. *Gastroenterol* 2001;120:1183–92.
80 Bugianesi E, Gastaldelli A, Vanni E, et al. Insulin resistance in non-diabetic patients with non-alcoholic fatty liver disease: sites and mechanisms. *Diabetologia* 2005;48:634–42.
81 Reaven GM. The individual components of the metabolic syndrome: is there a raison d'être? *J Am Coll Nutr* 2007;3:191–5.
82 Bruce KD, Byrne CD. The metabolic syndrome: common origins of a multifactorial disorder. *Postgrad Med J* 2009:614–21.
83 Gale EAM. The myth of the metabolic syndrome. *Diabetologia* 2005;48:1679–83.
84 Kahn R, Buse J, Ferrannini E, Stern M. The metabolic syndrome: time for a critical appraisal. *Diabetes Care* 2005;28:2289–304.
85 Marchesini G, Bugianesi E, Forlani G, et al. Nonalcoholic fatty liver, steatohepatitis, and the metabolic syndrome. *Hepatol* 2003;37:917–23.
86 Ryan MC, Wilson AM, Slavin J, et al. Associations between liver histology and severity of the metabolic syndrome in subjects with nonalcoholic fatty liver disease. *Diabetes Care* 2005;28:1222–4.
87 Olynyk JK, Knuiman MW, Divitini ML, et al. Serum alanine aminotransferase, metabolic syndrome, and cardiovascular disease in an Australian population. *Am J Gastroenterol* 2009;104:1715–22.
88 Saito T, Nishise Y, Makino N, et al. Impact of metabolic syndrome on elevated serum alanine aminotransferase levels in the Japanese population. *Metab Clin Experiment* 2009;58:1067–75.
89 Suh S-Y, Choi S-E, Ahn H-Y, et al. The association between normal alanine aminotransferase levels and the metabolic syndrome: 2005 Korean National Health and Nutrition Survey. *Metab Clin Experiment* 2009;58:1731–6.
90 Hsiao P-J, Kao K-K, Shin S-J, et al. Significant correlations between severe fatty liver and risk factors for metabolic syndrome. *J Gastroenterol Hepatol* 2007;22:2119–23.
91 Musso G, Gambino R, Bo S, et al. Should nonalcoholic fatty liver disease be included in the definition of metabolic syndrome? *Diabetes Care* 2008;31: 562–8.
92 Postic C, Girard J. Contribution of de novo fatty acid synthesis to hepatic steatosis and insulin resistance: lessons from genetically engineered mice. *J Clin Invest* 2008;118:829–38.
93 Lee K. Relationship between uric acid and hepatic steatosis among Koreans. *Diab Metab* 2009;35: 447–51.
94 Fan JG, Farrell GC. Epidemiology of non-alcoholic fatty liver disease in China. *J Hepatol* 2009;50:204–10.
95 Popkin BM. Global nutrition dynamics: the world is rapidly shifting toward a diet linked with noncommunicable diseases. *Am J Clin Nutr* 2006;84:289–98.
96 Heath GW. Physical activity transitions and chronic diseases. *Am J Lifestyle Med* 2009;3:27S–31S.
97 Zelber-Sagi S, Nitzan-Kaluski D, Goldsmith R, et al. Long-term nutritional intake and the risk for nonalcoholic fatty liver disease (NAFLD): a population-based study. *J Hepatol* 2007;47:711–7.
98 Oya J, Nakagami T, Sasaki S, et al. Intake of n-3 polyunsaturated fatty acids and non-alcoholic fatty liver disease: a cross-sectional study in Japanese men and women. *Eur J Clin Nutr* 2010;xx:1–7.
99 Abdelmalek MF, Suzuki A, Guy C, et al., for the NASH-CRN. Increased fructose consumption is associated with fibrosis severity in patients with non-alcoholic fatty liver disease. *Hepatol* 2010;51:1961–71.
100 Musso G, Gambino R, De Michieli F, et al. Dietary habits and their relations to insulin resistance and postprandial lipemia in nonalcoholic steatohepatitis. *Hepatol* 2003;37:909–16.
101 Zelber-Sagi S, Nitzan-Kaluski D, Goldsmith R, et al. Role of leisure-time physical activity in nonalcoholic fatty liver disease: a population-based study. *Hepatol* 2008;48:1791–8.
102 Dunn W, Xu R, Schwimmer JB. Modest wine drinking and decreased prevalence of suspected non-alcoholic fatty liver disease. *Hepatol* 2008;47:1947–54.
103 Suzuki A, Angulo P, St. Sauver J, et al. Light to moderate alcohol consumption is associated with lower frequency of hypertransaminasemia. *Am J Gastroenterol* 2007;102:1912–9.
104 Argo CK, Northup PG, Al-Osaimi AMS, Caldwell SH. Systematic review of risk factors for fibrosis

progression in non-alcoholic steatohepatitis. *J Hepatol* 2009;51:371–9.
105 Musso G, Gambino R, Cassader M, Pagano G. Meta-analysis: Natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. *Ann Med* 2010; ePub:1–33; doi: 10.3109/07853890.2010.518623.
106 Dunn W, Xu R, Wingard DL, et al. Suspected non-alcoholic fatty liver disease and mortality risk in a population-based cohort study. *Am J Gastroenterol* 2008;103:2263–71.
107 Adams LA, Lymp JF, St. Sauver J, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. *Gastroenterol* 2005;129:113–21.
108 Ong JP, Pitts A, Younossi ZM. Increased overall mortality and liver-related mortality in non-alcoholic fatty liver disease. *J Hepatol* 2008;49:608–12.
109 Söderberg C, Stål P, Askling J, et al. Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. *Hepatol* 2010;51:595–602.
110 Angulo P. Long-term mortality in nonalcoholic fatty liver disease: is liver histology of any prognostic significance? *Hepatol* 2010;51:373–5.
111 Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3:e442.
112 Karelis AD. Metabolically healthy but obese individuals. *Lancet* 2008;372:1281–3.
113 Ruderman N, Chisholm D, Pi-Sunyer X, Schneider S. The metabolically obese, normal-weight individual revisited. *Diabetes* 1998;47:699–713.
114 Stefan N, Kantartzis K, Machann J, et al. Identification and characterization of metabolically benign obesity in humans. *Arch Intern Med* 2008;168:1609–16.
115 Villanova N, Moscatiello S, Ramilli S, et al. Endothelial dysfunction and cardiovascular risk profile in nonalcoholic fatty liver disease. *Hepatol* 2005;42:473–82.
116 Targher G, Bertolini L, Padovani R, et al. Relation of nonalcoholic hepatic steatosis to early carotid atherosclerosis in healthy men. *Diab Care* 2004;27:2498–500.
117 Sookoian S, Pirola CJ. Non-alcoholic fatty liver disease is strongly associated with carotid atherosclerosis: A systematic review. *J Hepatol* 2008;49:600–7.
118 Petit JM, Guiu B, Terriat B, et al. Nonalcoholic fatty liver is not associated with carotid intima-media thickness in type 2 diabetic patients. *J Clin Endocrinol Metab* 2009;94:4103–6.
119 Fraser A, Harris R, Sattar N, et al. Gamma-glutamyltransferase is associated with incident vascular events independently of alcohol intake: analysis of the British Women's Heart and Health Study and meta-analysis. *Arterioscler Thromb Vasc Biol* 2007;27:2729–35.
120 Haring R, Wallaschofski H, Nauck M, et al. Ultrasonographic hepatic steatosis increases prediction of mortality risk from elevated serum gamma-glutamyl transpeptidase levels. *Hepatology* 2009;50:1403–11.
121 Sanyal AJ, Banas C, Sargeant C, et al. Similarities and differences in outcomes of cirrhosis due to non-alcoholic steatohepatitis and hepatitis C. *Hepatology* 2006;43:682–9.
122 Targher G, Bertolini L, Padovani R, et al. Prevalence of non-alcoholic fatty liver disease and its association with cardiovascular disease in patients with type 1 diabetes. *J Hepatol* 2010;53:713–8.
123 McKimmie RL, Daniel KR, Carr JJ, et al. Hepatic steatosis and subclinical cardiovascular disease in a cohort enriched for type 2 diabetes: the Diabetes Heart Study. *Am J Gastroenterol* 2008;103:3029–35.
124 Targher G, Day CP, Bonora E. Risk of cardiovascular disease in patients with nonalcoholic fatty liver disease. *New Eng J Med* 2010:363:1341–50.
125 Ghouri N, Preiss D, Sattar N. Liver enzymes, nonalcoholic fatty liver disease, and incident cardiovascular disease: a narrative review and clinical perspective of prospective data. *Hepatology* 2010;52:1156–61.
126 Bae JC, Rhee EJ, Lee WY, et al. Combined effect of nonalcoholic fatty liver disease and impaired fasting glucose on the development of type 2 diabetes: a 4-year retrospective longitudinal study. *Diab Care* 2011;34: 727–9.
127 Yamada T, Fukatsu M, Suzuki S, et al. Fatty liver predicts impaired fasting glucose and type 2 diabetes mellitus in Japanese undergoing a health check-up. *J Gastroenterol Hepatol* 2010;25:352–6.
128 Fraser A, Harris R, Sattar N, et al. Alanine aminotransferase, γ-glutamyltransferase, and incident diabetes. The British Women's Heart and Health Study and meta-analysis. *Diab Care* 2009;32:741–50.
129 Hanley AJG, Wagenknecht LE, Festa A, et al. Alanine aminotransferase and directly measured insulin sensitivity in a multiethnic cohort. The Insulin Resistance Atherosclerosis Study. *Diab Care* 2007;30:1819–27.
130 Wallace TM, Utzschneider KM, Tong J, et al. Relationship of liver enzymes to insulin sensitivity and intra-abdominal fat. *Diab Care* 2007;30:2673–8.

References for Fig 32.1

1 Fan JG, Zhu J, Li XJ, et al. Prevalence of and risk factors for fatty liver in a general population of Shanghai, China. *J Hepatol* 2005;43:508–14.

2 Zhou YJ, Li YY, Nie YQ, et al. Prevalence of fatty liver disease and its risk factors in the population of South China. *World J Gastroenterol* 2007;47:6419–24.
3 Li H, Wang YJ, Tan K, et al. Prevalence and risk factors of fatty liver disease in Chengdu, Southwest China. *Hepatobiliary Pancreat Dis Int* 2009;8:377–82.
4 Zelber-Sagi S, Nitzan-Kaluski D, Halpern Z, Oren R. The prevalence of primary non-alcoholic fatty liver disease in a population-based study and its association with biochemical and anthropometric measures. *Liv Int* 2006;26:856–63.
5 Bedogni G, Miglioli M, Masutti F, et al. Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos Nutrition and Liver Study. *Hepatology* 2005;42:44–52.
6 Hamaguchi M, Kojima T, Takeda N, et al. The metabolic syndrome as a predictor of non-alcoholic fatty liver disease. *Ann Intern Med* 2005;143: 722–8.
7 Jimba S, Nakagami T, Takahashi M, et al. Prevalence of non-alcoholic fatty liver disease and its association with impaired glucose metabolism in Japanese adults. *Diabet Med* 2005;22:1141–5.
8 Omagari K, Kadokawa Y, Masuda J, et al. Fatty liver in non-alcoholic non-overweight Japanese adults: incidence and clinical characteristics. *J Gastroenterol Hepatol* 2002;17:1098–105.
9 Kim HJ, Kim HJ, Lee KE, et al. Metabolic significance of nonalcoholic fatty liver disease in nonobese, nondiabetic adults. *Arch Intern Med* 2004;164:2169–75.
10 Lizardi-Cervera J, Laparra DI, Chavez-Tapia NC, et al. Prevalence of NAFLD and metabolic syndrome in asymptomatic subjects. *Rev Gastroenterol Mex* 2006; 71:453–9.
11 Chen CH, Huang MH, Yang JC, et al. Prevalence and risk factors of nonalcoholic fatty liver disease in an adult population of Taiwan: metabolic significance of nonalcoholic fatty liver disease in nonobese adults. *J Clin Gastroenterol* 2006;40:745–52.
12 Chen CH, Huang MH, Yang JC, et al. Prevalence and etiology of elevated serum alanine aminotransferase level in an adult population in Taiwan. *J Gastroenterol Hepatol* 2007;22:1482–9.
13 Park SH, Jeon WK, Kim SH, et al. Prevalence and risk factors of non-alcoholic fatty liver disease among Korean adults. *J Gastroenterol Hepatol* 2006;21:138–43.
14 Browning JD, Szczepanaik LS, Dobbins R, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 2004;40:1387–95.
15 Dassanayeke AS, Kasturirante A, Rajindrajith S, et al. Prevalence and risk factors for non-alcoholic fatty liver disease among adults in an urban Sri Lankan population. *J Gastroenterol Hepatol* 2009;24:1284–8.
16 Das K, Das K, Mukherjee PS, et al. Nonobese population in a developing country has a high prevalence of nonalcoholic fatty liver and significant liver disease. *Hepatology* 2010;51:1593–602.
17 Mohan V, Farooq S, Deepa M, et al. Prevalence of non-alcoholic fatty liver disease in urban south Indians in relation to different grades of glucose intolerance and metabolic syndrome. *Diab Res Clin Pract* 2009;84: 84–91.
18 Sung KC, Ryan MC, Kim BS, et al. Relationships between estimates of adiposity, insulin resistance, and nonalcoholic fatty liver disease in a large group of nondiabetic Korean adults. *Diab Care* 2007;30:2113–18.
19 Riquelme A, Arrese M, Soza A, et al. Non-alcoholic fatty liver disease and its association with obesity, insulin resistance and increased serum levels of C-reactive protein in Hispanics. *Liv Int* 2009;29:82–8.
20 Völzke H, Nauck M, Rettig R, et al. Association between hepatic steatosis and serum IGF1 and IGFBP-3 levels in a population-based sample. *Eur J Endocrinol* 2009;161:705–13.
21 Pendino GM, Mariano A, Surace P, et al. Prevalence and etiology of altered liver tests: a population-based survey in a Mediterranean town. *Hepatology* 2005;41:1151–9.
22 Clark JM, Bracanti FL, Diehl AM. The prevalence and etiology of elevated aminotransferase levels in the United States. *Am J Gastroenterol* 2003;98:960–7.
23 Ioannou GN, Boyko EJ, Lee SP. The prevalence and predictors of elevated serum aminotransferase activity in the United States in 1999–2002. *Am J Gastroenterol* 2006;101:76–82.
24 Saito T, Nishise Y, Makino N, et al. Impact of metabolic syndrome on elevated serum alanine aminotransferase levels in the Japanese population. *Metab Clin Experiment* 2009;58:1067–75.
25 Liu CM, Tung TH, Liu JH, et al. A community-based epidemiological study of elevated serum alanine aminotransferase levels in Kinmen, Taiwan. *World J Gastroenterol* 2005;11:1616–22.
26 Jamali R, Khonsari M, Merat S, et al. Persistent alanine aminotransferase elevation among the general Iranian population: Prevalence and causes. *World J Gastroenterol* 2008; 4:2867–71.

Answers to multiple choice questions

1. D
2. B
3. C

33 Epidemiology of common tropical GI diseases

Magnus Halland[1], *Rodney Givney*[2], *& Anne Duggan*[3]
[1]University of Newcastle, Callaghan, NSW, Australia
[2]Division of Microbiology, Hunter Area Pathology, Pathology North & Newcastle University, New South Wales, Australia
[3]Department of Gastroenterology, John Hunter Hospital, Newcastle, NSW, Australia

Key points

- Tropical sprue is an illness manifested by chronic diarrhea and malabsorption of uncertain etiology.
- Parasitic infections are numerically among the most significant causes of intestinal and hepatic disease worldwide.
- Leishmaniasis, trypanosomiasis, schistosomiasis, echinococcosis, amoebiasis, and liver flukes cause severe gastrointestinal morbidity and significant mortality.
- Although predominantly tropical diseases, migration has resulted in these diseases occurring more frequently in temperate climates.
- Immunosuppression of asymptomatically infected patients can result in life-threatening disease.

Tropical sprue

Tropical sprue is an illness manifested by chronic diarrhea and malabsorption of uncertain etiology. It is an important cause of malnutrition in endemic areas. It has a strictly limited geographical distribution between 30 degrees north and south of the equator [1]. Not all regions are affected equally, raising the possibility of genetic susceptibility in some populations [2]. It is more common in India, Haiti, Cuba, Puerto Rico, and the Dominican Republic, but rarely observed in Africa, the Bahamas, and Jamaica. It affects local inhabitants as well as visitors to tropical regions, but usually a 4–6 week stay in a high-risk area is required to acquire the disease. Ingestion of contaminated water and rancid fats has been associated with the disease in some studies. Recent data from India indicate that tropical sprue remains an important cause of malabsorption as 22–29 % of patients evaluated in tertiary malabsorption clinics are diagnosed with this disease [3].

Diseases caused by trypanosomatida

Leishmaniasis

This disease is caused by the flagellate protozoa of the genus *Leishmania*, of which at least 20 pathogenic species have been identified. The global impact of the disease is enormous, with an estimated prevalence of 12 million cases worldwide with a global yearly incidence of 2 million [4]. The illness is spread by the female phlebotomine and lutzomyia sandflies in tropical regions. Rarely, the disease can be acquired through injecting drug use, blood transfusion, solid organ transplantation, or *in utero* transmission.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Figure 33.1 Distribution of visceral leishmaniasis. Data source and map production: WHO/NTD/IDM HIV/AIDS, Tuberculosis and Malaria (HTM) World Health Organization, October 2010. Source: http://www.who.int/leishmaniasis/leishmaniasis_maps/en/. Reproduced with permission of the World Health Organization.

Gastrointestinal manifestations occur in visceral leishmaniasis (kala azar), and mucocutuaneous leishmaniasis. Epidemics can also occur when conditions are favorable (malnutrition and mass human movement), as recently observed in Sudan in 1997 [5]. Different clinical manifestations observed in various geographical regions are set out in Table 33.1.

Visceral leishmaniasis (VL)

The two dominant species causing VL are *L. donovani* and *L. infantum*. *L. donovani* occurs in South Asia and East Africa and transmission is mostly person-to-person, as opposed to *L. infantum* where the major reservoir is domestic dogs. *L. infantum* is more common in children aged less than 10 years. Ninety percent of cases occur in Bangladesh, India, Nepal, Sudan, and Brazil (see Figure 33.1). HIV infection increases the relative risk of developing active VL after exposure by more than 100 times [6]. The major gastrointestinal manifestations include hepatic and splenic enlargement as well as anemia. Attempts at disease

Table 33.1 Variation in common manifestations and geographic region of leishamaniasis

Manifestation	Region
Visceral	Bangladesh, Brazil, India, Nepal, Sudan
Cutaneous	Afghanistan, Brazil, Iran, Peru, Saudi Arabia, Syria
Mucocutaneous	Bolivia, Brazil, Peru

control are mainly aimed at vector control by use of indoor insecticide spraying or insecticide-treated bed nets. Early diagnosis and treatment also decreases the human reservoir available for disease transmission, but many affected persons cannot access nor afford treatment.

Mucocutaneous leishmaniasis

Mucocutaneus leishmaniasis or espundia is seen chiefly in persons infected by *L. braziliensis* and related species, which are endemic from Mexico to Argentina. The mucosal manifestations can occur months, years, or decades after the initial infection and can be locally destructive to all mucosal surfaces and surrounding structures. The disfigurement that results adds to the disease burden and social stigma.

American trypanosomiasis (Chagas disease)

Chagas disease is caused by the protozoan *Trypanosoma cruzi* and is a major health problem in Latin American countries with estimates ranging from 10 million to close to 20 million people affected [7]. It is estimated that 10,000 persons die from the disease annually. The disease is active in 18 countries in the American continent but also affects close to 0.5 million people outside the endemic region with most cases recorded in the United States and Spain. An estimated 50,000 new cases occur annually in endemic regions. Rural areas are high risk, with virtually no vectorial transmission observed in urban, developed countries. Chronic infection can lead to cardiac and digestive tract malfunction with lethal outcome for the human host. The most common gastrointestinal manifestations of this disease are pseudoachalasia and megacolon, which occurs in approximately 10 % of chronically infected persons.

The infection is transmitted by a vector, usually a species of the triatomine bug, an insect that survives on blood meals from humans, wild rodents, and small animals. Direct contact with the insect or its urine and feces remains the dominant mode of transmission. Other modes of infection include:

- congenital (estimated 15,000 cases annually)
- contaminated blood products
- organ transplantation
- contaminated foodstuff.

Changing trends in epidemiology primarily relate to reducing vector exposure by improving housing and sanitary conditions as well as insecticide treatment of human dwellings. Also, major improvements with screening of donated blood for *T. cruzi* has transformed this major route of disease spread (formerly the second most common mode of transmission) to minor, isolated events over the last decade [8–10]. Latin American countries have now been declared free of vectorial and transfusional transmission. Although no vaccine is currently available, advances with vaccine technology, particularly DNA vaccines, may emerge as future options for both disease prevention and treatment. Migration from endemic regions to developed countries has led to emergence of congenital Chagas disease in such countries and transfusional spread is emerging as a potential health hazard requiring routine screening of blood donors. Vector-borne spread in urban, developed countries is exceedingly rare.

Diseases caused by other protozoa

Entamoeba histolytica

This protozoal infection can lead to intestinal and extraintestinal complications, most commonly dysentery and liver abscesses respectively. Although over 90 % of infected persons are asymptomatic, more than 40 million people worldwide develop clinical disease annually and it is estimated that 100,000 deaths can be attributed to *E. histolytica* [11,12]. Two nonpathogenic species exist (*E. dispar* and *E. moshkovskii*) and antigen, serologic, and molecular testing can distinguish these from *E. histolytica*. Although amebiasis occurs worldwide the prevalence is much higher in developing countries due to poor sanitation. Transmission occurs via ingestion of amebic cysts, usually contaminated water or food. Venereal transmission (mainly fecal-oral contact) is relevant mainly to men who have sex with men in both developed and developing countries [13]. The main targets of reduction of disease relates to avoidance of untreated water and undercooked food in endemic areas, and avoiding high-risk sexual behaviors.

Cystoisospora belli (formerly *Isospora belli*)

This coccidian causes self-limiting disease in immunocompetent individuals, but can lead to chronic diarrhea and weight loss among persons with

immunodeficiency, mainly AIDS. Although reports of infection occur worldwide, tropical, and subtropical areas have the highest prevalence [14].

Cyclospora cayetanensis

This coccidial infection shares characteristics of being a cause of food and waterborne diarrheal illness in developing countries, but also has the potential to cause major food- related outbreaks due to contaminated food from endemic areas. Several such outbreaks occurred in the 1990s in the United States and Canada, but also more recently, an outbreak related to contaminated salad in central Europe occurred [15]. Increasing globalization and international trade of fresh food raises the risk of future outbreaks unless sanitary systems are improved in developing countries.

Sarcocystis

A rare cause of symptomatic human infection which occurs due to ingestion of undercooked meat containing sporocysts. High rates of asymptomatic carriage are observed on stool analysis of healthy persons in endemic regions such as central and southeast Asia [16].

Diseases caused by trematoda

Schistosomiasis [17]

Schistosomiasis is a parasitic infection which affects over 230 million people worldwide with more than 90 % of infected persons residing in Africa [15,17]. Twenty million people suffer severe consequences of the disease leading to more than 200,000 deaths annually in sub-Saharan Africa alone. Therefore its morbidity and mortality ranks third after malaria and intestinal helminths. The number of persons accessing treatment rose from 12.4 million in 2006 to 33.5 million in 2010. Several species cause disease, but the majority of human infections causing gastrointestinal disease are caused by the intestinal schistosomes: *S. mansoni* and *S. japonicum*.

The major intermediate host for *Schistosoma* are snails. Exposure to infested fresh water through agricultural work, domestic chores, and recreational activities is the dominant method of acquiring the infection. In endemic areas 60–100 % of children and adolescents become infected.

Immunologic reaction to *Schistosoma* eggs lodged in tissues lead to a wide range of symptoms depending on egg burden and location. Hepatic fibrosis, portal hypertension, gastrointestinal bleeding, gastrointestinal obstruction, and malnutrition are the commonest gastrointestinal manifestations of disease.

Changing trends in epidemiology has seen the disease eradicated from Japan and the Lesser Antilles islands. Transmission has been halted or markedly reduced in Tunisia, Morocco, Saudi Arabia, and Venezuela. Unfortunately targeted interventions to prevent water infestation, control of the intermediate snail host and diagnosis and treatment of infected persons has not yet decreased disease magnitude significantly in Africa, although incidence and prevalence is decreasing in Brazil, China, and Egypt. Changes in agricultural practices and irrigation along with migration are emerging as contributors to the disease spreading to previously uninfected regions and populations.

Clonorchiasis and opisthorchiasis [18]

Clonorchis sinensis (East Asia), *Opisthorchis viverrini* (Southeast Asia), and *Opisthorchis felineus* (former Soviet Union) have similar life cycles and clinical manifestations. Eggs ingested by snails hatch into miracidia which develop into the cercaria that are released into water and infect carp, other freshwater fish, and crustaceans (*C. sinensis* only). Humans are infected by eating uncooked fish or crustaceans. Adult worms mature in the bile ducts, causing cholangitis that can be complicated by bacterial sepsis, liver abscess, and pancreatitis. Cholangiocarcinoma is strongly associated with *O. viverrini* in Thailand. Millions are thought to be infected, and in endemic regions, 20–80 % of the population may be infected.

Fascioliasis [18,19]

Fasciola hepatica, the common liver fluke, infects an estimated 2.5 million people across 50 different countries in temperate regions. The reported incidence is rising, but it is likely that this is due to improved diagnostics and reporting in endemic regions. The main hosts are sheep and cattle and infected persons usually reside in regions where close interaction between humans and ruminants occur. Humans are

infected by ingesting metacercariae in uncooked food such as watercress or ingesting infected water. Gastrointestinal manifestations are mainly hepatic, where migration of flukes through the liver parenchyma causes necrosis and fibrosis. Bile ducts are often affected.

Disease prevalence is high in Latin America (Peru, Bolivia, and Ecuador) where 10–80 % of persons are infected. Fascioliasis was rarely reported in humans in Vietnam until 1997. Currently, between 2000 and 4000 cases are diagnosed there every year. It is an important cause of anemia in these regions, particularly among children. In other regions human infestation is sporadic, and in developed countries the disease is limited to travelers or immigrants.

Fasciolopsis

Fasciolopis buski the giant intestinal fluke, can cause anemia, bowel obstruction, and chronic diarrhea. It infects humans and pigs, and snails serve as the intermediate host.

Infection reaches endemic levels in China, Taiwan, Indonesia, and Malaysia. India is similarly affected. It is estimated that the prevalence is over 10 million persons, commonly school-aged children. Disease prevention focuses on avoiding eating raw vegetables in endemic regions, proper sanitation, and snail control. The impact on incidence and changing trends in epidemiology are largely unknown.

Echinostomatidae

These tiny flukes are common in Southeast Asia, and as opposed to many other trematodes, echinostomiasis occurs mainly in urban areas. Humans are infected by ingesting the intermediate host, molluscs, snails, and crustaceans. Gastrointestinal manifestations depend on fluke load and range from malnutrition and diarrhea to severe anemia and intestinal perforation. Epidemologic data are limited but a recent study from Cambodia found that 23 % of schoolchildren were positive for *Echinostoma* based on stool testing for eggs and worms [20,21].

Heterophydae

These tiny trematodes infect fish-eating animals including humans. Infection occurs when undercooked fish or shrimps containing metacercariae are ingested, and leads to diarrhea, abdominal pain, and nausea. Highest prevalence is observed in eastern Asia where raw fish (sushi) is popular and disease control is aimed at avoiding undercooked fish from endemic regions [20].

Diseases caused by nematodes

Ascaris lumbricoides

This roundworm causes human infection worldwide, but the prevalence of infection is highest in temperate regions with high rainfall. The global impact of disease is enormous with an estimated 1 billion humans infected worldwide [22]. Most infected persons live in Asia (73 %), Africa (12 %), and South America (8 %). *Ascaris* commonly infects infants and children leading to malnutrition, growth retardation, and anemia. Further gastrointestinal manifestations of disease include gastrointestinal obstruction from high worm load and pancreato-biliary disease from migration into the biliary tree. It is estimated that 20,000 deaths from complications of gastrointestinal obstruction occur annually.

Transmission occurs via ingestion of water and food contaminated with ova from human stool. The lifecycle requires the ingested ova to develop into larvae which penetrate the intestine and migrate to the lung. Within the alveoli maturation occurs and migration through the bronchial tree to the pharynx facilitates a return to the small intestine where the further maturation occur leading to production of ova. Multiplication within the host does not occur.

Preventative efforts focus on improved sanitation and changing farming practices to prevent contamination of soil with human feces. Boiling of drinking water in endemic areas can lessen worm burden. Mass treatment among school-aged children has been found to reduce disease intensity but not prevalence.

Hookworms

The two main hookworm species infecting humans are *Necator americanus* and *Ancylostoma duodenale*. The geographic distribution of these two worms is shown in Table 33.2. It is estimated that up to 740 million people worldwide are infected [23,24], most in developing nations. Transmission mostly occurs through

Table 33.2 The geographic distributions of *Necator americanus* and *Ancylostoma duodenale*

Necator americanus	*Ancylostoma duodenale*
North and South America	Mediterranean countries
Central Africa	Iran, India, Pakistan
Indonesia	
Parts of India	Far East

direct contact with larvae that can penetrate skin (often entering through soles of feet, arms, or legs) but ingestion of larvae can also lead to disease. The lifecycle is similar to *Ascaris* (described earlier) with maturation in pulmonary vasculature and return through the GI tract through migration up the bronchial tree and subsequent swallowing. Gastrointestinal manifestations of infection include nausea, diarrhea, and vomiting at the time of larval infestation of the small intestine. Chronic malnutrition and anemia can develop as a result of chronic blood loss due to worms feeding on mucosal capillaries. Hookworm (typically transmitted from a family dog) can cause eosinophilic ileocolitis. Disease control focuses on preventing contamination of soil with human feces and avoidance of direct contact between soil and human skin in endemic areas.

Strongyloides species

Infection caused by *Strongyloides stercoralis* (and rarely *S. fuelleborni*) occurs in all continents except Antarctica but occurs mainly in tropical, subtropical, and warm temperate regions. WHO estimates conservatively that 30–100 million people are infected [25,26]. Strongyloidiasis has almost disappeared in countries where sanitation and human waste disposal have improved. In the United States, a series of small studies in select populations have shown that between 0–6 % of persons sampled were infected. Studies in immigrant populations have shown a much higher percentage of infected persons ranging from 0–46 %. Infection by penetration of the skin by filariform larvae occurs through direct contact with contaminated soil. The larvae migrate via the circulatory system to the lung alveolae, thence by the bronchial tree to the pharynx where they are swallowed and come to the small intestine. There the larvae develop into adult worms, the eggs of which produce rhabditiform larvae which are either excreted or develop into autoinfecting filariform larvae which again cycle back through the lungs. Infection is usually asymptomatic but may cause abdominal, pulmonary and skin complaints. Disseminated hyperinfection in the immunosuppressed is life-threatening.

Diseases caused by cestoda [27]

Four species of adult cestode tapeworms commonly inhabit the human small intestine: *Taenia saginata*, *Taenia solium*, *Hymenolepis nana*, and *Diphyllobothrium latum*. More rarely adult worm human intestinal infections are caused by *Hymenolepis diminuta* and *Dipylidium caninum*. These infections are often asymptomatic but may result in bloating, abdominal pain, diarrhea, and obstruction.

Larval cestode infections that result from consumption of parasite eggs include cystic hydatid disease (*Echinococcus granulosus*), alveolar hydatid disease (*Echinococcus multilocularis*), and cysticercosis (*Taenia solium*). Rarer larval cestode diseases affecting humans include coenurosis (*Taenia multiceps*), sparganosis (*Spirometra mansonoides*), and cysticercosis caused by *Taenia crassiceps*. Clinical presentation depends on the location of the cyst which, with the notable exception of hydatid liver disease, is most often outside the gastrointestinal tract.

Taenia saginata

Beef tapeworm is endemic worldwide. It is especially prevalent in some parts of Africa, Central and South America, eastern and western Asia, and some countries in Europe. Ingestion of eggs from contaminated pasturelands by grazing cattle results in development in cattle tissues of the infective cysticercus stage. After ingestion of the cysticercus in raw or inadequately cooked beef, it takes approximately 2–3 months for the infection to become patent in the human host. The Southeast Asia species (*Taenia saginata asiatica*) is morphologically similar to *T. saginata* but the cysticercus stage occurs in the liver of pigs more frequently than in cattle. Patients usually exhibit no symptoms with these *Taenia species* but may notice passing proglottids. The mature worm can also cause abdominal discomfort, diarrhea, and occasionally intestinal obstruction.

Taenia solium

Pig tapeworm is endemic in Mexico, Central and South America, southern Europe, Africa, India, Southeast Asia, and the Philippines. Adult tapeworm infection in the human intestine is asymptomatic unless cysticercosis, caused by fecal-oral autoinfection with parasite eggs and subsequent larval infection of extraintestinal sites occurs.

Diphyllobothrium latum

While this fish tapeworm is generally associated with Northern Hemisphere sub-Arctic climates, cases also occur on the southwest coast of South America. In the early 1970s, diphyllobothriosis was estimated to affect million humans globally, with 5 million in Europe, 4 million in Asia, and the rest in America. More recent data indicate that 20 million people are infected worldwide, but no recent estimation concerning the global prevalence of this parasitosis is available. Clinical manifestations are mechanical bowel obstruction, diarrhea, abdominal pain, and, mainly in northern European populations, pernicious anemia resulting from vitamin B12 deficiency [28].

Hymenolepis nana

H. nana is normally a parasite of mice with worldwide distribution. Humans may acquire the infection by ingestion of infected beetles (e.g. in dry cereals) but direct infection is more common. *H. nana* usually occurs in institutional and familial settings in which hygiene is substandard. Internal autoinfection with the parasite also occurs. Hymenolepiasis may cause abdominal pain, diarrhea, headaches, or irritability, probably in infections with heavier worm burdens.

Echinococcus species [29]

Human cystic hydatid disease caused by *E. granulosus* an important cause of human morbidity, requiring costly surgical and medical treatment. Liver and lungs are commonly affected and more rarely heart, brain, bone, or other locations. There are areas of high endemicity in southern South America, the Mediterranean coast, the southern part of the former Soviet Union, the Middle East, southwestern Asia, northern Africa, Australia, Kenya, New Zealand, and Uganda. Sporadic local transmission occurs in Alaska and other states in the United States. In many countries of the endemic regions, national diagnostic incidence rates of cystic echinococcus range from 5 to 20 per100,000 population but the risk is small for urban populations. Sheep are infected with the larval stage (cystic hydatid) by ingesting infective eggs dispersed from the feces of the tapeworm-infected dog. The major risks are uncontrolled dogs living closely with people, uncontrolled slaughter of livestock, and insanitary living conditions amongst populations involved with sheep-raising. The adult worm of *E. multilocularis*, the cause of human alveolar echinococcosis, lives in the small intestine of the definitive host, commonly Northern Hemisphere wild predators in parts of Europe, Asia, Japan, and North America, including Alaska. However, human infections (alveolar hydatid disease) occur by accidental ingestion of the oncosphere by contamination with dog feces. Incidence of disease even in endemic areas is low, ranging from 0.02–1.4 per 100,000. The coexistence of the sylvatic cycle can make control difficult.

Conclusions

Despite aggressive efforts to halt transmission and treat tropical GI infections, the magnitude and impact of disease remains high. A multitarget approach, focusing on improving sanitation, improving access to diagnosis and treatment, as well as the development of vaccines are required to lessen current disease burden

Multiple choice questions

1 A 24-year-old US business traveler is reviewed 1 month after he returned from overseas travel. Four weeks ago he spent 48 hours at a conference in Kuala Lumpur. He rapidly develops diarrhea upon return. There is no associated vomiting, fever, or chills. After a few days his symptoms settle somewhat, but mild diarrhea persists for another month. Stool samples are collected and results are pending. He has no significant previous medical background, and only occasionally takes a multivitamin. He is worried about whether he has contracted tropical sprue.

Which of the following options is most likely?

A He has contracted tropical sprue and should commence 6 months treatment with antibiotics and folic acid supplementation

B He has probably attracted tropical sprue, but a gastroscopy and colonoscopy should be performed to exclude other causes prior to commencing treatment

C He is unlikely to have contracted tropical sprue and should be evaluated for other causes of diarrhea

D He is not at risk of tropical sprue due to the short duration of stay in a tropical region

E His diarrhea is most likely due to a parasitic gut infection and empirical treatment is warranted

2 A 27-year-old woman from rural Brazil is diagnosed with American trypanosomiasis (Chagas disease.). She lives on a rural property along with her parents and works part-time at a farm. She recently received a blood transfusion after a postpartum hemorrhage.

Which of the following alternatives is the least likely explanation for her acquiring Chagas disease?

A Transfusional spread due to the blood transfusion

B Congenital spread

C Bitten by triatomine bug

D Contact with triatomine bug urine or feces

E Ingestion of contaminated foodstuff

3 A 26-year-old male presents with diarrhea after a two-week holiday in Mexico. He admits to high-risk sexual behavior with men during his holiday. He is afebrile, but is having six loose stools daily. He has no previous medical history and was screened for HIV 3 months ago, which was negative. He takes no regular medication or over-the-counter therapy. Stool culture shows *Entamoeba dispar*.

Which of the following is the most appropriate next step in management?

A Proceed with further investigation of the diarrhea

B Treat empirically for traveler's diarrhea with oral antibiotic

C Advice on gluten and lactose avoidance, reassess symptoms in 2 weeks

D Treat *E. dispar* with parasitic therapy as most likely cause of symptoms

E Perform flexible sigmoidoscopy to assess for CMV colitis

4 A 9-year-old boy living in an Indian village develops anemia, lethargy, and marked hepatosplenomegaly. Investigations reveal infection with visceral leishamaniasis (species *L. infantum*).

Which of the following is the major reservoir for the *L. infantum*?

A Domestic dogs

B Humans

C Domestic cattle

D Bats

E Sandflies

5 Schistosomiasis is a parasitic infection that affects over 230 million people worldwide. Which of the following statements correctly describes the transmission and life-cycle of schistosomiasis?

A Transmission occurs via ingestion of contaminated water and food. The life-cycle requires a migration from the intestine to the lung, and a return to the GI tract via swallowing

B Exposure to contaminated fresh water through agricultural work or similar is the most common route of transmission, and the major intermediate hosts are snails

C Humans are infected when eating undercooked fish or crustaceans, and snails are the intermediate host

D Humans are infected when eating undercooked beef from regions where grazing cattle are exposed to grass contaminated by eggs

References

1 Gray GM. (1995) Tropical sprue, in *Infections of the Gastrointestinal Tract* (eds. MJ Blaser, PD Smith, JI Ravdin), Raven Press, New York, NY.

2 Walker MM. What is tropical sprue? *J Gastroenterol Hepatol* 2003;18:887–90.

3 Yadav P, Das P, Mirdha BR, et al. Current spectrum of malabsorption syndrome in adults in India. *Indian J Gastroenterol* 2011;30:22–8.

4 Murray HW, Berman JD, Davies CR, Saravia NG. Advances in leishmaniasis. *Lancet* 2005;366:1561.

5 World Health Organization. Leishmaniasis: Epidemics. http://www.who.int/leishmaniasis/epidemic/en/ (last accessed May 24, 2013).

6 World Health Organization. Leishmaniasis: Burden of disease. http://www.who.int/leishmaniasis/burden/en/ (last accessed May 24, 2013).

7 Rassi A Jr, Rassi A, Marin-Neto JA. Chagas disease. *Lancet* 2010;375(9723):1388.

8 World Health Organization (2004) *The World Health Report 2004: Changing History*. WHO, Geneva, Switzerland.
9 Schmunis GA, Cruz JR. Safety of the blood supply in Latin America. *Clin Microbiol Rev* 2005;18:12.
10 Moncayo A, Ortiz Yanine MI. An update on Chagas disease (human American trypanosomiasis). *Ann Trop Med Parasitol* 2006;100:663.
11 Entamoeba taxonomy. *Bull World Health Organ* 1997;75:291–4.
12 Li E, Stanley SL, Jr. Protozoa. Amebiasis. *Gastroenterol Clin North Am* 1996;25:471.
13 James R, Barrett J, Marriott D, et al. Seroprevalence of Entamoeba histolytica infection among men who have sex with men in Sydney, Australia. *Am J Trop Med Hyg* 2010;83:914–16.
14 Lindsay DS, Upton SJ, Weiss LM. *Isospora*, *Cyclospora* and *Sarcocystis* in *Manual of Clinical Microbiology*, 10th edn (Chief Ed. J Versalovic), ASM Press, Washington, DC, pp. 2172–9.
15 Döller PC, Dietrich K, Filipp N, et al. Cyclosporiasis outbreak in Germany associated with the consumption of salad. *Emerg Infect Dis* 2002;8:992–4.
16 Suh NS, Kozarsky P, Keystone JS. *Cyclospora cayatensis*, *Isospora belli*, *Sarcocystis* species, *Balantidium coli* and *Blastocystis hominis* in *Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases*, 7th edn (eds. GL Mandell, JE Bennett, R Dolin), Churchill Livingston/Elsevier, Philadelphia, PA, pp. 3561–8.
17 World Health Organization. Schistosomiasis: Epidemiological situation. http://www.who.int/schistosomiasis/epidemiology/en/ (last accessed May 24, 2013).
18 Maguire JH. (2010) Trematodes (Schistosomes and Other Flukes) in *Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases*, 7th edn (eds. GL Mandell, JE Bennett, R Dolin), Churchill Livingston/Elsevier, Philadelphia, PA, pp. 3595–3605.
19 Report of the WHO Expert Consultation on Foodborne Trematode Infections and Taeniasis/Cysticercosis http://www.who.int/neglected_diseases/preventive_chemotherapy/WHO_HTM_NTD_PCT_2011.3.pdf (last accessed August 6, 2013).
20 Mas-Coma MS, Esteban JG, Bargues MD. Epidemiology of human fascioliasis: a review and proposed new classification. *Bull World Health Organ* 1999;77:340–6.
21 Sohn WM, Chai JY, Yong TS, et al. Echinostoma revolutum infection in children, Pursat Province, Cambodia. *Emerg Infect Dis* 2011;17(1):117.
22 Tietze PE, Tietze PH. The roundworm, Ascaris lumbricoides. *Prim Care* 1991;18:25.
23 Hotez PJ, Brooker S, Bethony JM, et al. Hookworm infection. *New Engl J Med* 2004;351:799.
24 de Silva NR, Brooker S, Hotez PJ, et al. Soil-transmitted helminth infections: updating the global picture. *Trends Parasitol* 2003;19:547.
25 World Health Organization. Neglected tropical diseases: Strongyloidiasis. http://www.who.int/neglected_diseases/diseases/strongyloidiasis/en/ (last accessed May 24, 2013).
26 Centers for Disease Control and Prevention (CDC). Parasites: *Strongyloides*. http://www.cdc.gov/parasites/strongyloides/epi.html (last accessed May 24, 2013).
27 Garcia HH, Jimenez JA, Escalante H. (2012) Cestodes, in *Manual of Clinical Microbiology*, 10th edn (Chief Ed. J Versalovic), ASM Press, Washington, DC, pp. 2222–9.
28 Scholz T, Garcia HH, Kuchta R, Wicht B. Update on the human broad tapeworm (Genus *Diphyllobothrium*), including clinical relevance. *Clin Microbiol Rev* 2009:146–60.
29 Schantz PM. Progress in diagnosis, treatment and elimination of echinococcosis and cysticercosis. *Parasitol Int* 2006;55:S7–1.

Answers to the multiple choice questions

1. C

This gentleman is unlikely to have contracted tropical sprue, which usually affects travelers who spend 4–6 weeks in high-risk regions. Reports of disease with shorter exposure do occur and hence option D is incorrect. Treatment for tropical sprue is inappropriate at this point in time and he should be evaluated for other causes of chronic diarrhea. Parasitic gut infections could be a cause for his symptoms, but rarely affect business travelers who visit urban areas. A stool sample should guide management.

2. A

Latin American countries have now been declared virtually free of transmission from blood products, a major success towards halting transmission. The other methods of transmission remain active although improved sanitary conditions and pesticide use in human dwellings is reducing transmission.

3. A

The most appropriate option is to further investigate his diarrhea. Many men who have sex with men are colonized with *E. dispar* and it is rarely the cause for symptoms. Treating empirically for traveler's diarrhea is sometimes appropriate, but in this patient with

high-risk sexual exposures further investigation is warranted. Advice on gluten and lactose is inappropriate in this patient. Flexible sigmoidoscopy is an appropriate test if CMV colitis is considered. This patient was HIV negative 3 months ago, and this makes CMV colitis from immune failure very unlikely.

4. A

L. infantum is more common in children aged 10 and above, and the major reservoir is domestic dogs. Infections with *L. donivani*, the other major species of VL, is most commonly spread via direct person-to-person contact as humans are the reservoir. Domestic cattle and bats are not a reservoir for leishamaniasis; sandflies are a common vector, but not a reservoir.

5. B

Exposure to contaminated fresh water through agricultural work or domestic chores is the most common mode of transmission. The immunologic response to ova deposited in tissues lead to disease. Snails are the major intermediate hosts.

Option A is incorrect as this describes transmission and life-cycle or ascaris, a roundworm. Although snails are the major intermediate host option C is incorrect as this describes transmission of clonorchiasis and opisthorchiasis, a common infection in Southeast Asia. Option D is incorrect as this describes characteristics of *Tania saginata*, the beef tapeworm.

34 Nutritional epidemiology and GI cancers

Linda E. Kelemen[1] *& Ilona Csizmadi*[2]

[1]Department of Population Health Research, Alberta Health Services-Cancer Care, Departments of Medical Genetics and Oncology, University of Calgary, Calgary, AB, Canada
[2]Department of Population Health Research, Alberta Health Services-Cancer Care, Departments of Community Health Sciences and Oncology, Faculty of Medicine, University of Calgary, Calgary, AB, Canada

Key points

- The choice of dietary assessment method depends on the research question and the pathophysiology of the disease.
- Long-term dietary patterns are most relevant to estimate chronic disease risk.
- Most risk models will need to be adjusted for total energy intake.
- Knowledge of potential sources of error in nutritional assessment is essential.
- Statistical methods exist to estimate and to minimize measurement error.

Introduction

One of the earliest applications of nutritional epidemiology was to the study of gastrointestinal diseases. In the late 1960s, Burkitt [1] observed differences in fecal bulk between individuals in rural Africa compared to industrialized Western countries and hypothesized that this was due to the high fiber intake of the former. Subsequently, he hypothesized that dietary fiber protects against the development of colorectal cancer. To date, there are more than 1000 published accounts on this topic.

Nutritional epidemiology is the assessment of diet and its relationship to the causes of diseases in populations. Broadly speaking, diet includes the intake of essential nutrients (e.g. vitamins, minerals, and amino acids), energy sources (protein, carbohydrate, fat, and alcohol), naturally occurring food compounds (e.g. plant fiber, cholesterol, and caffeine), or derived compounds such as the intake of chemicals formed in cooking (e.g. heterocyclic aromatic amines formed in well-done or charred meats) or from food processing (e.g. *trans* fatty acids).

The investigator's choice of dietary assessment method will depend on his or her knowledge of the disease pathology. Events that are acute and occur over a relatively short period of time, such as maternal dietary folate intake and risk of fetal neural tube defects, requires methods that accurately and precisely assess an individual's intake over the course of a few days. In contrast, events such as cancer that are chronic and are complicated by exposure time, require methods that capture patterns of consumption among populations over a period of years, since dietary intake several years prior to disease manifestation likely represents the more relevant exposure period for understanding these diseases. Diet-disease associations may be confounded or modified by several factors, including body size, physical activity, other dietary factors and genetic susceptibility. Understanding the interplay among these factors is crucial to derive unbiased estimates of disease risk.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Dietary assessment instruments

Two methods of dietary assessment typically used in clinical settings have been modified for use in epidemiologic studies. Both the 24-hour recall of dietary intake and the diet record assess short-term dietary intake, but when used as repeated measurements, can inform *usual patterns* of intake over a longer period.

The *24-hour recall* interview is administered in person or by telephone. Subjects report their exact intake in the preceding 24 hours guided by the interviewer's standard questions that may also rely on visual aids to assist with recall of portion size. In its favor, memory of recent intake may be more precise and quantities may be estimated with greater accuracy with minimal participant burden. Well-trained interviewers are required, however, and the nutrient coding and analysis of food intake can be laborious. Because individual diets vary from day to day, a single day's dietary recall does not represent usual dietary intake for some nutrients.

The *diet record* is similar to the 24-hour recall, except that the subject records actual food and beverage intake prospectively over several days. Subjects are asked to provide detailed descriptions of preparation methods and food quantities, which are assessed by weighing, volume/dimension measurements, or estimation assisted by the use of photographs. The prospective nature of diet recording reduces errors associated with recall and minimizes omission of foods consumed. However, the method requires a high level of subject literacy, motivation and training, and can be costly to code and analyze. Furthermore, consecutive days of dietary recording may result in food intake that is highly correlated from day to day (due to consumption of leftover meals or intentional alteration of usual diet), possibly introducing bias. A trade-off is to collect fewer records per subject on a greater number of individuals. Like the 24-hour recall, multiple days of records over several months or one year can reduce day-to-day correlation of intake, improve accuracy and precision of individual intake, and capture seasonal variation in food intake.

For investigations of several hundred or thousand individuals such as those participating in large, prospective studies, *food frequency questionnaires (FFQs)* are a viable option to assess long-term diet. These questionnaires consist of a pre-determined list of foods and beverages that represent the major contributors to the macro- and micronutrient content of the diet of the population under study. Thus, they are population or ethnic-specific [2,3]. For each food or beverage item, the subject selects one of several options that best defines their frequency of intake, typically over the past year, with or without a selection for a portion size option (Figure 34.1). Photographs of different serving sizes can be used to assist with portion size recall. FFQs are easily administered in person or by mail, provide information on the intake of a large number of foods, food groups and individual nutrients, and are substantially less expensive to analyze particularly if optical scanning is used for data entry. Repeated FFQ administrations over several years can capture dietary changes over time.

Nutritional epidemiology of gastrointestinal cancers

According to the most recent publication by the World Cancer Research Fund/American Institute for Cancer Research [4], "convincing" evidence between the association of foods, nutrients or dietary constituents and various cancers affecting the GI system are few, while evidence is weaker for other dietary components. The WCRF/AICR panel comprised leading scientists in the field of cancer epidemiology, as well as leaders in nutrition and the biology of cancer. Strength of the evidence was based on a systematic approach to review all relevant observational and experimental evidence, as well as expert judgment. Table 34.1 summarizes the strength of evidence between dietary components and various GI cancers based on the WCRF/AICR report [4]. This section reviews selected findings from the panel that have been updated with recent studies.

Esophagus cancer

Alcohol

Alcoholic beverages and ethanol in alcoholic beverages are carcinogenic to humans [5]. Convincing evidence exists to support the association that alcohol increases the risk of esophagus cancer. In a meta-analysis of one cohort and 20 case-control studies by the WCRF/AICR scientific panel, consumption of one

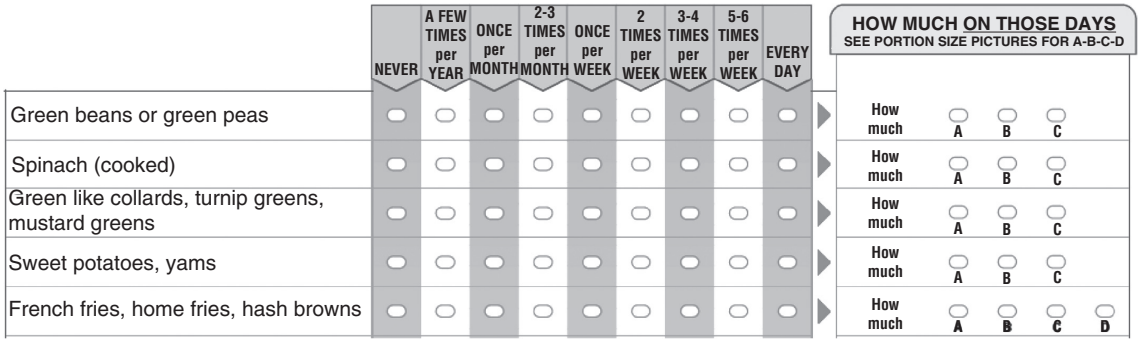


Figure 34.1 Example of the format of the Block Food Frequency Questionnaire 2005.
Source: www.nutritionquest.com/. Reproduced with permission of Block Dietary Data Systems.

alcoholic beverage per week was associated with a 4 % increase in risk of esophagus cancer (RR 1.04; 95 % CI 1.03–1.05) [4]. The panel observed statistically significant differences in the relative risks across studies that appeared to be attributable to the size, rather than the direction, of the estimated effect.

In more recent prospective investigations with refined phenotyping and adjustment for smoking, a 10 g day^{-1} increment of ethanol (approximately 9–10 oz beer, 4 oz wine, or 1 oz spirits) was associated with a 22–28 % increase in risk of squamous cell carcinoma among European men and a 31–62 % increase in risk of squamous cell carcinoma among European women [6,7]. Alcohol did not appear to be associated with esophageal adenocarcinoma [7,8]. Increased risks seemed to be associated similarly with beer, wine, and spirits [4,8]. At intake ranges of approximately 40–99 g day^{-1} ethanol, increased risks appeared to be

Table 34.1 Dietary components and their strength of evidence for association with GI cancers.

	Dietary component and its strength of evidence*	
Cancer site	"Convincing" evidence	"Probable" evidence
Colon and rectum	Alcohol (men) ↑ Red and processed meats ↑	Alcohol (women) ↑ Foods containing dietary fiber ↓ Garlic ↓ Milk ↓ Calcium ↓
Esophagus	Alcohol ↑	Non-starch vegetables ↓ Fruits ↓ Foods containing beta-carotene ↓ Foods containing vitamin C ↓ Hot maté intake ↑
Liver	Alflatoxins ↑	Alcohol ↑
Pancreas		Foods containing folate ↓
Stomach		Non-starch vegetables ↓ Fruits ↓ Allium vegetables ↓ Salt, and salted and salty foods ↑

*The direction of the association is denoted beside each dietary component to be either an increase (↑) or decrease (↓) in risk. Information is abstracted from Reference [4].

higher among men of Asian compared to European ancestry [6,9–11].

Risk estimates and their statistical significance are influenced by the choice of reference category and the cutpoints chosen for categories of consumption. Two studies used more flexible modeling approaches to estimate how the risk of esophagus cancer varies across different levels of alcohol consumption. Polesel et al. [12] showed that both linear and cubic (Figure 34.2(a)) regression splines fit the data better than categories of alcohol intake, with the highest levels of consumption (>150 g day^{-1} ethanol) associated with a relative risk of 10 among 343 men with unspecified esophagus cancer histology. Rota et al. [13] employed a two-step process of fitting two-term fractional polynomial models to individual studies followed by pooled random-effects analysis of squamous cell carcinoma using 3000 cases. They reported that >100 g day^{-1} ethanol intake was associated with a relative risk of 11 (Figure 34.2(b)). As can be seen in both figures, statistically significant risks are observed for intermediate doses of alcohol, with no apparent threshold. Furthermore, the relative risk is underestimated by both the step function model of approximate quintile categories of ethanol consumption (Figure 34.2(a)), and the linear model (Figure 34.2(b)) particularly at moderate (<50 g day^{-1}) intake levels.

Alcohol, ADH and ALDH polymorphisms

Alcohol dehydrogenase (ADH) and acetaldehyde dehydrogenase (ALDH) enzymes play critical roles in the metabolism of ethanol in alcoholic beverages to acetaldehyde, and of acetaldehyde to acetate, respectively [14]. Polymorphisms in these genes, particularly *ADH1B* (previously called *ADH2*) and *ALDH2*, though rare in Caucasians, are common in individuals of Asian ancestry [15] and can modify the speed at which ethanol is metabolized. Carriers of the *ADH1B**1/*1 variant genotype are slow metabolizers of ethanol and experience prolonged exposure to ethanol after drinking. Carriers of the *ALDH2**2/*2 variant genotype experience unpleasant flushing responses and nausea induced by severe acetaldehydemia [16]. In a meta-analysis of Asian populations, the *ADH1B**1/*1 variant vs *2/*2 wild-type genotype was associated with a twofold increase in risk of esophagus cancer of unspecified histology (RR 2.17; 95 % CI 1.08–4.34), with a higher risk observed among heavy drinkers (RR 3.22; 95 % CI 2.27–4.57) [17]. The risk associated with the

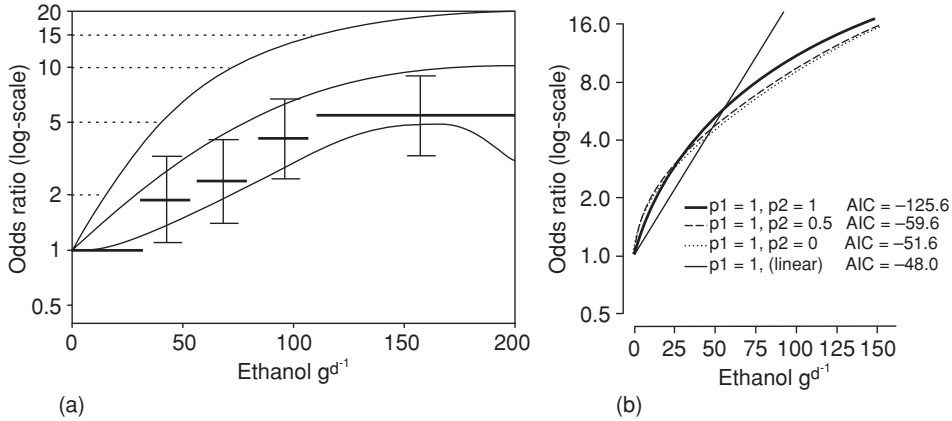


Figure 34.2 Nonlinear models of the association between ethanol consumption and esophagus cancer. (a) Estimates of ORs and 95 % CIs (upper and lower curves) using cubic regression spline (—) and step function (—) models. The referent category was never drinkers for spline models and was the first quintile for the step function models [12]. (b) Estimates of ORs and 95 % CIs by the linear model (solid straight line) and the three second-order fractional polynomials (curved lines) resulting in smaller Akaike's Information Criterion (AIC) values. Smaller AIC values denote the best-fitting model compared to the linear model [13].
Source: Comprised of data from: (a) Polesel et al. 2005 [12], and (b) Rota et al. 2010 [13].

*ALDH2**2/*2 variant vs *1/*1 wildtype genotype was not statistically significant (RR 1.31; 95 % CI 0.52–3.34), yet it was elevated among heavy drinkers (RR 7.12; 95 % CI 4.67–10.86) [17]. Thus, the main effect of the genotypes without consideration of the modifying influence of alcohol intake underestimated the risk associations at high alcohol intakes.

The mechanisms of ethanol-induced carcinogenesis are closely related to the metabolism of ethanol, and include the generation of DNA point mutations, DNA adducts, and other mutagens [14]. The upper and lower GI tract are particularly vulnerable to high acetaldehyde concentrations owing to the oxidation of ethanol by bacteria in saliva and the large intestine following moderate alcohol intake [14].

Colorectal cancer

Alcohol

The WCRF/AICR panel determined that convincing evidence exists to support the association between alcohol intake and colorectal cancer [4]. A meta-analysis of nine cohort studies from Europe, the United States, and Japan reported that each 10 g day^{-1} increment in ethanol consumption was associated with an increase in risk of colon cancer (RR 1.09; 95 % CI 1.03–1.14) and colorectal cancer (RR 1.06; 95 % CI 1.01–1.12) [4]. Similar relative risks were reported from a large prospective cohort in eight European countries of over 2000 colorectal cancer cases [18]. Stratified meta-analyses for colorectal cancer for each 10 g day^{-1} ethanol increase gave summary effect estimates of 1.09 (95 % CI 1.02–1.15) for seven studies for men and 1.00 (95 % CI 0.89–1.40) for three studies for women [4]. A pooled analysis of data from eight cohort studies of women and men from North America and Europe, including over 4600 with colorectal cancer followed for 6 to 16 years, reported a relative risk of 1.41 (95 % CI 1.16–1.72) for those who consumed $\geq$45 g day^{-1} ethanol from alcohol [19]. No increased risk was observed below intakes of 30 g day^{-1}, and no significant statistical heterogeneity was observed by study, sex, cancer site, or specific alcoholic beverage [19], suggesting that the positive association is attributable to ethanol intake itself rather than to a specific beverage. Some studies reported stronger associations of higher alcohol intake among men than women [18–20] despite a similar dose-response relationship [20]. The weaker association among women may be because of the generally lower consumption of alcohol among women [4]. Others found stronger associations among men than women for rectal cancer [20,21].

Red and processed meat

Differences in findings even among cohort studies between the association of red and processed meat with colorectal cancer may be due to various reasons, including differences in the populations or in the endpoints studied (e.g. colon vs. colorectal carcinomas), follow-up duration, and the range of intake captured by the FFQ. For example, two national US cohorts examined the ratio of red to white meat intake with risk of colon cancer among women [22,23] (Figure 34.3). Higher intakes increased risk in both cohorts, which were similar in sample size and duration of follow-up (6–9 years). The overlap in intake distributions, however, suggests the full extent of increased risk is observed only at very high intakes.

The WCRF/AICR panel classified the strength of evidence between red and processed meat intake and increased risk for colorectal cancer as "convincing" [4]; however, this was not without controversy [24,25]. The panel based their decision on a meta-analysis of seven cohort studies with a summary relative risk of 1.43 (95 % CI 1.05–1.94) for each time per week that red meat was consumed, and a meta-analysis of three cohort studies with a relative risk of 1.29 (95 % CI 1.04–1.60) for each 100 g day^{-1} red meat intake [4]. In support of the panel's recommendation is a meta-analysis of 19 cohort studies including almost 8000 cases from the United States, Europe, Australia, and Japan relating an increase in risk of colon or colorectal cancer with 120 g day^{-1} red meat intake (RR 1.28; 95 % CI 1.18-1.39) and with 30 g day^{-1} processed meat intake (RR 1.09; 95 % CI 1.05–1.13) [26]. No significant between-study statistical heterogeneity was found. Although the relative risks were lower among women compared to men, the overlapping confidence intervals suggested no significant differences in risk estimates [26]. Three large investigations reported higher relative risks associated with cancer of the rectum than of the colon, but all report that the differences were either marginally or not statistically significant [26–28].

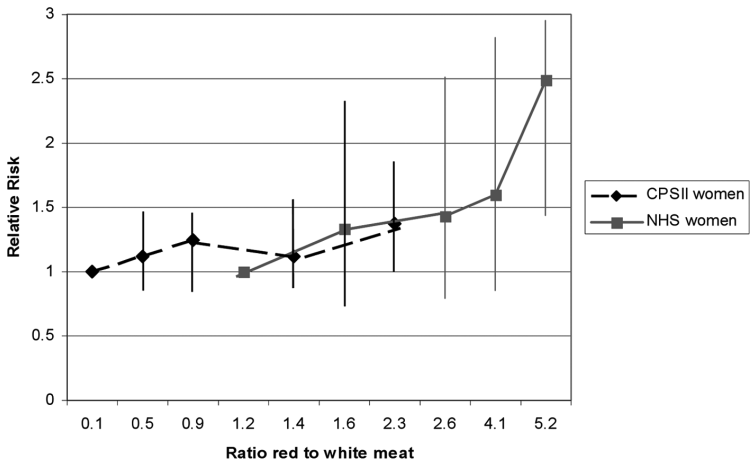


Figure 34.3 Ratio of red to white meat intake and risk of colon cancer among women in two national US cohorts. NHS, Nurses' Health Study [22]; CPSII, Cancer Prevention Study II [23]. Vertical bars represent 95 % confidence intervals.
Source: Comprised of data from Willett et al. 1990 [22], and Chao et al. 2005 [23].

The proposed mechanisms by which red and processed meats may influence colorectal cancers are, among other factors, through potentially mitogenic fatty acid components, exposure to *N*-nitroso compounds from food processing, the ingestion of potent carcinogens formed during cooking, and/or the increase in fecal iron that initiates the generation of hydroxyl radicals [29].

Vitamin D

Dietary sources of vitamin D are largely from fortified foods including dairy products and breakfast cereals, and also from fatty fish, fish liver oils, and liver. Vitamin D is also synthesized from cholesterol precursors in the skin by the action of UV B radiation. Exposure to sunlight for 15 minutes at wavelengths between 290 and 315 nm (found at the equator or in peak summer months) is sufficient to significantly increase the levels of pre-vitamin D_3, which is the form ingested from foods [30]. Pre-vitamin D_3 is metabolized to vitamin D_3, which then requires two obligate hydroxylations in the liver (25(OH)D) and kidney to form the biologically active hormone, 1,25-dihydroxyvitamin D ($1,25(OH)_2D$) [31]. Consequently, serum 25(OH)D concentration is the best indicator for determining adequacy of vitamin D of an individual since it represents a summation of the total cutaneous production of vitamin D as well as ingested vitamin D.

The WCRF/AICR panel concluded that the evidence was "limited or suggestive" to support an association between foods containing vitamin D and decreased risk with colorectal cancer [4]. Analyses that rely solely on dietary estimates of vitamin D can be confounded by other correlated food components, such as calcium and animal fats that are also found in dairy products, and which have been associated with colorectal cancer. These variables are not always adjusted in analyses. Different fortification practices between countries can further confound the interpretation of associations. On the other hand, analyses of serum vitamin D should take into account physical activity level. Because physical activity may often be outdoors, it is correlated with sunlight exposure.

Because case-control studies measure serum 25(OH)D after the diagnosis of cancer, when tumor growth may have altered serum concentrations, these studies were not summarized here. A meta-analysis of 10 prospective cohort studies that measured plasma 25(OH)D concentration prior to cancer incidence reported a 15 % decrease in risk for colorectal cancer risk for a 10 ng mL^{-1} increase in serum 25(OH)D (RR 0.85; 95 % CI 0.79–0.92) after controlling for factors such as body mass index and physical activity [32]. There was statistically significant heterogeneity in risk estimates across the studies, and this appeared to be due to the magnitude, rather than the direction, of the effects [32]. Results from two recently published prospective cohort studies reported conflicting results: up to a 46 % decrease in risk of colorectal cancer for plasma 25(OH)D above the second quintile ($\geq$16.8 ng mL^{-1}) in one study [33] compared to no association with colon cancer in another study [34].

Clinical trials of vitamin D supplementation and risk of colorectal cancer have been limited, and results are equivocal for a beneficial effect at higher exposure [35,36]. In the Women's Health Initiative, the hazard ratio for colorectal cancer incidence after 7 years of supplementation with vitamin D (400 IU) and calcium was 1.08 (95 % CI 0.86–1.34); however, participants were permitted to self-supplement at up to 600 IU per day, complicating the trial's interpretation [35]. Supplementation with vitamin D (1,100 IU) and calcium was associated with a lower risk of developing various nonskin cancers including colorectal cancer after four years (RR 0.40; 95 % CI 0.20–0.82). The resulting increase in 25(OH)D, per 1 ng mL^{-1}, was suggestive of lower risk for all cancers combined (RR 0.98; 95 % CI 0.97–1.0) [36].

Vitamin D plays a role in calcium homeostasis, and higher calcium intake has been associated with reduced risk of colorectal cancer [37]. Higher vitamin D status may protect against cancer by reducing cellular proliferation and angiogenesis or inducing differentiation and apoptosis [38]. Vitamin D could also act locally in the colon to inhibit carcinogenesis, since both 1-α-vitamin D hydroxylase, the enzyme that metabolizes 25(OH)D to $1,25(OH)_2D$, and the vitamin D receptor, which binds the active $1,25(OH)_2D$ hormone, are expressed in the colon and elsewhere [38].

Methodologic issues in nutritional epidemiology

Statistical adjustment for total energy

Statistical adjustment for energy intake in models of diet and disease is important for several reasons. Because intakes of nutrients, particularly macronutrients, are correlated with total energy intake, these nutrients may be noncausally associated with disease from confounding by total energy intake [39]. Residual confounding from factors difficult to measure or measured with error that are associated with energy intake, including body size, physical activity and metabolism, can attenuate associations with disease risk. Failure to account for total energy intake can obscure associations between nutrient intakes and disease risk or possibly reverse the direction of the association. Several disease-risk models are described to control for energy intake in epidemiologic studies [39], although studies show the superiority of one or two statistical models over others [40].

Measurement error correction

While advances in nutritional epidemiology have increased our understanding of the role that diet plays in the etiology of a number of chronic diseases, conflicting and inconsistent results have lead to controversy, particularly with regard to the accuracy of dietary intake methods that rely on self-reports [41–44]. This controversy is perhaps most relevant to the role of diet in cancer prevention [45]. Mounting evidence suggests important bioactive roles for dietary components as determinants of cancer risk and tumor behavior [46], yet few associations have been classified as "convincing" or "probable" by the joint WCRF/AICR review panel [4]. The large, well-conducted prospective studies that have failed to find a consistent relation between dietary components (such as fat, fiber, and fruits and vegetables) and cancers of the breast, colon, or rectum, or overall [47–50] may be explained by a true lack of diet–cancer associations or, alternatively, by serious methodological limitations of the studies themselves, including measurement error associated with self-reporting methods such as the FFQ [44].

Over the years, investigators have recognized that self-reporting methods of dietary intake are subject to substantial systematic and random error and both have profound implications for the design, analysis, and interpretation of nutritional epidemiologic studies [51,52]. Dietary measurement error often attenuates (biases toward one) the estimates of disease relative risks and reduces the statistical power to detect their significance. Therefore, an important relation between diet and disease may be obscured [51,53,54]. This concern has prompted researchers involved in large epidemiologic investigations to re-examine current dietary assessment methods and to explore the use of new technologies and innovative analytical strategies and methodologies [46,55,56]. Several options that promise to advance nutritional epidemiology are emerging [55]. These options include the use of: (i) novel designs and new analytical methods to improve the quality of dietary intake estimates; and (ii) objective biologic markers to calibrate diet–disease estimates of risk.

An area of great interest is the development of cost-effective, self-administered tools that can provide more valid measures of dietary intake. While there is great enthusiasm for the use of cell phones, cameras and other objective forms of capturing dietary intake, these technologies may not be valid for the ascertainment of habitual intake because they may be "reactive", meaning that they may promote behavior change with their use. This assumption will need to be tested as these technologies become available for use in large studies with diverse populations [57]. Researchers at the US National Cancer Institute have responded to the need for better cost-effective dietary intake instruments by creating the web-based Automated Self-administered 24-hour Dietary Recall (ASA24^{TM}) software program. The ASA24 builds on the strengths of the 24-hour dietary intake method and uses state-of-the-art automated computer technology with graphic enhancements and animated characters to guide participants in reporting previous day's food intake [58,59]. Audio language and cues to enhance use in low-literacy populations are also incorporated. The ASA24 is available on the NCI website (http://riskfactor.cancer.gov/tools/instruments/asa24/).

Another advancement is the development of analytical methods to combine dietary intake data from multiple instruments. For example, statistical modeling is currently being tested to incorporate the precision of the 24-hour recall with estimates of usual or habitual intake captured by the FFQ. Methods are being developed to improve upon dietary intake estimates at the individual [60] and population levels [61–64].

The use of objective biologic markers that directly reflect the intake of specific dietary constituents under study has been proposed for correction of the attenuation of diet–disease risk that results from measurement error in diet [65]. This was shown in the Women's Health Initiative randomized controlled trial and observational study. In a sub-sample of the subjects, the authors compared self-reported dietary intake with the objective recovery biomarkers, doubly labeled water and urinary nitrogen excretion [65]. Recovery biomarkers are based on the concept of the metabolic balance between intake and excretion over a fixed period and provide an estimate of absolute intake levels [66]. The mean self-reported total energy intake was estimated as 1477 kcal day^{-1} from the FFQ, which grossly underestimated the mean intake of 2141 kcal day^{-1} estimated using doubly labeled water [65]. By "calibrating" the self-reported dietary intakes with the biomarkers in regression analyses that also included several other subject characteristics, "adjusted" risk estimates of total energy and protein intake could be obtained in the larger cohort that were closer to the true risk estimates. The WHI study showed, for example, that prior to calibration of risk estimates, the association of a 20 % increase in energy intake with breast cancer risk was not statistically significant (RR 0.99; 95 % CI 0.95–1.02). Following calibration, the association was RR = 1.25 (95 % CI 1.07–1.47), indicating that measurement error (random and systematic) is common, substantial and may have obscured diet–disease associations in previous analyses [47–50].

Conclusions

Nutritional epidemiology has contributed significantly to our understanding of the relationships between diet and disease over the past four decades. Ongoing investigations that further characterize important exposure periods (early life, *in utero*), clarify associations within the context of genetic susceptibility, and incorporate both biomarkers of exposure and outcome will continue to elucidate our understanding of the pathophysiology of complex diseases. Recent advances in dietary intake measurement promise to facilitate our understanding of the role that this modifiable lifestyle factor can play in reducing the burden of various chronic diseases including cancer.

Multiple choice questions

1 The 2007 WCRF/AICR report concluded that "convincing" evidence exists to support the association between colorectal cancer and which of the following dietary factors:

A Alcohol
B Red meat
C Foods containing dietary fiber
D A and B only
E A, B, and C

2 The best indicator for determining adequacy of vitamin D status of an individual is:

A Serum 25(OH)D concentration

B Serum $1,25(OH)_2D$ concentration
C Dietary intake
D A and B
E A, B, and C

3 To improve the accuracy of dietary assessment, researchers have proposed:

A New electronic self-administered versions of dietary assessment instruments
B Analytical approaches that combine dietary intake data from more than one instrument
C Incorporation of biologic markers into diet–disease risk models
D B only
E A, B, and C

References

1 Burkitt DP. Epidemiology of cancer of the colon and rectum.*Cancer* 1971;28(1):3–13.
2 Kolonel LN, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;151(4):346–57.
3 Kelemen LE, et al. Development and evaluation of cultural food frequency questionnaires for South Asians, Chinese, and Europeans in North America. *J Am Diet Assoc* 2003;103(9):1178–84.
4 World Cancer Research Fund/American Institute for Cancer Research (2007) *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*, AICR, Washington, DC.
5 World Health Organization/International Agency for Research on Cancer (2010) Alcohol consumption and ethyl carbamate, in *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, IARC, Lyon, France, p. 1424.
6 Weikert C, et al. Lifetime and baseline alcohol intake and risk of cancer of the upper aero-digestive tract in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Int J Cancer* 2009;125(2):406–12.
7 Steevens J, et al. Alcohol consumption, cigarette smoking and risk of subtypes of oesophageal and gastric cancer: a prospective cohort study. *Gut* 2010;59(1):39–48.
8 Freedman ND, et al. A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. *Am J Epidemiol* 2007;165(12):1424–33.
9 Fan Y, et al. Alcohol, tobacco, and diet in relation to esophageal cancer: the Shanghai Cohort Study. *Nutr Cancer* 2008;60(3):354–63.
10 Kimm H, Kim S, Jee SH. The independent effects of cigarette smoking, alcohol consumption, and serum aspartate aminotransferase on the alanine aminotransferase ratio in Korean men for the risk for esophageal cancer. *Yonsei Med J* 2010;51(3):310–17.
11 Oze I, et al. Alcohol drinking and esophageal cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. *Jpn J Clin Oncol* 2011;41(5):677–92.
12 Polesel J, et al. Estimating dose-response relationship between ethanol and risk of cancer using regression spline models. *Int J Cancer* 2005;114(5):836–41.
13 Rota M, et al. Random-effects meta-regression models for studying nonlinear dose-response relationship, with an application to alcohol and esophageal squamous cell carcinoma. *Stat Med* 2010;29(26):2679–87.
14 Seitz HK, Stickel F. Molecular mechanisms of alcohol-mediated carcinogenesis. *Nat Rev Cancer* 2007;7(8):599–612.
15 International HapMap Project, http://www.hapmap.org (accessed July 2013).
16 Harada S, Agarwal DP, Goedde HW. Aldehyde dehydrogenase deficiency as cause of facial flushing reaction to alcohol in Japanese. *Lancet* 1981;2(8253):982.
17 Yang SJ, et al. Relationship between genetic polymorphisms of ALDH2 and ADH1B and esophageal cancer risk: a meta-analysis. *World J Gastroenterol* 2010;16(33):4210–20.
18 Schutze M, et al. Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study. *BMJ* 2011;342:d1584.
19 Cho E, et al. Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. *Ann Intern Med* 2004;140(8):603–13.
20 Moskal A, et al. Alcohol intake and colorectal cancer risk: a dose-response meta-analysis of published cohort studies. *Int J Cancer* 2007;120(3):664–71.
21 Bongaerts BW, et al. Alcohol consumption, type of alcoholic beverage and risk of colorectal cancer at specific subsites. *Int J Cancer* 2008;123(10):2411–7.
22 Willett WC, et al. Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women. *New Engl J Med* 1990;323(24):1664–72.
23 Chao A, et al. Meat consumption and risk of colorectal cancer. *JAMA* 2005;293(2):172–82.
24 Boyle P, Boffetta P, Autier P. Diet, nutrition and cancer: public, media and scientific confusion. *Ann Oncol* 2008;19(10):1665–7.
25 Truswell AS. Problems with red meat in the WCRF2. *Am J Clin Nutr* 2009;89(4):1274–5; author reply 1275–6.
26 Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. *Int J Cancer* 2006;119(11):2657–64.

27 Cross AJ, et al. A prospective study of red and processed meat intake in relation to cancer risk. *PLoS Med* 2007;4(12):e325.
28 Norat T, et al. Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition. *J Natl Cancer Inst* 2005;97(12): 906–16.
29 Cross AJ, Sinha R. Meat-related mutagens/carcinogens in the etiology of colorectal cancer. *Environ Mol Mutagen* 2004;44(1):44–55.
30 Holick MF. The cutaneous photosynthesis of previtamin D3: a unique photoendocrine system. *J Invest Dermatol* 1981;77(1):51–8.
31 Holick MF. McCollum Award Lecture, 1994: Vitamin D – new horizons for the 21st century. *Am J Clin Nutr* 1994;60(4):619–30.
32 Gandini S, et al. Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. *Int J Cancer* 2011;128(6):1414–24.
33 Woolcott CG, et al. Plasma 25-hydroxyvitamin D levels and the risk of colorectal cancer: the multiethnic cohort study. *Cancer Epidemiol Biomarkers Prev* 2010;19(1):130–4.
34 Weinstein SJ, et al. Serum 25-hydroxyvitamin D and risks of colon and rectal cancer in Finnish men. *Am J Epidemiol* 2011;173(5):499–508.
35 Wactawski-Wende J, et al. Calcium plus vitamin D supplementation and the risk of colorectal cancer. *New Engl J Med* 2006;354(7):684–96.
36 Lappe JM, et al. Vitamin D and calcium supplementation reduces cancer risk: results of a randomized trial. *Am J Clin Nutr* 2007;85(6):1586–91.
37 Chung M, et al. Vitamin D and calcium: a systematic review of health outcomes. *Evid Rep Technol Assess* 2009;183:1–420 [full report].
38 Holick MF. Vitamin D deficiency. *New Engl J Med* 2007;357(3):266–81.
39 Willett WC. (1998) *Nutritional Epidemiology*, 2nd edn, Oxford University Press, New York, NY.
40 Michels KB, et al. The effect of correlated measurement error in multivariate models of diet. *Am J Epidemiol* 2004;160(1):59–67.
41 Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? *Cancer Epidemiol Biomarkers Prev* 2005;14(12):2826–8.
42 Freedman LS, et al. Abandon neither the food frequency questionnaire nor the dietary fat-breast cancer hypothesis. *Cancer Epidemiol Biomarkers Prev* 2007;16(6):1321–2.
43 Kelemen LE. Food frequency questionnaires: not irrelevant yet. *Cancer Epidemiol Biomarkers Prev* 2006;15(5):1054.
44 Schatzkin A, V. Kipnis V. Could exposure assessment problems give us wrong answers to nutrition and cancer questions? *J Natl Cancer Inst* 2004;96(21): 1564–5.
45 Prentice RL, et al. Nutrition and physical activity and chronic disease prevention: research strategies and recommendations. *J Natl Cancer Inst* 2004;96(17):1276–87.
46 Milner JA. Diet and cancer: facts and controversies. *Nutr Cancer* 2006;56(2):216–24.
47 Hung HC, et al. Fruit and vegetable intake and risk of major chronic disease. *J Natl Cancer Inst* 2004;96(21):1577–84.
48 George SM, et al. Fruit and vegetable intake and risk of cancer: a prospective cohort study. *Am J Clin Nutr* 2009;89(1):347–53.
49 Hunter DJ, et al. Cohort studies of fat intake and the risk of breast cancer – a pooled analysis. *New Engl J Med* 1996;334(6):356–61.
50 Fuchs CS, et al. Dietary fiber and the risk of colorectal cancer and adenoma in women. *New Engl J Med* 1999;340(3):169–76.
51 Kipnis V, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* 2003;158(1):14–21; discussion 22–6.
52 Prentice RL, et al. Statistical aspects of the use of biomarkers in nutritional epidemiology research. *Stat Biosci* 2009;1(1):112–23.
53 Kipnis V, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol* 2001;153(4):394–403.
54 Kipnis V, et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr* 2002;5(6A):915–23.
55 Schatzkin A, et al. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev* 2009;18(4):1026–32.
56 Thompson FE, et al. Need for technological innovation in dietary assessment. *J Am Diet Assoc* 2010;110(1):48–51.
57 Six BL, et al. Evidence-based development of a mobile telephone food record. *J Am Diet Assoc* 2010;110(1):74–9.
58 Subar AF, et al. Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. *J Am Diet Assoc* 2010;110(1): 55–64.
59 Zimmerman TP, et al. Challenges in converting an interviewer-administered food probe database to self-administration in the National Cancer Institute Automated Self-administered 24-Hour Recall (ASA24). *J Food Compost Anal* 2009;22(Suppl 1):S48–51.

60 Kipnis V, et al. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics* 2009;65(4):1003–10.
61 Tooze JA, et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc* 2006;106(10):1575–87.
62 Dodd KW, et al. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *J Am Diet Assoc* 2006;106(10):1640–50.
63 Midthune D, et al. Binary regression in truncated samples, with application to comparing dietary instruments in a large prospective study. *Biometrics* 2008;64(1):289–98.
64 Subar AF, et al. The food propensity questionnaire: concept, development, and validation for use as a covariate in a model to estimate usual food intake. *J Am Diet Assoc* 2006;106(10):1556–63.
65 Prentice RL, et al. Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women. *Am J Epidemiol* 2009;169(8):977–89.
66 Jenab M, et al. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum Genet* 2009;125(5–6):507–25.

Answers to multiple choice questions

1. D
2. A
3. E

35 The epidemiology of obesity among adults

Cynthia L. Ogden[1], *Brian K. Kit*[1], *Tala H.I. Fakhouri*[1,2], *Margaret D. Carroll*[1], *& Katherine M. Flegal*[3]

[1]Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, USA
[2]Epidemic Intelligence Service, Centers for Disease Control and Prevention, Atlanta, GA, USA
[3]Office of the Director, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, USA

Key points

- Obesity is defined using body mass index (BMI), a proxy measure of body fat based on weight and height.
- Waist circumference, another anthropometric measure, is used as a proxy for abdominal adiposity.
- The prevalence of obesity in the United States, based on BMI, has not changed in recent years.
- Between 1999 and 2008 there was no change in mean energy intake among US adults.
- In 2000, adults were less likely to walk to work or have jobs that required high levels of physical activity than in 1950.

Introduction

Reducing obesity prevalence is a public health priority in the United States (USA) [1] and around the world [2]. The prevalence of obesity among adults in the USA doubled between 1980 and 2000 [3]. Although the rate of increase has slowed, the prevalence of obesity remains high at over one third of the population [4]. Obesity has been linked to a variety of conditions including type 2 diabetes mellitus, nonalcoholic fatty liver disease (NAFLD), certain types of cancer, and coronary artery disease [5]. Obesity has also been shown to increase the risk for premature mortality [6,7].

In this chapter we present a review of the epidemiology of obesity among US adults. Obesity prevalence in the USA is presented and trends are compared to those in Canada and England. A discussion of the determinants of obesity, which include dietary intake, physical activity, and the environment, is also included. The chapter ends with a summary of the consequences of obesity. All data on obesity and diet shown in the figures are based on previously published results from the US National Health and Nutrition Examination Survey (NHANES), a nationally representative survey containing direct measurements of weight, height and body fat, and 24-hour dietary recall interviews [8]. Except for body fat results, which are only available for 1999–2004, the data presented in the figures are from 2007–2008. Detailed definitions for variables can be found in the publications.

Definition of obesity

Obesity is defined as the excessive and abnormal accumulation of fat in adipose tissue [9], although the definitions of excessive and abnormal are not agreed upon. Measuring body fat in individuals or populations is challenging.

Hydrodensitometry or underwater weighing is considered the gold standard in determining body fatness.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology

Hydrodensitometry estimates body density by comparing the weight of an individual out of water (dry weight) with that while fully immersed in water. An individuals' underwater weight will be smaller than his or her dry weight because fat is more buoyant in water compared to fat-free mass. The greater the difference between immersed weight and dry, the greater the percentage of body fat. The technique is time-consuming, cumbersome and expensive, making it difficult to implement in epidemiologic studies [10].

Body fatness can also be obtained from dual-energy X-ray absorptiometry (DXA) imaging. The technology is based on the concept of differential attenuation in the intensity of X-ray as it traverses different tissue types. These differences are then captured via a detector and are used to estimate lean body mass, fat mass, and bone mineral density. A key advantage of DXA methodology is its ability to provide information on body fatness with high accuracy with less time and expense than underwater weighing [11].

Because direct measures of body fatness using the methods described earlier are difficult and expensive, proxy measures based on anthropometric indices that correlate with the measures mentioned previously are often used in epidemiologic studies. One of the most popular and widely used proxy measures for body fat is body mass index (BMI, weight [kg]/height $[m]^2$). Weight and height are easy to measure and low cost making these appealing for large population studies. A major limitation of BMI is it fails to discriminate between lean and fat mass.

BMI can be calculated based on measurements of weight and height or based on self-reported values of weight and height. Systematic bias in self-reported weight and height has been shown to compromise the validity of BMI [12]. Differences in reporting have been seen by sex, race-ethnicity, weight status, and age.

The National Institutes of Health and the World Health Organization recommend the use of BMI categories to classify individuals as underweight (<18.5 kg m^{-2}), normal weight (18.5–24.9 kg m^{-2}), overweight (25–29.9 kg m^{-2}), obese (≥ 30 kg m^{-2}) and extremely (class III) obese (≥ 40 kg m^{-2}) [9,13].

Waist circumference (WC), another anthropometric method, is also used to estimate abdominal adiposity. Because WC is a better estimate of abdominal or central obesity than BMI, and because central adiposity is more predictive of morbidity and mortality as opposed to peripheral adiposity, some studies have recommended the use of WC as a substitute for BMI [14,15]. However, BMI and WC have been shown to be equally predictive of the risk for disease and mortality in other studies [16,17].

A major limitation of WC estimation is that a common location for the measurement of WC is not agreed upon. Different studies measure at different locations [18–20]. These differences complicate inter-study comparisons and make universal cut points for excess WC difficult to identify. Nonetheless, the recommended sex-specific cut-offs for WC are >102 cm for men and >88 cm for women [9].

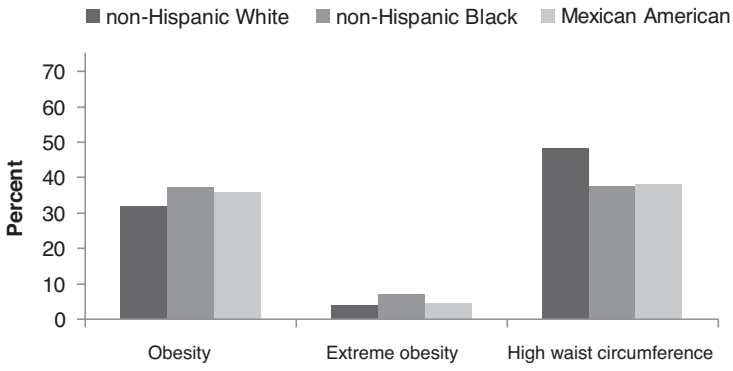
In this chapter, obesity estimates are presented based on the recommended BMI and WC cut points described above. Mean percentage body fat as determined by DXA measurements within BMI categories is also presented.

Obesity prevalence and trends

Published data from 2007–2008 indicate that approximately 34 % or 72 million US adults are obese [4]. More than 5 % of the US population are extremely obese [4], and 53 % have high WC [21]. Sex-specific estimates of obesity, extreme obesity and high WC are shown in Figure 35.1 and Figure 35.2 for each race/ethnic group. Racial or ethnic differences in prevalence of obesity or high WC are not the same among men and women. Among men, the prevalence of high WC is higher among non-Hispanic White men than among non-Hispanic Black men. The prevalence of obesity, however, did not differ significantly between race/ethnic groups. Among women, on the other hand, the prevalence of obesity, extreme obesity, and high WC was significantly lower among non-Hispanic Whites compared to other race/ethnic groups [4,21].

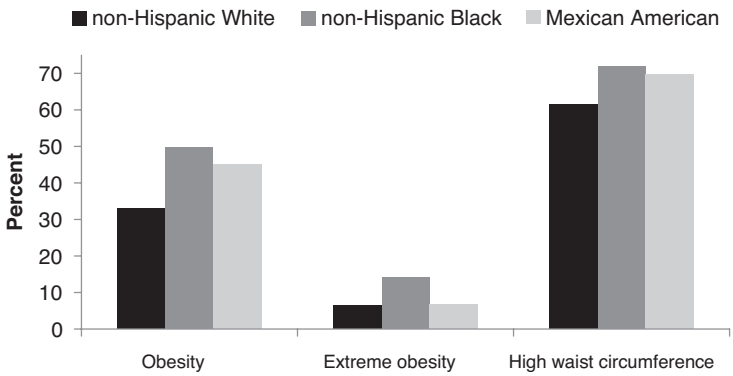
Mean percentage body fat varies by race/ethnicity within BMI categories. Figure 35.3 and Figure 35.4 show mean percentage body fat for men and women within BMI categories. Except in the highest BMI category (BMI ≥ 35), non-Hispanic Black men and women have lower mean percentage body fat than do non-Hispanic White individuals [22].

During the period 1999–2008 there was no significant trend in obesity prevalence among women; however, among men there was a significant increase in



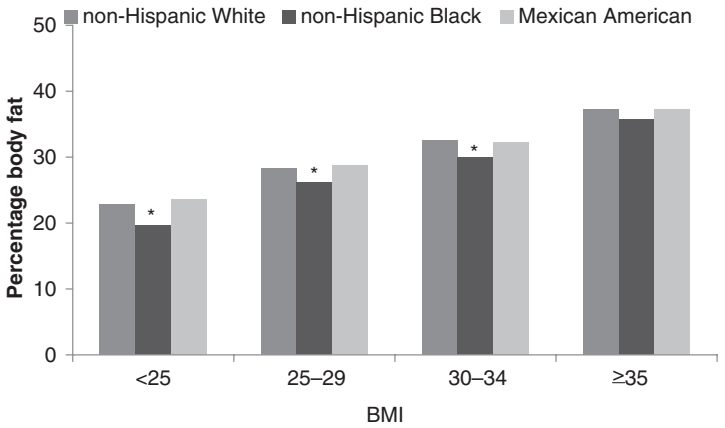
Notes: Obesity defined as body mass index (BMI) ≥30; Extreme obesity defined as BMI ≥40; High waist circumference is >88 cm for women and >102 cm for men. All estimates age -adjusted by the direct method to the 2000 US Census population using age groups 20–39, 40–59, 60+years.

Figure 35.1 Prevalence of obesity, extreme obesity and high waist circumference by race/ethnicity, US men, 20+ years, 2007–2008.
Source: CDC/NCHS, National Health and Nutrition Examination Survey, http://www.cdc.gov/nchs/nhanes.htm; Flegal et al. 2010 [4]; Ford et al. 2011 [21].



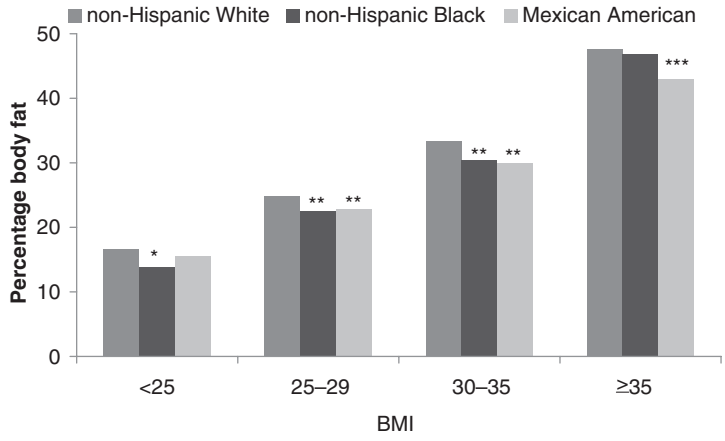
Notes: Obesity defined as body mass index (BMI) ≥30; Extreme obesity defined as BMI ≥40; High waist circumference is >88 cm for women and >102 cm for men. All estimates age-adjusted by the direct method to the 2000 US Census population using age groups 20–39, 40–59, 60+years.

Figure 35.2 Prevalence of obesity, extreme obesity and high waist circumference by race/ethnicity, US women, 20+ years, 2007–2008.
Source: CDC/NCHS National Health and Nutrition Examination Survey, http://www.cdc.gov/nchs/nhanes.htm; Flegal et al. 2010 [4]; Ford et al. 2011 [21].



BMI: body mass index.
*Significantly different from non-Hispanic Whites and Mexican Americans (P<0.05).

Figure 35.3 Mean percentage body fat among US men by BMI category, 20+ years, 1999–2004.
Source: CDC/NCHS, National Health and Nutrition Examination Surveys, http://www.cdc.gov/nchs/nhanes.htm; Li et al. 2009 [22].



BMI: body mass index.
*Significantly different from non-Hispanic Whites and Mexican Americans; (P<0.05).
**Significantly different from non-Hispanic Whites (P<0.05).
***Significantly different from non-Hispanic Whites and non-Hispanic Blacks.

Figure 35.4 Mean percentage body fat among US women by BMI category, 20+ years, 1999–2004.
Source: CDC/NCHS, National Health and Nutrition Examination Surveys, http://www.cdc.gov/nchs/nhanes.htm; Li et al. 2009 [22].

obesity [4]. Figure 35.5 contains obesity prevalence estimates for 1999–2000, 2001–2002, 2003–2004, 2005–2006, and 2007–2008. Estimates are shown for each sex and age group.

An international context

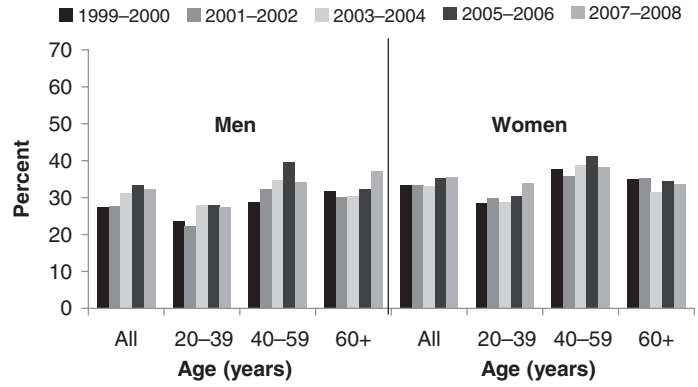
The USA is not alone in experiencing concern about obesity prevalence. Although the prevalence in Canadian adults is not as high as among US adults (Figure 35.6), there have been similar increases in prevalence in Canada as seen in the USA [23]. Obesity prevalence in England is lower than in the USA, yet prevalence estimates appear to be leveling off in England [24] similar to the trend seen in US women (Figure 35.7).

Determinants of obesity

Obesity is a consequence of a prolonged energy imbalance related to dietary intake and physical activity. Dietary intake and physical activity can in turn be influenced by environmental factors [25].

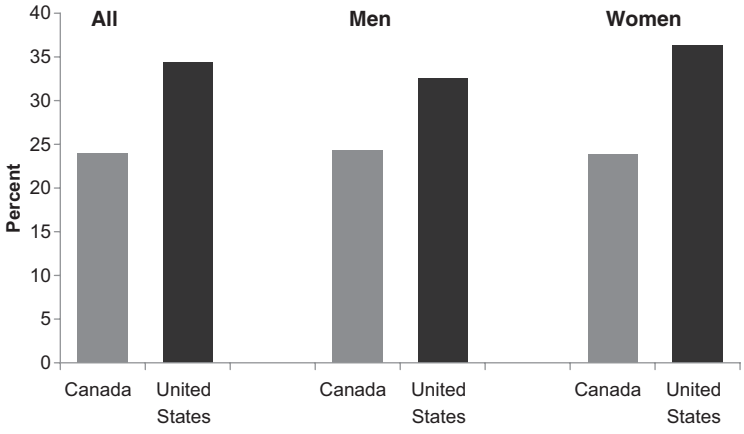
Dietary intake

The supply of dietary energy, measured in kilocalories (kcal), available to the US population has increased



Notes: Obesity defined as body mass index (BMI) ≥30. All estimates age-adjusted by the direct method to the 2000 US Census population using age groups 20–39, 40 –59, 60+ years. Significant increasing trend for all men (P<0.05).

Figure 35.5 Trends in prevalence of obesity, US adults 20+ years, 1988–1994 to 2007–2008.
Source: CDC/NCHS, National Health and Nutrition Examination Surveys, http://www.cdc.gov/nchs/nhanes.htm; Flegal et al. 2010 [4].



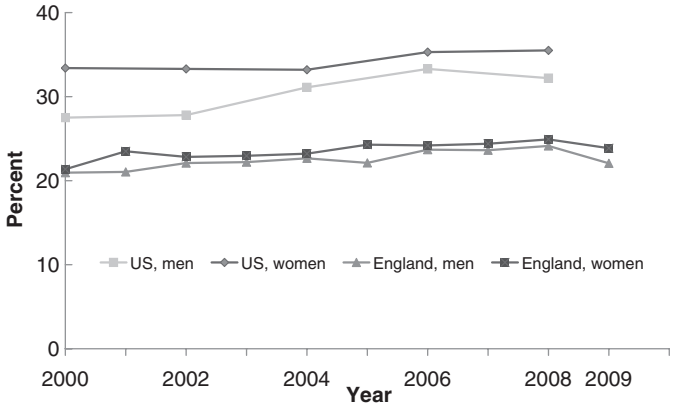
Notes: Estimates were age-standardized by the direct method to the 2000 United States Census population using age groups 20–39, 40 –59, 60 –79 years. Obesity defined as body mass index (BMI) ≥30.

Figure 35.6 Prevalence of obesity, by sex, adults aged 20–79 years, United States (2007–2008) and Canada (2007–2009).
Source: US: CDC/NCHS, National Health and Nutrition Examination Survey, http://www.cdc.gov/nchs/nhanes.htm; Canada: Canadian Health Measures Survey, http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5071&Item_Id=129548&lang=en; Shields et al. 2011 [23].

during the years that obesity prevalence increased in the USA. The US Department of Agriculture reports that the per capita daily food energy supply was 3200 kcal in 1980 and 3900 kcal in 2006 [26].

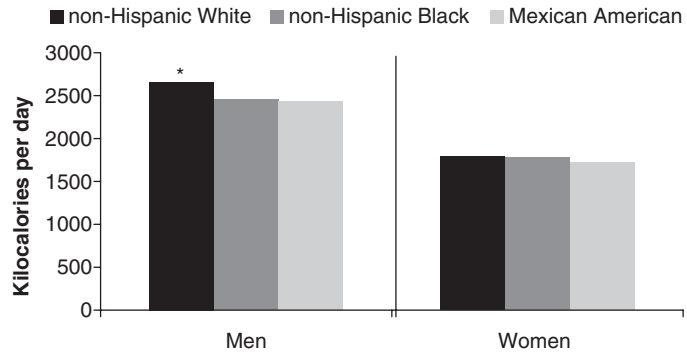
NHANES data from 2007–2008 on dietary intake, indicate that the mean energy intake for men and women in the USA was 2504 kcal and 1771 kcal [27]. Non-Hispanic White men had higher energy intake than non-Hispanic Black and Mexican American men; there were no differences by race/ethnicity among women (Figure 35.8). As a percentage of total energy intake, carbohydrates comprise approximately 50 % of the macronutrients consumed by adults (Figure 35.9). Among both men and women, total fat comprised approximately one third of calories and saturated fat comprised approximately 11 %. The Dietary Guidelines for Americans recommend that total fat comprise 20–35 % of energy and saturated fat no more than 10 % of energy in the diet of adults [28].

Secular trends between 1999–2000 and 2007–2008 suggest there were no statistically significant trends in total energy intake (Figure 35.10) [27]. However, over this time period, there was a decrease in average carbohydrate intake and an increase in saturated fat among non-Hispanic Black men and non-Hispanic White women [27].



Notes: US estimates age-adjusted by the direct method to the year 2000 US Bureau of the Census estimates using the age groups 20–39, 40–59, 60+ years. Obesity defined as body mass index (BMI) ≥30.

Figure 35.7 Trends in prevalence of obesity among adults, United States and England.
Source: US: CDC/NCHS, National Health and Nutrition Examination Surveys, http://www.cdc.gov/nchs/nhanes.htm; Flegal et al. 2010 [4]; England: Health Survey for England 2009, http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles-related-surveys/health-survey-for-england/health-survey-for-england–2009-trend-tables.



Notes: All estimates age-adjusted by the direct method to the 2000 US Census population using age groups 20–39, 40 –59, 60+ years.
*Significantly different from non -Hispanic Black and Mexican American.

Figure 35.8 Mean total daily kilocalorie intake, by sex and race/ethnicity, US adults 20+ years, 2007–2008.
Source: CDC/NCHS, National Health and Nutrition Examination Survey, http://www.cdc.gov/nchs/nhanes.htm; Wright & Wang 2010 [27].

Physical activity

Physical activity data from the USA show that between 1950 and 2000 adults became less likely to walk to work or have jobs that require high levels of physical activity [29]. On the other hand, trends in leisure time physical inactivity based on self-reported data show a decrease in inactivity from about 30 % to about 25 % between 1988 and 2008 [30]. Significant age differences in physical activity levels exist. Data from the USA based on measurements from accelerometers show that physical activity decreases with age and women have lower levels than men [31]. Other countries have also reported declines in physical activity related to employment [32,33].

The environment

A growing body of evidence is implicating environmental factors in promoting overconsumption of nutrient-poor foods, discouraging the intake of nutrient-rich foods and hindering adequate levels of physical activity [25,34].

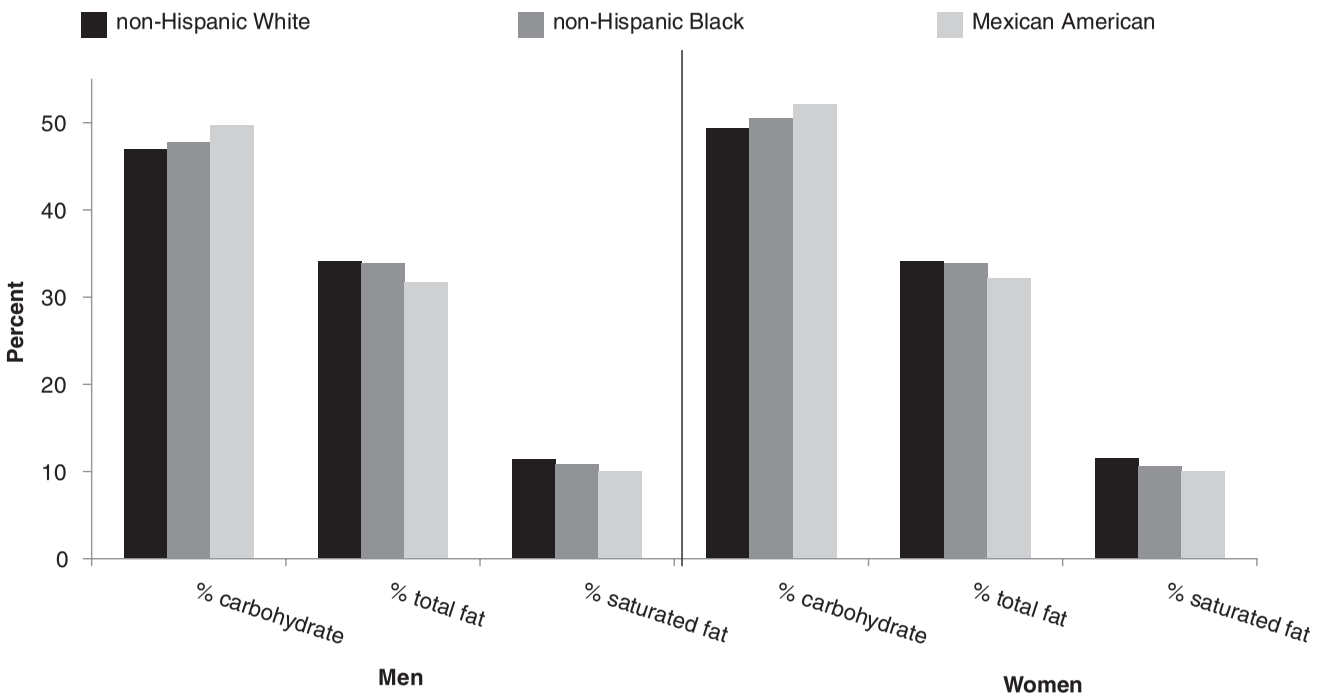
Most of the evidence supporting an association between the food environment and obesity comes from cross-sectional studies [34]. Residential proximity to sources of affordable nutrient-rich foods may be protective against obesity [25]. In one study, the presence of chain supermarkets, which have a larger selection of healthy food items at low prices as compared to small individually owned grocery stores [35], was associated with a lower prevalence of obesity [36]. Conversely, increased fast-food density has been positively associated with weight status [25,37]. However, differences are small and some studies have reported no association between residential proximity to fast-food restaurants and obesity [38–40]. Lack of methodological homogeneity and the absence of clear metrics to define the food environment hinder cross-study comparisons and may confound our understanding of the role of environmental factors in obesity [34].

The built environment can promote or restrict physical activity [25,41,42]. The availability of accessible recreational spaces and walkable built environments that are perceived to be safe has been correlated with an increased level of physical activity and low prevalence of obesity [25,43]. However, the evidence remains inconclusive with some studies reporting no significant associations [34]. As is the case with studies of the food environment and obesity, it is difficult to establish causality from observational studies of the built environment and its effect on physical activity and obesity.

Consequences of obesity

Obesity is associated with a variety of health conditions including hypertension, hypercholesterolemia, diabetes, cancer, and NAFLD [9,44]. Premature mortality can also be a consequence of obesity [45].

Hypertension is a risk factor for the development of a range of diseases including coronary heart, cerebrovascular, and renal diseases [46]. Prevalence of hypertension in the USA increased approximately 5 % between 1988–1994 and 2005–2008 (25.5 % to 30.9 %) [47]. As a point of reference, prevalence of



Note: All estimates age-adjusted by the direct method to the 2000 US Census population using age groups 20–39, 40 –59, 60+ years.

Figure 35.9 Mean daily percent of kilocalories from carbohydrate, total fat, and saturated fat by sex and race/ethnicity, US adults, 2007–2008. Source: CDC/NCHS, National Health and Nutrition Examination Survey, http://www.cdc.gov/nchs/nhanes.htm; Wright & Wang 2010 [27].

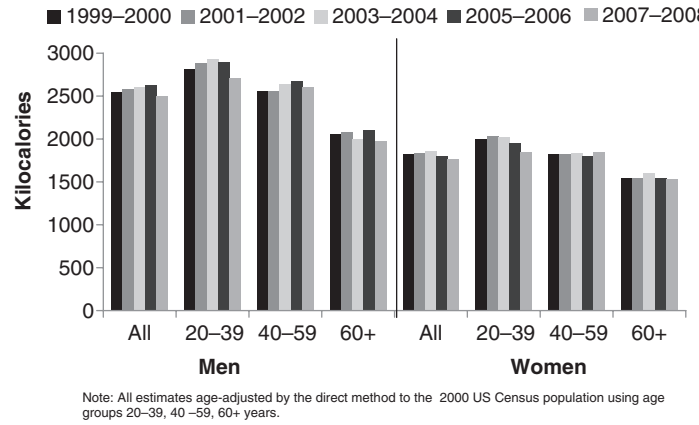


Figure 35.10 Mean total daily calorie intake, US adults 20+ years, 1988–1994 to 2007–2008.
Source: CDC/NCHS, National Health and Nutrition Examination Surveys, http://www.cdc.gov/nchs/nhanes.htm; Wright & Wang 2010 [27].

obesity increased approximately 13 % and 10 % for men and women between 1988–1994 and 2007–2008 [47]. Adults, and particularly younger adults, who are obese have higher odds of hypertension than those who are normal weight [48].

Adverse serum lipid concentrations, including elevated low-density lipoprotein (LDL) and triglycerides, and low high-density lipoproteins (HDL) cholesterol, contribute to the process of artherosclerosis [49]. Hypercholesterolemia among adults *increased* between 1988–1994 and 2005–2008 from 22.8 % to 27.5 % [47]. Obese persons, and especially younger obese persons, are more likely to have adverse serum lipid concentrations when compared with those of a normal weight [48].

Diabetes is more common among those with a higher weight [50]. The age-adjusted prevalence of physician-diagnosed diabetes was 5.5 % in 1988–1994 and 7.9 % in 2005–2008 [47]. Previous analyses of trends in diabetes prevalence from the 1970s to the early 2000s had suggested the increase in diabetes prevalence during this time (5.1 % to 8.8 %) was disproportionally comprised of individuals with BMI $\geq$35 [51].

Obesity has been associated with increased risk of diseases such as NAFLD and certain cancers including uterine, kidney, gallbladder, breast (postmenopausal women), esophageal, and colon [9,52]. Some evidence suggests that there may be obesity-induced hormonal changes [53], including sex steroids, insulin, and insulin-like growth factors that could lead to these cancers [54]. Population-based prevalence estimates of NAFLD are not available. However, one review article estimates its prevalence between 3–24 % [55]. The prevalence of NAFLD is higher among obese adults than normal weight adults [44].

Obese adults have higher all-cause mortality than normal weight adults, an association largely driven by increased mortality from cardiovascular disease (CVD) [45]. In addition to higher mortality from CVD, compared to normal weight adults, obese adults have higher cause-specific mortality attributed to diabetes/kidney diseases and cancers that have been epidemiologically linked to obesity [45]. However, obese and normal weight adults do not differ in mortality from cancers not epidemiologically associated with obesity or in mortality from noncancer, non-CVDs [45].

An emerging area of interest among both researchers and clinicians is the so-called "obesity paradox". The obesity paradox refers to the better prognosis among overweight and obese, compared to normal weight, adults with similar health conditions, including cardiovascular diseases [56]. The mechanism for improved prognosis is not yet clear but several theories have been cited including differences in nutritional reserves and origins of disease [57].

Conclusion

Obesity has been linked to increased morbidity and mortality in the United States and worldwide [2]. In 2008, an estimated 502 million adults globally were obese [58]. Obesity is estimated based on BMI, which is not a perfect measure of body fat. There are

significant differences in body fat by race/ethnicity within BMI categories [22]. Racial/ethnic disparities in the prevalence of obesity in the USA may, in part, reflect these differences in body fat. Per capita supply of food energy has increased while employment-related physical activity has decreased in recent decades. Moreover, the environment has been implicated in promoting the consumption of foods high in sugar and fat while hindering the ability to be physically active.

Multiple choice questions

1 Which method is considered the "gold standard" in determining body fatness?
- **A** BMI
- **B** Hydrodensitometry
- **C** Waist circumference
- **D** All of the above

2 Mean percentage body fat varies by race/ethnicity within BMI categories.
- **A** True
- **B** False

3 Obesity is a consequence of:
- **A** Increased energy intake
- **B** Physical inactivity
- **C** Environmental factors such as accessibility to healthy food options
- **D** All of the above

References

1 U.S. Department of Health and Human Services (2010) *The Surgeon General's Vision for a Healthy and Fit Nation*, U.S. Department of Health and Human Services, Office of the Surgeon General, Rockville, MD.

2 International Obesity Taskforce (IOTF) (2011). http://www.iaso.org/iotf/ (last accessed July 2103).

3 Flegal KM, Carroll MD, Ogden CL, Johnson CL. Prevalence and trends in obesity among US adults, 1999–2000. *JAMA* 2002;288(14):1723–7.

4 Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999–2008. *JAMA* 2010;303(3):235–41.

5 Must A, Spadano J, Coakley EH, et al. The disease burden associated with overweight and obesity. *JAMA* 1999;282(16):1523–9.

6 Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. *JAMA* 2005;293(15):1861–7.

7 Jee SH, Sull JW, Park J, et al. Body-mass index and mortality in Korean men and women. *New Engl J Med* 2006;355(8):779–87.

8 Centers for Disease Control and Prevention/National Center for Health Statistics (CDC/NCHS). National Health and Nutrition Examination Survey, 2011; http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm (last accessed July 2013).

9 National Institutes of Health (1998) Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults – The Evidence Report. *Obes Res* 6(Suppl 2). 1998/11/14 ed; http://www.nhlbi.nih.gov/guidelines/obesity/ob_gdlns.pdf (last accessed July 2013).

10 Ellis KJ. Selected body composition methods can be used in field studies. *J Nutr* 2001;131(5):1589S–95S.

11 Lohman TG CZ. (2005) Dual-energy X-ray absorptiometry, in *Human Body Composition*, 2nd edn (eds. Heymsfield SB LT, Wang Z, Going SB), Human Kinetics Books, Champaign, IL, p. xii.

12 Gorber SC, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev* 2007;8(4):307–26.

13 World Health Organization (1995) *Physical Status: The Use and Interpretation of Anthropometry*, WHO, Geneva.

14 Wang Y, Rimm EB, Stampfer MJ, et al. Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. *Am J Clin Nutr* 2005;81(3):555–63.

15 Yusuf S, Hawken S, Ounpuu S, et al. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet* 2005;366(9497):1640–9.

16 Flegal KM, Graubard BI. Estimates of excess deaths associated with body mass index and other anthropometric variables. *Am J Clin Nutr* 2009.

17 Vazquez G, Duval S, Jacobs DR, Jr., Silventoinen K. Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiol Rev* 2007;29:115–28.

18 Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-94. Series 1: programs and collection procedures. *Vital Health Stat 1* 1994;32:1–407.

19 Lohman TG, Roche AF, Martorell R (1988) *Anthropometric Standardization Reference Manual*, Human Kinetics Books, Champaign, IL.

20 Matsushita Y, Tomita K, Yokoyama T, Mizoue T. Optimal waist circumference measurement site for assessing the metabolic syndrome. *Diabetes Care* 2009;32(6):e70.
21 Ford ES, Li C, Zhao G, Tsai J. Trends in obesity and abdominal obesity among adults in the United States from 1999–2008. *Int J Obes (Lond)* 2011;35(5):736–43.
22 Li C, Ford ES, Zhao G, et al. Estimates of body composition with dual-energy X-ray absorptiometry in adults. *Am J Clin Nutr* 2009;90(6):1457–65.
23 Shields M, Carroll MD, Ogden CL. Adult obesity prevalence in Canada and the United States. *NCHS Data Brief* 2011;56:1–8. Available at: http://www.cdc.gov/nchs/data/databriefs/db56.htm (accessed March, 2011).
24 National Health Service. Health Survey for England – 2009: Trend tables 2011.
25 Popkin BM, Duffey K, Gordon-Larsen P. Environmental influences on food choice, physical activity and energy balance. *Physiol Behav* 2005;86(5):603–13.
26 U.S. Department of Agriculture ERS. Nutrient Availability; http://www.ers.usda.gov/Data/FoodConsumption/NutrientAvailIndex.htm (last accessed July 2013).
27 Wright JD, Wang C-Y. Trends in intake of energy and macronutrients in adults from 1999–2000 through 2007–2008. *NCHS Data Brief* 2010;49:1–8. Available at: http://www.cdc.gov/nchs/data/databriefs/db49.htm (accessed November 2010).
28 U.S. Department of Agriculture (2010) *Dietary Guidelines for Americans, 2010*. Vol. 7, U.S. Government Printing Office, Washington, DC.
29 Brownson RC, Boehmer TK, Luke DA. Declining rates of physical activity in the United States: what are the contributors? *Ann Rev Public Health* 2005;26:421–43.
30 Centers for Disease Control and Prevention (2008) 1988–2008 No Leisure-Time Physical Activity Trend Chart; http://www.cdc.gov/nccdphp/dnpa/physical/stats/leisure_time.htm (last accessed July 2013).
31 Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* 2008;40(1):181–8.
32 Stamatakis E, Ekelund U, Wareham NJ. Temporal trends in physical activity in England: the Health Survey for England 1991 to 2004. *Prev Med* 2007;45(6):416–23.
33 Popkin BM. Will China's nutrition transition overwhelm its health care system and slow economic growth? *Health Aff (Millwood)* 2008;27(4):1064–76.
34 Feng J, Glass TA, Curriero FC, et al. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place* 2010;16(2):175–90.
35 Sallis JF, Nader PR, Rupp JW, Atkins CJ, Wilson WC. San Diego surveyed for heart-healthy foods and exercise facilities. *Public Health Report* 1986;101(2):216–19.
36 Morland K, Wing S, Diez Roux A. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *Am J Public Health* 2002;92(11):1761–7.
37 Mehta NK, Chang VW. Weight status and restaurant availability a multilevel analysis. *Am J Prev Med* 2008;34(2):127–33.
38 Lopez RP. Neighborhood risk factors for obesity. *Obesity (Silver Spring)* 2007;15(8):2111–19.
39 Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? *Int J Behav Nutr Phys Act* 2006;3:2.
40 Wang MC, Kim S, Gonzalez AA, et al. Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *J Epidemiol Community Health* 2007;61(6):491–8.
41 Saelens BE, Sallis JF, Frank LD. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Ann Behav Med* 2003;25(2):80–91.
42 Centers for Disease Control and Prevention (2011) The Guide to Community Preventive Services. Obesity Prevention and Control: Interventions in Community Settings; http://www.thecommunityguide.org/obesity/communitysettings.html (last accessed July 2013).
43 Frank LD, Saelens BE, Powell KE, Chapman JE. Stepping towards causation: do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Soc Sci Med* 2007;65(9):1898–914.
44 Argo CK, Caldwell SH. Epidemiology and natural history of non-alcoholic steatohepatitis. *Clin Liver Dis* 2009;13(4):511–31.
45 Flegal KM, Graubard BI, Williamson DF, Gail MH. Cause-specific excess deaths associated with underweight, overweight, and obesity. *JAMA* 2007;298(17):2028–37.
46 National High Blood Pressure Education Program (2004) *The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure*, National Institutes of Health, Bethesda, MD.
47 CDC/National Center for Health Statistics. Health, United States, 2010: With Special Feature on Death and Dying. Feb 2011.
48 Brown CD, Higgins M, Donato KA, et al. Body mass index and the prevalence of hypertension and dyslipidemia. *Obes Res* 2000;8(9):605–19.
49 Stamler J, Wentworth D, Neaton JD. Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? Findings in 356,222 primary screenees of the Multiple Risk Factor Intervention Trial (MRFIT). *JAMA* 1986;256(20):2823–8.

50 Hartz AJ, Rupley DC, Jr., Kalkhoff RD, Rimm AA. Relationship of obesity to diabetes: influence of obesity level and body fat distribution. *Preventive Medicine* 1983;12(2):351–7.
51 Gregg EW, Cheng YJ, Narayan KM, et al. The relative contributions of different levels of overweight and obesity to the increased prevalence of diabetes in the United States: 1976–2004. *Prev Med* 2007;45(5):348–52.
52 Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *New Engl J Med* 2003;348(17):1625–38.
53 The World Health Organization (2002) *The World Health Report 2002: Reducing Risks, Promoting Healthy Life*; http://www.who.int/whr/2002/en/whr02_en.pdf (last accessed May 13, 2013).
54 International Agency for Research on Cancer (1997) *IARC Handbooks of Cancer Prevention*, IARC Press, Lyon, France, p. v.
55 Clark JM. The epidemiology of nonalcoholic fatty liver disease in adults. *J Clin Gastroenterol* 2006;40(Suppl 1):S5–10.
56 Lavie CJ, De Schutter A, Patel D, et al. Body composition and coronary heart disease mortality: an obesity or a lean paradox? *Mayo Clin Proc* 2011;86(9):857–64.
57 Lavie CJ, Milani RV, Ventura HO. Obesity and the "obesity paradox" in cardiovascular diseases. *Clin Pharmacol Ther* 2011;90(1):23–5.
58 Finucane MM, Stevens GA, Cowan MJ, et al. National, regional, and global trends in body mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet* 2011;377(9765):557–67.

Answers to multiple choice questions

1. B
2. A
3. D

Index

Page numbers in *italics* denote figures, those in **bold** denote figures.

GI Epidemiology: Diseases and Clinical Methodology, Second Edition. Edited by Nicholas J. Talley et al.

Published 2014 by John Wiley & Sons, Ltd. Companion website: www.wiley.com/go/talley/giepidemiology