Jacob S. Siegel

The Demography and Epidemiology of Human Health and Aging

Springer

The Demography and Epidemiology of Human Health and Aging

Jacob S. Siegel

The Demography and Epidemiology of Human Health and Aging

In consultation with S. Jay Olshansky

Springer

Jacob S. Siegel
J. Stuart Siegel Demographic Services
Nicholson Lane 5809
20852-5710 North Bethesda
Maryland
Unit 808
USA
jssiegel@verizon.net

In Consultation with
S. Jay Olshansky
School of Public Health
University of Illinois at Chicago
West Taylor Street 1603
60186 Chicago Illinois
Room 885
USA
sjayo@uic.edu

ISBN 978-94-007-1314-7 e-ISBN 978-94-007-1315-4
DOI 10.1007/978-94-007-1315-4
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011929938

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Irv, Ronnie, and Dave and our parents

Preface and Acknowledgments

Health demography and its companion, the epidemiology of aging, are not well defined fields among the academic areas of study. A subdiscipline of demography labeled health demography has been identified and it essentially applies the materials, methods, and perspectives of demography to the study of health. Health demography overlaps greatly with the epidemiology of aging, and the distinction mainly reflects the difference in the academic training of the analyst. This book tries to set forth the basic materials of this area of study and to give it more formal definition and scope.

Health demography and the epidemiology of aging incorporate materials from a number of established disciplines, mainly sociology, demography, epidemiology, public health, gerontology, and actuarial science, but also bioethics, genetics, biostatistics, biology, medicine, geography, and computer science. Although the list above is long, the natural home of the subject is in departments of gerontology, demography, and epidemiology. The scope of the subject, as I envision it, is suggested by the Table of Contents of the book. It is apparent that I cover a broad area. The field ranges from the highly theoretical to the very practical. I deal with the theories regarding human aging and longevity, derived from molecular and evolutionary biology and biodemography, as well as with the issues involved in the expansion of hospital facilities in a city by local health administrators, derived from applied demography and spatial epidemiology. I consider the range of theories regarding the question of how long humans can live, and I consider the practical problems of how to bring demographic and epidemiological methods, materials, and perspectives to serve the health needs of local communities.

The book describes the latest sources of data on mortality and population health and presents the principal research findings based on these sources. It brings together the latest methods of measuring population health phenomena, including those for mortality, morbidity, and combinations of measures of mortality and morbidity. In this connection it considers such varied measures as the many types of life tables, the design of health surveys, and geographic information systems for health research and administration. While maintaining a primary focus on the demographic and epidemiological aspects of the subjects treated, the book covers

socioeconomic variations in mortality and health, reproductive health, the health determinants and consequences of migration, health policy and administration, and selected bioethical issues. The measures described are often illustrated with examples in the form of calculation paradigms, so that the reader has a model for replicating the measures, and I suggest software packages for implementing some of the more elaborate methods described.

The geographic scope of the text, the tabular material, and the references and suggested readings relate to both the national and international area. The health demography of the less developed countries is given a substantial share of attention, in deference to the size and distinctive character of their populations and the extent and complexity of their health problems. At the same time, primary attention is given to the population health of the industrialized countries, especially the United States, where most of the readers of this volume work or will work as professional demographers, gerontologists, or epidemiologists. If no other geographic reference is given, the reader may assume that the area referred to is the United States. Nevertheless, readers from other countries and students of international comparative population epidemiology will find much in it of interest for them.

Since this book is intended to be comprehensive and interdisciplinary, it can be used as a text for courses taught in a broad range of scientific disciplines relating to population health and the epidemiology of aging, and as a reference work and handbook for demographers, epidemiologists, gerontologists, and public health specialists who work in this area. The book is offered as a text for advanced undergraduate and graduate students in medical sociology, the demography and epidemiology of aging, health demography, and social gerontology, and as supplementary reading in the instructional programs of these disciplines. It should serve the needs of the many instructors who prefer a single integrated comprehensive text in the area of health demography for use in their courses. There is sufficient material for a two-semester course, but with selective omissions, it can be adapted for use in a one-semester course. For shorter and less advanced study of the subject, such sections of the book may be omitted as those on mortality models in Chap. 3, multistate life tables in Chap. 8, and the methods of preparing local population estimates in Chap. 16, and Chaps. 12 and 17.

Inasmuch as human aging/longevity is the subject of study, the reader may look for material on the characteristics of various health conditions, the risks of succumbing to them, and ways of treating them, even specific recommendations for how to overcome them. This book *is* concerned with the prevalence of diseases and the risks associated with various health conditions, but it is not a how-to-do-it handbook for treating specific illnesses or for achieving a long life. It does, however, describe the research relating to longevity/aging and so implicitly it provides general guidelines for extending human longevity.

Readers may note the absence of measures of reliability or statistical significance. Although much is said qualitatively about the accuracy of the underlying data, discussion and presentation of measures of statistical significance are omitted, partly in the interests of simplifying the writing and tabular presentation. Computer software for calculating many of the measures in this book include the associated measures of significance.

It is always a matter of great interest and concern to the prospective reader to ascertain what prior technical preparation is required to take most advantage of a book of this kind. Training in some areas is essential, training in others in advisable, and training in still other areas is helpful. Training is essential in the fundamentals of statistics, and in mathematics up to the equivalent of a basic course in mathematical analysis. Courses in demographic techniques and social research methods are advisable but not essential, and a basic course in epidemiology, biostatistics, or medical sociology is helpful. A course in biostatistics provides a strong background for some of the topics treated here. Much of the material "needed" is covered in this very book, however, since it is intended to be a relatively free-standing and self-contained text.

While I have tried to make this book rather comprehensive in its field, it provides only an introduction to many topics. Other related works that will extend the knowledge of the reader on topics basic to the subject matter of this volume include:

Jacob S. Siegel & David A. Swanson (Eds.), *Methods and materials of demography* (2nd ed. Elsevier/Academic Press, 2004): Provides a comprehensive discussion of the basic techniques of demographic analysis.

Richard G. Rogers, Robert A. Hummer, & Charles B. Nam, *Living and dying in the United States* (Academic Press, 2000): Gives a comprehensive presentation of the social and economic differentials in adult mortality in the United States at the end of the 20th century.

S. Jay Olshansky and Bruce A. Carnes, *The Quest for Immortality* (W.W. Norton and Company, 2001): Presents a highly readable discussion of the facts and fictions regarding longevity.

Louis G. Pol and Richard K. Thomas, *The Demography of Health and Health Care* (Second Edition, Kluwer Academic Publishers, Norwall, MA, 2000): Focuses on applied health demography and provides additional material on health geodemography.

Leon Gordis, *Epidemiology* (Third Edition, Philadelphia: Elsevier Saunders, 2004).

Stanton A. Glantz *Primer of Biostatistics* (Sixth Edition. New York: McGraw-Hill, 2007).

These books can be referred to for further discussion of the tools of demography, the methods of epidemiological analysis, and the basic issues of "longevity science." I cover many of the same topics, but bring them together under one cover as an integrated, continuous narrative, even if in less detailed fashion.

In the interest of expediting the publication of the book, I drew on some material that I previously authored and that was published by Elsevier/Academic Press, to whom I express my thanks. The publications of the U.S. Census Bureau, the Population Reference Bureau, the World Bank, the World Health Organization, and the United Nations have been especially helpful as documentation for many of the topics covered in this book and I gratefully acknowledge my use of them also.

In the preparation of this work, I was fortunate in being able to consult with S. Jay Olshansky, Professor of Epidemiology and Biostatistics, School of Public

Health, University of Illinois at Chicago. He would have been a co-author of the book, had his schedule in the last few years permitted it. As it turned out, I planned and prepared the initial version of the entire work, Dr. Olshansky then reviewed and commented on this draft, and I completed the text for publication. I have full responsibility for the form and content of the text submitted to the publisher. I owe Dr. Olshansky a debt of gratitude for his willingness to contribute his expertise in making the book both more authoritative and more useful for students of public health and gerontology.

Professor Murray Gendell, my former colleague at Georgetown University, was helpful in reviewing Chap. 15, Health Policy, and, especially, in identifying and interpreting recent sources of material in this area. Dr. Hazel Denton, of Johns Hopkins University, very graciously agreed to review Chap. 9, Reproductive Health, an area of her special expertise. Both Dr. Gendell and Dr. Denton provided many valuable comments on these particular chapters that contributed to refining the text.

Professor David Swanson, of the University of California, Riverside, kindly provided illustrative material on litigation between hospitals for Chap. 16. Cheryl Siegel Boyd, of Jefferson Medical University, provided practical insights into the bioethical issues of organ transplantation. Rose V. Siegel, formerly a health statistician with the National Center for Health Statistics and the U.S. Navy, prepared the initial draft of the subject index of the book, which I then edited and completed for publication. Trisha Moslin, of the Population Reference Bureau, collaborated in preparing a number of the figures.

Writing a book of this type is possible only because many scholars and analysts came before, conducted their research, and shared their specialized knowledge with others in their publications, research papers, and lectures. I profited greatly from exchanges with current contributors to the field of health demography, biodemography, biogerontology, and population epidemiology as I proceeded in my writing. In particular, I took special advantage of the writings, seminar presentations, and scholarly notes of Mark Hayward of the University of Texas, Teresa Seaman of the University of California at Los Angeles, Maxine Weinstein of Georgetown University, Jane Menken of the University of Colorado, James Vaupel of the Max Planck Research Institute, Bruce Carnes of the University of Oklahoma, Tommy Bengsston of the University of Lund, Kathleen O'Connor of the University of Washington, Leonard Hayflick of the University of California at San Francisco, Eileen Crimmins of the University of Southern California, Robert Binstock of Case Western Reserve University, and Jean-Marie Robine of the University of Montpellier, and many others.

I also want to acknowledge the contribution of the following persons in the U.S. Federal government, who provided either advice or materials: Robert Anderson, Elizabeth Arias, Lois Fingerhut, Ellen Kramerow, Michael Molla, and Harry Rosenberg of the National Center for Health Statistics; Charles Heller of the Agency for International Development; Howard Iams of the Social Security Administration; Celia Bortlein, Frederick Hollmann, Dean Judson, and David Waddington of the Census Bureau; Anne Martin of the Centers for Medicare and Medicaid; and Mitra Toosi of the Bureau of Labor Statistics. Several analysts in the National Institutes

of Health were kind enough to discuss special topics with me: Rebecca Clark of the National Institute of Child Health and Human Development; John Haaga and Donald K. Ingram of the National Institute on Aging; and Catherine Norton of the National Institute of Mental Health. Anthony Mann of the National Academy of Sciences provided guidance as to relevant publications from his organization on a variety of subjects. Keith Horvath, Robert Ledermann, and Annette Stine of the National Heart, Lung, and Blood Institute played the dual role of informing me of current research in cardiovascular medicine while incidentally attending to my personal health when I served as a volunteer patient in the Institute.

I gratefully acknowledge the support and encouragement of J. Scott Bentley, Senior Executive Editor, Elsevier/Academic Press, who oversaw the project from its beginning to its mid-stage. In its later stages, Evelien Bakker, Senior Publishing Editor, Springer, with the able assistance of Bernadette Deelen-Mans guided the work to publication with grace and efficiency.

Finally, I offer the usual disclaimer to exempt the many persons and organizations cited above from responsibility for any errors of omission or commission in this book and I accept full responsibility for its contents. It is *de trop* to add that this freedom from responsibility applies to my former longtime employers, the U.S. Census Bureau and Georgetown University, and to my several short-term academic employers around the country. Professor Olshansky and the School of Public Health, University of Illinois at Chicago, are also free of any responsibility for the contents of this book, including any opinions offered. At the same time, I wish to express my gratitude to all these parties for the opportunity of working with them or under their auspices and for advancing my knowledge of the materials that are the subject of this book.

Jacob S. Siegel

Contents

Part I
Scope of Health Demography, Sources, and Quality of Data

Chapter 1
Introduction

In including the names of two academic disciplines in the title of this book, demography and epidemiology, I have admittedly carved out a rather broad area for treatment, but hybrid and joint disciplines are common in science today. I am calling this hybrid health demography. Health demography has emerged as a subdiscipline of demography in recent decades and it essentially applies the materials, methods, and perspectives of demography to the study of health. At the same time, a subdiscipline of epidemiology has emerged labeled population epidemiology that essentially applies the materials, methods, and perspectives of epidemiology to the study of population. These two subdisciplines overlap to a large extent, and any distinctions mainly reflect the differences in the academic training of the analysts. With its focus on health demography, this book deals with the intersection of these subfields of demography and epidemiology.

Outline of the Book

The book is divided into five parts, each consisting of two or more of the 17 chapters included in the book and grouping together roughly related subjects. Part I presents basic conceptual material relating to health demography/population epidemiology (Chap. 1) and a comprehensive review of sources of data on health and mortality, including general discussions of the basic sources of data, collection methods, and limitations of the data (Chap. 2). Part II comprises several chapters concerned with the concepts of mortality and morbidity (Chaps. 3, 4, and 5), ways of measuring their levels, patterns, and trends (Chaps. 3–5 and 8), a description of U.S. and international mortality and morbidity experience (Chap. 6), an enumeration of the risk factors in disease and death (Chap. 6), and an analysis of group differences in mortality and morbidity (Chap. 7). Chapter 3, the first chapter in Part II, is devoted wholly to mortality, describing a range of measures of mortality and the

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_1,

methods of reformulating them for comparison and analysis. Chapter 4 carries the treatment of mortality measurement further with a discussion of the life table and its attributes as a model paradigm for human survival over a lifetime. The concepts of health status and functioning and their measurement are the subject of Chap. 5. Chapter 6 deals with national and international trends in mortality and morbidity and the principal risk factors associated with these events. Chapter 7 is concerned with differences in mortality and morbidity among demographic and socioeconomic groups, including the methods of measuring them and the principal findings relating to them. The last chapter in this part, Chap. 8, describes measures that link mortality and morbidity, representing extensions of life tables to include health and disability.

The group of chapters in Part III is loosely associated by their concern for the relation between the components of population change (i.e., fertility, mortality, and migration) and population age structure, on the one hand, and health, on the other. Chapter 9, entitled "Reproductive Health," is the chapter on fertility; Chap. 10 is the chapter on migration; Chap. 11 is a special chapter for mortality and morbidity in less developed regions of the world; and Chap. 12 is the chapter on population age structure, describing how mortality molds the age distribution of a population and how the mortality and health characteristics of a population are influenced by its age structure. In addition to describing various measures of population aging, Chap. 12 discusses the demographic causes and consequences of population aging, the construction of models of mortality, and the derivation of population age structures from models of mortality.

Part IV contains two chapters, the first being concerned with concepts and measures of longevity and descriptions of long-lived populations (Chap. 13), and the second (Chap. 14) with theories of aging and methods of making projections of mortality and health status. The final part, Part V, considers the areas of health policy and health administration. Chapter 15 deals with health policy at the national and international levels, both in the more developed and the less developed countries. Chapter 16 looks at the health situation at the local level, and so is concerned with methods of developing local population and health estimates, applying geographic information systems, and planning for public health emergencies. Chapter 17 supplements the earlier discussion on health policy and administration by considering some ethical and legal aspects of health issues involving demographic elements. The book closes with several appendices that include notes on several general analytic devices used in health demography, lists of journals and internet sources providing material on population health, a glossary of terms and phrases, and lists of countries corresponding to statistical rubrics employed by the United Nations and other international organizations.

Definition and Scope of Health Demography and Related Fields

Demography and Epidemiology

Central to the discussion throughout this book are the scope of demography and the scope of epidemiology. The boundaries of these disciplines can be defined only loosely, but clear areas of common interest can be identified. Later in the chapter I explore the general meaning of the other two terms in the title of the book – health and aging.

Demography may be defined as the study of the size, structure, distribution, and composition of populations, the principal determinants of changes in these parameters, namely, fertility, mortality, and migration, and the determinants and consequences of changes in these factors of population change. As such, a central area of concern of demography is the survival patterns of populations, the factors contributing to the different survival patterns of populations and groups within them, and the social and economic consequences of these survival patterns. In recent decades demographers have extended their interests to encompass health as a determinant of mortality, a factor in population growth, and a population characteristic, in addition to their earlier interests in it as a factor influencing fertility and migration (Lamb and Siegel in Siegel and Swanson 2004).

Traditionally, epidemiology has been concerned with the study of diseases in populations, including their causes, distribution, incidence and prevalence, and modes of transmission and control. In recent decades the scope of epidemiology has been extended to encompass other phenomena besides diseases, in particular aging and the age patterns of mortality, and fertility and family planning – a field denominated population epidemiology (Omran 1974). It is evident that the two fields, demography and epidemiology, overlap to a fair degree. That overlap is now particularly evident in their mutual interest in disease, mortality, and aging. Two special subfields of the disciplines of demography and epidemiology, in particular health demography and population epidemiology, represent the main areas of common interest.

Health Demography, Population Epidemiology, and Related Fields

The field of health demography is relatively new, but reflections of the joint interests of demographers and health scientists are quite old. In fact, health demography may be said to have begun with the initial efforts by John Graunt in 1662 to analyze vital statistics and fashion a life table from burial records. Mortality has from the beginning been an essential part of the subject matter of demography. Demographers' regular forays into health began in the latter part of the twentieth century, as suggested by their many publications on family planning, reproductive health, maternal and child health, health policies and programs, emerging diseases,

and tables of healthy life. Recent societal events have been instrumental in their further involvement with health themes, including the aging of the population, the spread of HIV/AIDS, the dramatic extension of life expectancy at the higher ages, and the need to redesign national social security, disability, and health insurance systems (Martin and Preston 1994). Health is a common theme in all these events or programs. The aging of the population will inevitably bring with it a vast increase in the number of persons who are afflicted with major chronic diseases and severe disabilities. HIV/AIDS is not only a health issue, but an issue that is closely related to population growth, fertility behavior, and family structure. The planning of social security and disability, Medicare, and other age entitlement programs depends heavily on projections of population, mortality, and health status.

The study of health by demographers has usually been an appendage of the study of mortality, but health is a population characteristic worthy of demographic study in its own right. Questions on health now appear often on population censuses and are used to classify groups in the population as disabled/non-disabled or other health classes. Some health conditions are not closely linked to death but are only loosely related to it. Yet, they are important elements in the frequency, distribution, and severity of disability and dependency, and the need for long-term care. These include such conditions as sensory impairments (e.g., vision and hearing loss), musculoskeletal conditions (e.g., arthritis, osteoporosis), urinary problems (e.g. incontinence), and neuropsychiatric disorders (e.g., depression). While health can be considered as a primary factor affecting mortality, it too is influenced by demographic and socioeconomic variables, such as the age and sex structure of the population, marital status, income status, and educational level, and has demographic and socioeconomic consequences, particularly with respect to fertility, migration, and marital status. Clearly health is an appropriate subject of demographic inquiry.

Epidemiology has always been concerned with the study of diseases in populations, including their distribution, incidence and prevalence, modes of transmission, and modes of control. It is distinguished by its risk-factor approach and its use of statistical methods and comparative geographic analysis. Population epidemiology extends epidemiology to the application of epidemiological methods to population subjects such as fertility, migration, population growth, and age structure, including aging (Omran 1974). This very last area virtually corresponds to the subdiscipline of demography called the demography of population aging and now an area of considerable interest to many demographers (Siegel 1980, 1993).

Health Demography and Biodemography

For the most part, demographers have looked to other demographic and socioeconomic factors to explain the demographic attributes of a population. They have generally tended to neglect biological influences, although there is a history of demographers invoking biological explanations to account for some attributes of fertility and mortality. In incorporating health as a demographic field and broadening

its prior studies of aging and longevity, demography has of late enlisted biology directly to provide insights not gained from socioeconomic and demographic factors alone. The modern rise of a subfield of demography known as biodemography was delineated in the last few decades of the last century as an explicit effort to add biology to the toolkit for explaining attributes of health, aging, and longevity in humans as well as other species (Keyfitz 1984; Olshansky and Carnes 1994). Olshansky et al. (2005) have defined biodemography narrowly in relation to aging as the scientific study of common age patterns and causes of death observed among humans and other sexually reproducing species and the biological forces that contribute to them. Most demographers and gerontologists identifying themselves as biodemographers define the field much more broadly as biologically informed demography or as the interface of biology and demography (Wachter and Finch 1997; Myers and Eggers 1996). Accordingly, it deals with demographic questions partly through enlisting an array of biological indicators, or biomarkers, to explain individual and group differences in fertility, mortality, migration, aging, and longevity, and partly through a comparative analysis of these demographic factors and characteristics in humans and subhuman species (Crimmins and Seaman 2000; Carey 1993).

Several main questions are addressed in biodemography's concern with aging and longevity: Is there an upper limit to the life span of humans and other species? What biological factors influence life expectancy and life span? Is there a common pattern to the expression of disease and aging between humans and other species? Is aging a disease? These questions must be considered in biological as well as demographic terms since evolution may have set constraints on the extension of the human life span, and biological analysis of other species may suggest insights into how biomedical interventions may be effective in expanding the horizon on the length of human life. The conjuncture of biology and demography can provide a more profound understanding of demographic variations and processes, offering indications as to which variations and changes are common to living things and which are uniquely characteristic of humans (Wachter and Bulatao 2003).

Health Demography, Demography of Aging, and Medical Sociology

Health demography bears a close resemblance to several other disciplines and subdisciplines, particularly the demography of aging and medical sociology, in addition to overlapping greatly with biodemography and population epidemiology. The domains of these fields cannot always be distinguished, however. The discipline of epidemiology has been broadened to include the statistical study of various social and demographic phenomena, including their geographic and temporal variations. For example, the epidemiology of aging focuses on geographic and temporal variations in the characteristics of the elderly, including their health conditions. Thus, health demography and the epidemiology of aging share many areas (as is suggested by the etymology of the terms for the two disciplines), but these common areas may receive a different emphasis. In recent years there has been an interest in

promoting a dialogue between the two fields, particularly with respect to the study of aging, as evidenced by publications of the National Academy of Sciences and other scientific organizations (Weinstein et al. 2001) and by the proposed support for research and training linking the two fields by the National Institutes of Health (PAR-05-134).

Medical sociology also overlaps with health demography and population epidemiology. According to Zimmerman (2000: 1813), medical sociology embraces:

1. social epidemiology, or the socioeconomic, demographic, and behavioral factors in the etiology of disease and mortality;
2. study of the development and organizational dynamics of health occupations, hospitals, and other health maintenance organizations, including, for example, relations of doctor and patient;
3. the social construction of the illness experience by societies and individuals; and
4. social policies, social movements, and political and economic conditions that shape and are shaped by health and disease.

Among these areas of interest of medical sociology, item (1) may be considered as falling most clearly within the scope of health demography and population epidemiology. The relationship of items (2), (3), and (4) to health demography and epidemiology is much weaker and only tangentially related to issued addressed in this book.

Health Demography and Biostatistics, Actuarial Science, and Medical Demography

Health demography bears a close relationship to several other sciences that are primarily mathematical and statistical. They are biostatistics, actuarial science, and medical demography. Biostatistics and actuarial science focus mainly on the applied aspects of mathematical and statistical analysis of health and mortality, the former being mostly concerned with diseases and the development of treatments, and the latter being mostly concerned with the age patterns of deaths, prediction of group deaths, and their costs. Epidemiology and biostatistics are closely related. Biostatisticians have collaborated in the design of leading epidemiological surveys, such as the Framingham Study, and they have become increasingly involved in their conduct and interpretation. In fact, the distinction between applied biostatistics and theoretical epidemiology has become blurred. Actuarial science is the ultimate applied statistical science since virtually all actuaries work for insurance companies or the grand "insurance" government agencies, SSA and CMS. Actuaries develop actuarial tables based on official life tables to determine the periodic funds to be paid to annuitants or their survivors as well as the amounts collected to sustain the viability of the funds.

Medical demography, the last of the fields I distinguish here, is the study of chronic disease, disability, and mortality among the older population and of the processes related to age that affect disease. According to Manton and Stallard (1994),

who wrote the most comprehensive statement on its scope and methods, medical demography has roots in biostatistical and actuarial models of mortality and the insurance of health-contingent events. They note that "medical demography requires biomedically detailed models of the relationship of age to health, to change in the ability to function, and to mortality in individuals." The medical demographer exploits vital statistics, longitudinal data, and the results of experimental designs and randomized trials to investigate the effects of changing age on health, and "relies on multivariate, stochastic models to represent the interaction of disease, disability, and mortality as they change over time." The field of medical demography differs from the epidemiology of chronic disease in that the latter tries to determine causal relations between risk factors and disease usually by employing general statistical models to measure the strength of the relations (Manton and Stallard 1994).

Concepts of Health, Health Conditions, and Longevity

Health is a leading characteristic of a population, like other demographic and socioeconomic characteristics, and an important correlate of the other demographic and socioeconomic characteristics. It is often used in interpreting trends in fertility, mortality, marital status, socioeconomic status, and household characteristics. It has often been the subject of inquiry in population censuses and national sample surveys. Health has been used to develop assumptions for making projections of mortality and population.

In recent decades, the field of population health has been extended beyond the measurement of mortality (e.g., life expectancy, infant mortality) to include specific health measures such as the incidence and prevalence of chronic health conditions, and newly-developed summary measures of health, such as healthy life expectancy and health-related quality of life. This extension of population health is due in part to the spectacular success in extending life expectancy in modern times. This success has shifted the focus of population health from the *quantity* of life to the *quality* of life. The World Health Organization (WHO) and various national governments and academic research organizations have worked to develop and promote summary measures of population health that combine measures of mortality and measures of health conditions into single more informative measures of a population's health. Summary measures of population health have several potential applications. These applications include comparing the burden of disease of one population with that of another or of the same population at different dates; comparing inequalities in the burden of disease between groups within populations; analyzing the relative effects of non-lethal health conditions and lethal health conditions on the overall burden of disease; determining the degree to which added years of life contribute further to disabled years; and aiding in the development of priorities for health service delivery and planning, informing decisions on priorities for research and development in the health sector, and analyzing the benefits of health interventions for use in cost-effectiveness analyses (Murray et al. 1999:3–4).

The WHO has promulgated a new global health policy that is designed to deal with the major current and emerging health problems in the world. The health plan, "Health for All in the twenty-first century" (HFA), aims to improve global health and reduce health inequalities within and between countries through the development of health priorities and targets for the first two decades of the twenty-first century. Member countries have been encouraged to set their own goals in accordance with HFA guidelines. In the United States, "Healthy People 2010" has been launched to meet two basic goals: To increase the years of healthy life (i.e., quality of years lived) and to eliminate health disparities (U.S. Dept. of Health and Human Services 2000a:2). The U.S. secondary goals are couched in terms of leading health indicators, including ones for health conditions, lifestyle habits, and access to health care.

The incidence and prevalence of morbid conditions, particularly chronic/degenerative diseases, the age at onset of the conditions, their patterns of dissemination, and their causes (i.e., genetic, environmental, behavioral, and stochastic) and effects (i.e., physical, psychological, demographic, social, and economic) have become increasingly important concerns in understanding population health. Health is a complex variable with many dimensions, as we shall see. Accordingly, there is a need for a variety of health statistics and methods of measurement and analysis to represent the occurrence and distribution of health conditions and to characterize population health.

Health and Health Conditions

It is difficult to define health because health is, in part, a physical state, in part a social state, and in part a psychological state. It is a multidimensional concept representing a condition that is a continuum and steadily changing for the individual. Often the determination as to whether a health condition exists is a matter of degree; the problem of making this determination exists whether health is measured objectively or subjectively. The complexities of defining health are particularly great in the case of disabilities and mental conditions.

Health has been defined both broadly (World Health Organization 1947; Mahler 1981) and narrowly (Dubos 1968). The World Health Organization gave health a comprehensive positive definition in its Constitution in 1947: "Health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity." In 1953 the President's Commission on the Health Needs of the Nation expressed this same general idea in describing a healthy individual as a "well-integrated individual both as to his physical structure and as to his physiological and psychological functioning." Echoing the WHO's and the President's Commission's concept of health, Halbert L. Dunn (1957) urged that medicine's goal should go beyond the cure of disease and even its prevention, and strive for "positive health," or "wellness." This goal represents a theoretical

ideal and little progress has been made toward it in the half century since it was promulgated. Breslow wrote in 1999 about moving "from disease prevention to disease promotion." American society is still short of striving actively even toward preventive health.

In spite of a long-time interest in defining health status in positive terms, for the most part health is still commonly conceived in terms of its inverse, that is, the absence of disease, impairment, or disability. Several general terms are employed to indicate a lack of good health. They include ill health, illness, and morbidity. A number of other terms, such as disease, sickness, health condition, and morbid condition, in addition to injury and impairment, refer to more specific health conditions. The definitions of several of these terms overlap and clear distinctions cannot easily be made between them. Morbidity is both a general term referring to sickness and a specific term referring to a health condition. Morbidity, morbid condition, and health condition encompass diseases, injuries, and impairments. Disease and illness both refer to a state of ill health, but disease emphasizes the biophysical state of ill health whereas illness emphasizes the social experience of being sick. This concept of disease is suggested by the medical definition of disease, which is an interruption, cessation, and disorder of body functions, systems, and organs (Stedman 2006).

For analytic purposes a distinction must also be drawn between disease, on the one hand, and injuries and impairments, on the other. Injuries are acute conditions involving abnormalities of body structure and appearance resulting from accidents or attacks. Impairments are chronic conditions involving an abnormality of psychological function, physiological or anatomical structure or function, or physical appearance, resulting from a disease or injury. The most common impairments are chronic sensory and musculoskeletal conditions. Another group of terms emphasize the physical and social consequences of ill health in terms of ability to function. The leading examples are disability and handicap. The term disability has many meanings, but in general a disability is a restriction or inability to perform an activity in the range considered normal for a person. A handicap is a limitation or inability to perform one's social role(s) due to an impairment and/or disability. It is useful to distinguish those disabilities that can be managed in part or whole by technological devices and structural changes in the community and those that cannot. Such a distinction can aid in defining degrees of disability, from slight to moderate to severe, inasmuch as the availability of adaptive devices affects the effective level of disability.

In another scheme classifying types of morbidity, a distinction is made between acute conditions (including acute illnesses and injuries) and chronic conditions (including chronic illnesses and impairments). The terms acute and chronic do not have precise definitions, but are distinguished on the basis of the duration and the type of health condition. An acute condition typically has a rapid onset, has a relatively short duration, and usually ends with either recovery or death. A chronic condition usually involves a lengthy period of development and progression, has a long duration, and is considered relatively intractable to treatment. Infectious

diseases, for example, are usually classified as acute while cardiovascular diseases are classified as chronic. These distinctions are not rigid, however. HIV/AIDS is an infectious disease that is chronic.

It is important to recognize that in reality people are not merely healthy or unhealthy. There are gradations of ill health, levels of disability, and degrees of impairment. Alternatively, there are levels of the healthy state. Most people who are described as healthy by their physicians are not nearly as healthy as they could be or should be when health status is viewed in positive terms. Health status is in constant flux and individuals are always moving on a trajectory toward better or poorer health, with the general direction being downward as they age. It is likely that the majority of older persons who are described as not in poor health as commonly understood are living at a low level of wellness and are beginning to experience the precursors of future illness. One of the tasks of health demography/population epidemiology is to devise improved systems for classifying persons in the various categories of "illness" or "wellness," to determine the numbers in these categories, and to measure changes and variations in them. Health concepts are considered in more detail in Chap. 5. The terminal phase of ill health is death. The concepts and measurement of death are discussed in Chaps. 3 and 4, and measures combining health and death are considered in Chap. 8.

Aging and Longevity

The term aging included in the title of the book has a number of meanings like health. It has demographic, sociological, psychological, and biological dimensions. For now let us say, it is the process of growing older along the four trajectories listed. The demographic and biological dimensions, in particular, are considered in some detail in later chapters. I explore mainly the nature and implications of chronological and physiological aging.

The term longevity requires little explanation. It is a general term for length of life and so it encompasses the two main terms used to refer to length of life, namely, life expectancy and life span. Longevity can refer to the (average) length of life prevailing in a population, as does the term life expectancy, or it can refer to the length of life of individuals, as does the term life span. We are familiar with such statements as that the life expectancy of a nation is 75 years and that an individual died after a life span of 73 years. In another usage the term life span refers to the longevity of a species, such as, that the human life span (e.g., the age beyond which less than 1% of the original cohort survives) is about 100. There are other variants of the concept of longevity that will be explicated in the course of the book. Note that the concept of longevity differs from the concept of aging, but is closely related to it, as will be explained in Chaps. 12 and 13.

Selected Issues Relating to Human Health and Longevity

As the reader makes his or her way through this book, it will become apparent that many issues in health demography are subjects of debate and contention among scholars and remain unresolved. Proponents of one side of an issue are matched by proponents on the other side, and the two sides may feel equally fervent about their positions. Then too there are the disagreements between the traditional health scientists and others viewed by the former as the "lunatic fringe."

For some, this situation may be frustrating but I hope that the reader will find it fascinating and intriguing that health demography is a science brimming with issues to be resolved. I identify some of these issues in the following paragraphs.

There are continuing issues with respect to the optimum ways of summarizing age-specific death rates and comparing the relative mortality and morbidity of different populations. For example, the traditional measure for summarizing and comparing lifetime mortality for populations, life expectancy at birth, is under fire, as described in Chap. 4. It is charged with over- or underestimating "real" life expectation in a year depending on the direction and rate of change in mortality. Other analysts see the proposed cure as imposing an obscure set of metaphysical assumptions on the data, and at best as another esoteric summary measure of annual mortality.

Other related issues include: Proposals for an alternative classification of the causes of death than the conventional WHO classification; differences in interpretations of cause-elimination life tables as simply analytic summaries of current mortality or possibly also as predictive tools adumbrating increases in life expectancy; the optimum ways of measuring socioeconomic inequalities in mortality and morbidity and the degree to which such inequalities can be reduced; and the anomaly of lower mortality among first-generation immigrants than natives.

An issue touching the very core of biogerontological thinking is whether aging is or is not a disease. There is general agreement that many serious diseases are age-related and that age is a risk factor in the onset of these diseases. To some, aging is not a disease for a variety of reasons, among them that some characteristics of aging are clearly not diseases (e.g., graying of hair) and that many of the diseases of later life occur to younger persons as well as to older persons, even if for different reasons. For another, less numerous and vocal group of health scientists, if advancing age is accompanied inevitably with increasing organic dysfunctionality and organic dysfunctionality is a characteristic of disease, then from a pragmatic point of view, aging should be viewed as a disease.

Many mainstream health scientists are concerned that this view – that aging is a disease – has also been adopted by the "anti-aging" movement, which has been using it to support its argument that the aging process can be reversed and that life span can be extended indefinitely by growth hormones, special health potions, and megadoses of vitamins. Traditional biodemographers and biogerontologists vigorously reject the claims of the anti-aging movement. They believe that the means proposed by the anti-aging movement cannot achieve the goals it claims

and that the means are not now available for achieving these goals. Its proposed "treatments" are viewed by mainstream scientists as fraudulent ways of influencing the public to buy into a false hope. Yet, the SENS movement (de Grey 2005), which believes that the same goals can be achieved by more conventional means, has been winning support from many traditional biogerontologists, as described in Chap. 13.

Biogerontologists and geriatricians strongly disagree on the most effective allocation of research funds in the health area between research on the basic processes of aging at the molecular level and research on specific chronic diseases, such as heart disease, cancer, diabetes, and Alzheimer's disease. The bulk of research funds are now allocated for specific diseases, but this choice runs counter to new knowledge about the possibly common molecular origins of the leading chronic diseases, the role of early life influences on disease in later life, and the role of chance in the onset of disease. Politically the latter view is difficult to sell and, hence, some analysts recommend a compromise and support both lines of effort at the same time. Fortunately, new lines of research into specific diseases are employing biogerontological approaches, such as gene and stem cell therapy, thus combining the two research philosophies and mollifying the intensity of the disagreement.

Various biodemographers hold very different views as to the possibilities for extending human longevity in the next half century, whether life expectancy (i.e., average years lived by a population) or the maximum observed life span (i.e., average of the ages attained by persons over age 100). Some hold tenaciously to the view that life expectancy can increase only a few more years in this period while others insist that a life expectancy of 100 is the future lot of birth cohorts now being born. As explained in Chap. 13, the arguments on both sides are cogent, the relative weights of the arguments on each side are difficult to assess and predict, and, therefore, the future must remain a mystery for now. The bulk of professional opinion is on the side of the optimists, but few of these supporters are really familiar with the issues. This controversy spills over into the methods and assumptions used in making projections of mortality and morbidity, as described in Chap. 14. Yet, while most of the projection analysts side with the optimists, they continue to make only conservative assumptions of mortality for their projections!

Most of the information regarding health maintenance and treatment of disease that has been amassed through research studies pertains to adults of working age, i.e., pre-longevous persons, as described in Chap. 17. Yet, a large and increasing share of the population is surviving to advanced ages. As a result, health scientists often disagree on the management and treatment regimens for the risk factors and illnesses common at the very old ages (e.g., hypertension/high blood pressure, hypercholesteremia/high cholesterol, overweight) and on the priorities of treatment in the face of multiple morbidities and the problems of maintaining physiological homeostasis (i.e., equilibrium of bodily systems). There is a pressing need to broaden the range of ages of research subjects and to conduct clinical trials that specifically address these issues for the advanced aged. Some movement in this direction is now beginning and light is being focused on a number of "centenarian" populations as well.

Advances in the possibilities of implementing the wider use of electronic records and advances in genetic and genomic knowledge raise difficult issues of privacy for patients. Yet there is widespread agreement that electronic recordkeeping would advance the standard of patient care and reduce iatrogenic (physician-caused) and nosocomial (hospital-caused) illness and death. There is less agreement on whether it would be more cost-efficient. As of now, only a small percentage of medical records are managed electronically.

Another less obvious but real disagreement relates to the concepts describing the optimum transition to the older years. A plethora of terms is being employed by writers in the field with sometimes little thought as to their comparative implications and, in the tradition of science, the need to convey a common meaning. The phrases, successful aging, positive aging, aging well, healthy aging, productive aging, "aging in stride," and so on, are employed with little discrimination, in a way unworthy of scientific discourse. This variety is not helpful to precise communication, and clear operational definitions should be given to the phrases used. My view is that "successful aging" – probably the most widely used expression in the field – is inappropriate for professional use, mainly because it is too value-laden or judgmental and makes unrealistic promises. "Healthy aging" appears to be a better choice as a standard term, but even this term has an aura of unrealism. This terminological question merits further attention and debate.

Finally, I mention the problems of making reliable local estimates of the prevalence of various diseases, the merits of using the results of group research with biomarkers in predicting individual health trajectories, and the seemingly intractable problems of reducing food insecurity and controlling the spread of endemic infectious diseases, including HIV/AIDS, in the Less Developed Countries. Many more issues could be identified in an area of study as broad as I have mapped out in this book, but these can be discovered by proceeding with the rest of the book.

References and Suggested Readings

Armstrong, B. K., White, E., & Saracci, R. (1992). *Principles of exposure measurement in epidemiology*. New York: Oxford University Press.

Breslow, L. (1999). From disease prevention to disease prevention. *Journal of the American Medical Association*, 281, 1030–1033.

Caldwell, J. C. (2001, February 9–10). *Demography: Scope, perspectives, and theory*. Paper presented at the conference on demography and epidemiology: Frontiers in population and aging, Washington, DC: Center for Population and Health, Georgetown University.

Carey, J. R. (1993). *Applied demography for biologists: With special emphasis on insects*. New York, Oxford University Press.

Carnes, B. A., & Olshansky, S. J. (1993). Evolutionary perspectives on human senescence. *Population Development Review*, 19(4), 793–806.

Cohen, J. (1984). Demography and morbidity: A survey of some interactions. In N. Keyfitz (Ed.), *Population and biology: Bridge between disciplines. Proceedings of a conference* (pp. 199–220). Liège: Ordina Editions.

Crimmins, E. M., & Seaman, T. (2000). Integrating biology into demographic research on aging (with a focus on the McArthur study of successful aging). In C. E. Finch, J. W. Vaupel, & K. Kinsella (Eds.), *Cells and surveys: Should biological measures be included in social science research? national research council* (pp. 9–41). Washington, DC: National Academy Press.

de Grey, A. D. N. J. (2005). A strategy for postponing aging indefinitely. *Studies in Health Technology and Informatics*, 118, 209–219.

Dubos, R. (1968). *Determinants of health and disease in man, medicine and environment*. New York: Frederick A. Praeger.

Dunn, H. L. (1957). Points of attack for raising the levels of wellness. *Journal of the American Medical Association*, 49(4), 225–235.

Goldman, N. (2002). Incorporating health into demographic training. *Genus*, 58(3–4), 91–107.

Gordis, L. (2005). *Epidemiology*. Philadelphia: W.B.Saunders.

Gray, R. H. (1989). The integration of demographic and epidemiological approaches to studies of health in developing countries. In L. Ruzicka, G. Wunsch, & P. Kanes (Eds.), *Differential mortality: Methodological issues and biosocial factors* (pp. 36–63). Oxford: Clarendon Press.

Keyfitz, N. (Ed.). (1984). *Population and biology: Bridge between disciplines. Proceedings of a Conference*. Liège: Ordina Editions.

Mahler, H. (1981). The meaning of "health for all" by the Year 2000, *World Health Forum* 3(1), 5–22.

Manton, K. G., & Stallard, E. (1994). Medical demography: Interaction of disability dynamics and mortality. In L. G. Martin & S. H. Preston (Eds.), *Demography of aging* (pp. 217–278). Washington, DC: National Academy Press.

Martin, L. G., & Preston, S. H. (Eds.), (1994). *Demography of aging*. National Academy of Sciences. Washington, DC: National Academy Press.

Mausner, J. S., & Kramer, S. (1985). *Epidemiology: An introductory text*. Philadelphia: W.B. Saunders Company.

Murray, C. J. L., Salomon, J., & Mathers, C. A critical examination of summary measures of population health. (1999). *WHO global programme on evidence for health policy. Discussion Paper No.2*. Geneva: World Health Organization.

Myers, G. C., & Eggers, M. L. (1996). Demography. In J. Birren (Ed.), *Encyclopedia of gerontology* (Vol. 1). San Diego, CA: Academic.

New England Research Institute. (1994). *Network*. Summer/Fall.

Olshansky, S. J., & Carnes, B. A. (1994). Demographic perspectives on human senescence. *Population and Development Review,* 20(1), 57–80.

Olshansky, S. J., Grant, M., Brody, J., & Carnes, B. C. (2005). Biodemographic perspectives for epidemiologists In S. J. Olshansky, et al. (Eds.), *Emerging themes in epidemiology*. NCBI/ NLM/NIH. BioMed Central.2:10, Sept. 30.

Omran, A. R. (1974). Population epidemiology: An emerging field of inquiry for population and health students. *American Journal of Public Health*, 64(7), 674–679.

President's Commission on the Health Needs of the Nation. (1953). *America's health status, needs and resources. Building America's Health* 2:13. Washington, DC: U.S. Government Printing Office.

Rockett, I. (1999). *Population and health: An introduction to epidemiology: Vol. 54. Population bulletin*. Washington, DC: Population Reference Bureau.

Scott, S., & Duncan, C. J. (1998). *Human demography and disease*. New York: Cambridge University Press.

Siegel, J. S. (1980). On the demography of aging. *Demography*, 17(4), 345–364.

Siegel, J. S. (1993). *A generation of change: A profile of America's older population*. New York: Russell Sage Population.

Siegal, J. S., & Swanson, D.A. (Eds.), (2004) *Methods and materials of demography, eps. Chapter 14, Lamb, V.A., and Siegel, J.S., Health demography* (2nd ed., San Diego: Elsevier/Academic Press.)

Singer, B. H., & Ryff, C. D. (Eds.). (2001). *New horizons in health: An integrative approach*. Washington, DC: National Academy Press.

Stedman, T. L. (2006). *Medical dictionary* (28th ed.). Baltimore: Lippincott, Williams, and Wilkins.

Susser, M. (1987). Epidemiology in the United States after World War II: The evolution of technique. *Epidemiogic Review*, 7, 147–177.

U.S. Department of Health and Human Services. (2000a). *Healthy people 2010: Understanding and improving health* (2nd ed.). Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. (2000b). *International health data reference guide, 1999* (DHHS Publication No. (PHS) 2000–1007). Washington, DC: U.S. Government Printing Office.

U.S. National Center for Health Statistics. (2001). *DATA2010: The healthy people 2010 database*. <http://Wonder.cdc.gov/data2010/>.

Wachter, K., & Bulatao, R. A. (2003). *Offspring: Human fertility behavior in biodemographic perspective*. Washington, DC: National Academy Press.

Wachter, K. W., & Finch, C. E. (Eds.), (1997). *Between Zeus and the Salmon: The biodemography of aging*. Washington, DC: National Academy Press/National Academy of Sciences.

Weinstein, M., Hermalin, A., & Stoto, M. (Eds.), (2001). *Population health and aging: strengthening the dialogue between epidemiology and demography: Vol. 954. Annals of the New York Academy of Sciences*. New York: New York Academy of Sciences.

World Health Organization. (1947). Constitution of the World Health Organization. *Chronicle of the World Health Organization*, 1, 29–43.

World Health Organization. (1984, August 30–September 1). *The uses of epidemiology in the study of the elderly: Report of a scientific group on the epidemiology of aging. (Technical Report Series* 706). Geneva, Switzerland: World Health Organization.

World Health Organization. (1993). *Basic epidemiology*. Geneva, Switzerland: World Health Organization.

Zimmerman, M. K. (2000). Medical sociology. In *Encyclopedia of sociology* (2nd ed., Vol. 3).

Chapter 2
Sources and Quality of Data for Mortality and Health Studies

Mortality and health data can be obtained both from the general sources for demographic data, such as censuses, vital statistics registrations, and general sample surveys, and from specialized sources, such as national sample surveys on health, administrative records on health and mortality (e.g., hospital discharge records; records of hospital inpatient stays), epidemiological studies, and clinical trials. These sources provide quantitative macrodata (aggregate data) and microdata (individual data) on mortality and health. I consider each of them in turn and then briefly discuss qualitative sources of information on health.

Measures of the health of a population may be based on either subjective reports on health, that is, reports given by respondents on how they feel, or on objective information, that is, reports given by respondents on the basis of medical records, results of physical examinations, or medical records of health providers. More specifically, objective measures of health may be based on the results of clinical examinations of a sample of a population or on administrative records, such as hospital discharge or inpatient records. However, many health measures are based on a combination of sources, for example, respondents' reports of their health conditions and a health examination by a health worker or record of a hospital stay.

National Censuses

U.S. Censuses

Censuses have ventured into the area of obtaining health data in a limited way even though they are not well designed to secure such statistics. This area of inquiry usually calls for a battery of probing questions. The Census' more general contributions to health studies are to provide population bases for calculating various types of mortality and morbidity rates, the socioeconomic data used as correlates in analyzing geographic and other variations in mortality and health, and the population data needed for mapping health service areas.

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_2, © Springer Science+Business Media B.V. 2012

U.S. population censuses once obtained data regarding health conditions that are now termed sensory impairments and physical and mental disabilities. The 1830 U.S. Census was the first U.S. census to contain questions on the number of persons who were "deaf, dumb, or blind." The 1840 census added the categories "insane" and "idiots." These types of questions were continued through the 1890 census.

Specific questions on disabilities first appeared in the 1880 census, which asked about "temporary disability" as well as "chronic disability" associated with sensory, physical, or mental defects. Such inquiries were repeated in the 1890 census, but the questions on disability were rephrased to obtain data on "chronic" and "acute" diseases, by type and duration of disease. Several recent censuses of the United States (1970, 1980, 1990, and 2000) had questions on sensory disability, physical disability, and mental disability, and activity limitations with respect to work, transportation, and/or personal care. The 2000 U.S. Census, for example, asked the following yes/no questions:

Does this person have any of the following long-lasting conditions:

a. Blindness, deafness, or a severe vision or hearing impairment?
b. A condition that substantially limits one or more basic physical activities such as walking, climbing stairs, reaching, lifting, or carrying?

Because of a physical, mental, or emotional condition lasting 6 months or more, does this person have any difficulty in doing any of the following activities:

a. Learning, remembering, or concentrating?
b. Dressing, bathing, or getting around inside the home?
c. [Answer if this person is 16 YEARS OLD OR OVER.] Going outside the home alone to shop or visit a doctor's office?
d. [Answer if this person is 16 YEARS OLD OR OVER.] Working at a job or business?

The reader will note that these questions are designed to measure the prevalence of various types and degrees of disability in the population, distinguishing sensory disability, physical disability, mental disability, work disability, transportation disability, and personal-care disability. Currently and in 2010 the American Community Survey, the national sample survey replacing the traditional sample inquiries on the 2010 census, will ask essentially the same questions, with the following modifications: Separate questions on visual disability and hearing disability; omission of the question on work disability.

Several U.S. censuses in the second half of the nineteenth century – prior to the establishment of the vital registration system in the United States in 1900 – inquired about deaths in the household during the previous year. These censuses were those of 1850, 1860, 1870, 1880, and 1885 (four states and one territory). The separate mortality schedules were quite detailed, including questions on age, sex, race, marital status, occupation, place of birth, and month and cause of death. Demographers have used both historic (Costa 2000; Elman and Myers 1999, 1997; Preston and Haines 1991) and recent (Geronimus et al. 2001; Hayward and Heron 1999) U.S. censuses to examine trends in health and health expectancy in the United States.

U.S. Census Data Sets and Software to Access Census Data

U.S. census data and American Community Survey (ACS) data are available on the American FactFinder website. Both census and ACS data are available through Data Ferret (Federal Electronic Research, Review, and Extraction Tool), a data access tool that allows the user to customize tables, graphs, and maps. Data Ferret also allows the user to download user-specified portions of the data. Some other tools for accessing decennial census data include:

PDQ-Explore (a software tool developed by Public Data Queries, Inc.)
StudentCHIP and Beyond 20/20 (software tools for accessing the census)
The Demographic Data Center (www.usa-demographics.com): Commercial data-vending company

Censuses of Other Countries

A number of other countries have collected data on the health of their populations in their censuses. The inquiries have related mainly to disability and, in the less developed countries, have focused on sensory and physical disability. The list of more developed countries include the United Kingdom (1991), Poland (1988), and Australia (1976), and the list of less developed countries include Chile (1992), Turkey (1985), Botswana (1991), India (1981), and Pakistan (1981). Most developed countries secure their disability information through surveys, however, and even the less developed countries have been shifting from censuses to surveys to secure such information.

In 1988 the United Nations completed, DISTAT, the United Nations Disability Statistics Database. It contains disability statistics from population censuses, national household surveys, and population or civil registration systems for 55 countries. Based upon national statistics available in DISTAT, the United Nations published the *Disability Statistics Compendium* in 1990. In 2006 the United Nations Statistics Division initiated the collection of basic statistics on disability through the same data collection system by which it collects data for its *Demographic Yearbook*. Earlier, the 1993 U.N. *Demographic Yearbook,* a Special Issue on Population Aging and the Situation of Elderly Persons, had included a table on disabled persons.

Vital Statistics and Related Administrative Records

National Vital Statistics Systems

The various registrations comprising the vital registration system of a country produce several sets of data related to health, including data on births, deaths, fetal losses, induced terminations of pregnancy, and the health conditions associated

with these events. From each state government's point of view, these are primarily administrative records establishing the facts of birth and death for legal and related purposes. Public health officials view these data primarily as indicators of the changing health situation and a guide for administering public health programs. Demographers view these data primarily as basic demographic data for use in measuring population change. From the point of view of the biostatistician and epidemiologist, certificates of birth, death, and fetal loss are a major source of statistical data to identify public health problems, to monitor progress in public health, and to determine the allocation of funds for public health research. The various data products of vital statistics systems are the subject of a few other chapters in this book, namely deaths in Chap. 3, life tables, a derivative product of vital statistics, in Chap. 4, and births and fetal losses in Chap. 9.

U.S. Standard Certificates of Birth and Death, and Report of Fetal Death

Standard or model certificates of birth, death, fetal death, and induced terminations of pregnancy in the United States were developed to promote uniformity in data collection and tabulation across registration areas. The National Center for Health Statistics (NCHS) designs the standard forms for reporting and recording vital events in cooperation with state vital statistics officials. The standard certificates are generally revised every 10–15 years. They were last revised during the late 1990s and early 2000s, and the new certificates have gradually been adopted by the states beginning in 2003, with full implementation being phased in over several years. The previous standard certificates were introduced in 1989. The process involved in revising the content of the standard certificates is described in the "Executive Summary" of the *Report of the Panel to Evaluate the U.S. Standard Certificates and Reports*, available at the web site of NCHS (www.cdc.gov/nchs). Copies of the latest certificates of birth and death are displayed as Exhibits 2.1 and 2.2. Copies of the other reporting forms are shown as exhibits in the Appendix to Chap. 9. Copies of these documents are also available at the NCHS web site. NCHS has issued an updated version of its *Physicians' Handbook on Medical Certification of Death* (2003 Revision) and *Coroners' Handbook on Death Registration and Fetal Death Reporting* (2003 Revision).

Special Surveys on U.S. Mortality and Data Sets on Mortality

In addition to the basic products of the vital statistics system as described above, the National Center for Health Statistics generates special data sets on vital statistics for both research and administrative uses. The principle data sets are:

Linked Files of Live Birth and Infant Death Records. Links the information on the birth certificate and the infant death record for babies who died in their first year after birth.

Exhibit 2.1.-- U.S. STANDARD CERTIFICATE OF LIVE BIRTH

LOCAL FILE NO. BIRTH NUMBER:

CHILD

1. CHILD'S NAME (First, Middle, Last, Suffix) | 2. TIME OF BIRTH (24 hr) | 3. SEX | 4. DATE OF BIRTH (Mo/Day/Yr)

5. FACILITY NAME (If not institution, give street and number) | 6. CITY, TOWN, OR LOCATION OF BIRTH | 7. COUNTY OF BIRTH

MOTHER

8a. MOTHER'S CURRENT LEGAL NAME (First, Middle, Last, Suffix) | 8b. DATE OF BIRTH (Mo/Day/Yr)

8c. MOTHER'S NAME PRIOR TO FIRST MARRIAGE (First, Middle, Last, Suffix) | 8d. BIRTHPLACE (State, Territory, or Foreign Country)

9a. RESIDENCE OF MOTHER-STATE | 9b. COUNTY | 9c. CITY, TOWN, OR LOCATION

9d. STREET AND NUMBER | 9e. APT. NO. | 9f. ZIP CODE | 9g. INSIDE CITY LIMITS? ☐ Yes ☐ No

FATHER

10a. FATHER'S CURRENT LEGAL NAME (First, Middle, Last, Suffix) | 10b. DATE OF BIRTH (Mo/Day/Yr) | 10c. BIRTHPLACE (State, Territory, or Foreign Country)

CERTIFIER

11. CERTIFIER'S NAME: _____
TITLE: ☐ MD ☐ DO ☐ HOSPITAL ADMIN. ☐ CNM/CM ☐ OTHER MIDWIFE
☐ OTHER (Specify)_____

12. DATE CERTIFIED ____/____/_____ MM DD YYYY

13. DATE FILED BY REGISTRAR ____/____/_____ MM DD YYYY

INFORMATION FOR ADMINISTRATIVE USE

MOTHER

14. MOTHER'S MAILING ADDRESS: 9 Same as residence, or: State: City, Town, or Location:
Street & Number: Apartment No.: Zip Code:

15. MOTHER MARRIED? (At birth, conception, or any time between) ☐ Yes ☐ No
IF NO, HAS PATERNITY ACKNOWLEDGEMENT BEEN SIGNED IN THE HOSPITAL? ☐ Yes ☐ No

16. SOCIAL SECURITY NUMBER REQUESTED FOR CHILD? ☐ Yes ☐ No

17. FACILITY ID. (NPI)

18. MOTHER'S SOCIAL SECURITY NUMBER:

19. FATHER'S SOCIAL SECURITY NUMBER:

INFORMATION FOR MEDICAL AND HEALTH PURPOSES ONLY

MOTHER

20. MOTHER'S EDUCATION (Check the box that best describes the highest degree or level of school completed at the time of delivery)
☐ 8th grade or less
☐ 9th - 12th grade, no diploma
☐ High school graduate or GED completed
☐ Some college credit but no degree
☐ Associate degree (e.g., AA, AS)
☐ Bachelor's degree (e.g., BA, AB, BS)
☐ Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)
☐ Doctorate (e.g., PhD, EdD) or Professional degree (e.g., MD, DDS, DVM, LLB, JD)

21. MOTHER OF HISPANIC ORIGIN? (Check the box that best describes whether the mother is Spanish/Hispanic/Latina. Check the "No" box if mother is not Spanish/Hispanic/Latina)
☐ No, not Spanish/Hispanic/Latina
☐ Yes, Mexican, Mexican American, Chicana
☐ Yes, Puerto Rican
☐ Yes, Cuban
☐ Yes, other Spanish/Hispanic/Latina
(Specify)_____

22. MOTHER'S RACE (Check one or more races to indicate what the mother considers herself to be)
☐ White
☐ Black or African American
☐ American Indian or Alaska Native (Name of the enrolled or principal tribe)_____
☐ Asian Indian
☐ Chinese
☐ Filipino
☐ Japanese
☐ Korean
☐ Vietnamese
☐ Other Asian (Specify)_____
☐ Native Hawaiian
☐ Guamanian or Chamorro
☐ Samoan
☐ Other Pacific Islander (Specify)_____
☐ Other (Specify)_____

FATHER

23. FATHER'S EDUCATION (Check the box that best describes the highest degree or level of school completed at the time of delivery)
☐ 8th grade or less
☐ 9th - 12th grade, no diploma
☐ High school graduate or GED completed
☐ Some college credit but no degree
☐ Associate degree (e.g., AA, AS)
☐ Bachelor's degree (e.g., BA, AB, BS)
☐ Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)
☐ Doctorate (e.g., PhD, EdD) or Professional degree (e.g., MD, DDS, DVM, LLB, JD)

24. FATHER OF HISPANIC ORIGIN? (Check the box that best describes whether the father is Spanish/Hispanic/Latino. Check the "No" box if father is not Spanish/Hispanic/Latino)
☐ No, not Spanish/Hispanic/Latino
☐ Yes, Mexican, Mexican American, Chicano
☐ Yes, Puerto Rican
☐ Yes, Cuban
☐ Yes, other Spanish/Hispanic/Latino
(Specify)_____

25. FATHER'S RACE (Check one or more races to indicate what the father considers himself to be)
☐ White
☐ Black or African American
☐ American Indian or Alaska Native (Name of the enrolled or principal tribe)_____
☐ Asian Indian
☐ Chinese
☐ Filipino
☐ Japanese
☐ Korean
☐ Vietnamese
☐ Other Asian (Specify)_____
☐ Native Hawaiian
☐ Guamanian or Chamorro
☐ Samoan
☐ Other Pacific Islander (Specify)_____
☐ Other (Specify)_____

Mother's Name | Mother's Medical Record No.

26. PLACE WHERE BIRTH OCCURRED (Check one)
☐ Hospital
☐ Freestanding birthing center
☐ Home Birth: Planned to deliver at home? 9 Yes 9 No
☐ Clinic/Doctor's office
☐ Other (Specify)_____

27. ATTENDANT'S NAME, TITLE, AND NPI
NAME: _____ NPI:_____
TITLE: ☐ MD ☐ DO ☐ CNM/CM ☐ OTHER MIDWIFE
☐ OTHER (Specify)_____

28. MOTHER TRANSFERRED FOR MATERNAL MEDICAL OR FETAL INDICATIONS FOR DELIVERY? ☐ Yes ☐ No
IF YES, ENTER NAME OF FACILITY MOTHER TRANSFERRED FROM:
_____

REV. 11/2003

Exhibit 2.1 U.S. Standard Certificate of Live Birth

National Death Index (NDI). A computerized file of information on death records designed to assist researchers in determining whether specific study subjects have died. To do this, it tries to effect matches of information provided by qualified parties with the complete death certificate.

National Mortality Database. Presents information on national mortality derived from information on the original death certificates.

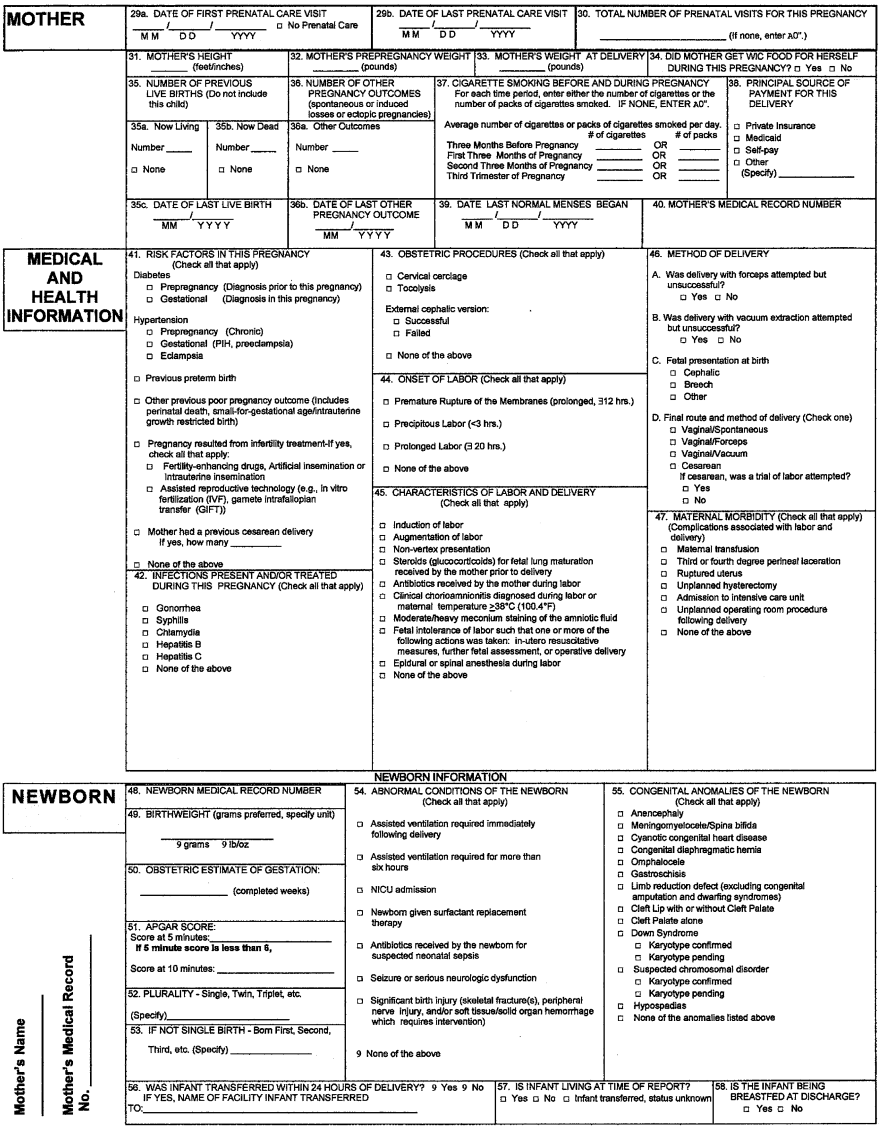
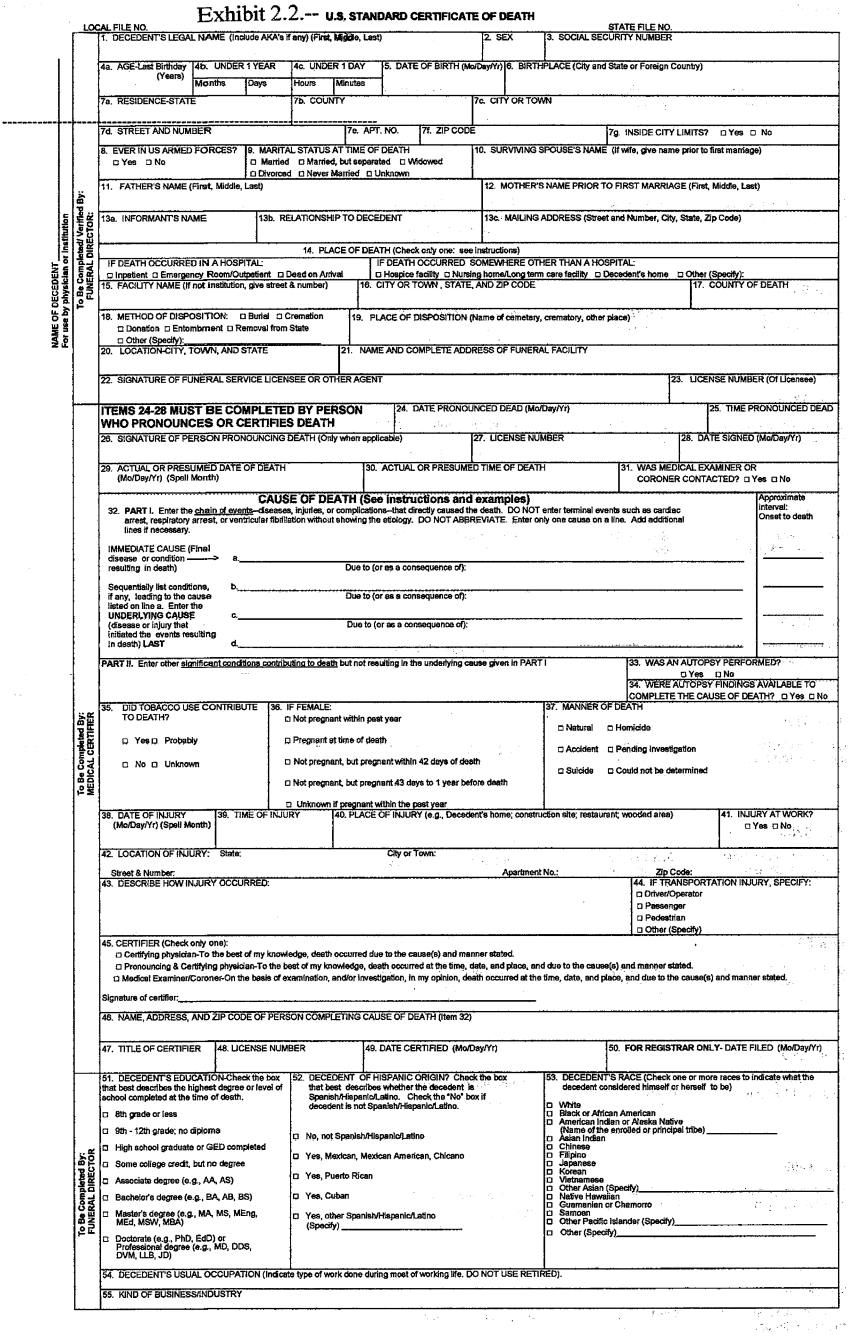


Exhibit 2.1 (continued)

National Longitudinal Mortality Survey (NLMS). Designed for the intensive study of mortality in the United States; provides information for the analysis of mortality differences between the races and their determinants, and on the quality of race and Hispanic-origin reporting on the death certificate.

Exhibit 2.2.-- U.S. STANDARD CERTIFICATE OF DEATH



LOCAL FILE NO. STATE FILE NO.

NAME OF DECEDENT — For use by physician or institution — To Be Completed/ Verified By: FUNERAL DIRECTOR:

1. DECEDENT'S LEGAL NAME (Include AKA's if any) (First, Middle, Last) | 2. SEX | 3. SOCIAL SECURITY NUMBER

4a. AGE-Last Birthday (Years) | 4b. UNDER 1 YEAR Months Days | 4c. UNDER 1 DAY Hours Minutes | 5. DATE OF BIRTH (Mo/Day/Yr) | 6. BIRTHPLACE (City and State or Foreign Country)

7a. RESIDENCE-STATE | 7b. COUNTY | 7c. CITY OR TOWN

7d. STREET AND NUMBER | 7e. APT. NO. | 7f. ZIP CODE | 7g. INSIDE CITY LIMITS? ☐ Yes ☐ No

8. EVER IN US ARMED FORCES? ☐ Yes ☐ No | 9. MARITAL STATUS AT TIME OF DEATH ☐ Married ☐ Married, but separated ☐ Widowed ☐ Divorced ☐ Never Married ☐ Unknown | 10. SURVIVING SPOUSE'S NAME (If wife, give name prior to first marriage)

11. FATHER'S NAME (First, Middle, Last) | 12. MOTHER'S NAME PRIOR TO FIRST MARRIAGE (First, Middle, Last)

13a. INFORMANT'S NAME | 13b. RELATIONSHIP TO DECEDENT | 13c. MAILING ADDRESS (Street and Number, City, State, Zip Code)

14. PLACE OF DEATH (Check only one: see instructions)

IF DEATH OCCURRED IN A HOSPITAL: ☐ Inpatient ☐ Emergency Room/Outpatient ☐ Dead on Arrival
IF DEATH OCCURRED SOMEWHERE OTHER THAN A HOSPITAL: ☐ Hospice facility ☐ Nursing home/Long term care facility ☐ Decedent's home ☐ Other (Specify):

15. FACILITY NAME (If not institution, give street & number) | 16. CITY OR TOWN, STATE, AND ZIP CODE | 17. COUNTY OF DEATH

18. METHOD OF DISPOSITION: ☐ Burial ☐ Cremation ☐ Donation ☐ Entombment ☐ Removal from State ☐ Other (Specify): | 19. PLACE OF DISPOSITION (Name of cemetery, crematory, other place)

20. LOCATION-CITY, TOWN, AND STATE | 21. NAME AND COMPLETE ADDRESS OF FUNERAL FACILITY

22. SIGNATURE OF FUNERAL SERVICE LICENSEE OR OTHER AGENT | 23. LICENSE NUMBER (Of Licensee)

ITEMS 24-28 MUST BE COMPLETED BY PERSON WHO PRONOUNCES OR CERTIFIES DEATH

24. DATE PRONOUNCED DEAD (Mo/Day/Yr) | 25. TIME PRONOUNCED DEAD

26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable) | 27. LICENSE NUMBER | 28. DATE SIGNED (Mo/Day/Yr)

29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month) | 30. ACTUAL OR PRESUMED TIME OF DEATH | 31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? ☐ Yes ☐ No

CAUSE OF DEATH (See instructions and examples)

32. **PART I.** Enter the chain of events--diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary. — Approximate interval: Onset to death

IMMEDIATE CAUSE (Final disease or condition ——> resulting in death) a. ____ Due to (or as a consequence of):

Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST
b. ____ Due to (or as a consequence of):
c. ____ Due to (or as a consequence of):
d. ____

PART II. Enter other significant conditions contributing to death but not resulting in the underlying cause given in PART I | 33. WAS AN AUTOPSY PERFORMED? ☐ Yes ☐ No | 34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ☐ Yes ☐ No

To Be Completed By: MEDICAL CERTIFIER

35. DID TOBACCO USE CONTRIBUTE TO DEATH? ☐ Yes ☐ Probably ☐ No ☐ Unknown

36. IF FEMALE: ☐ Not pregnant within past year ☐ Pregnant at time of death ☐ Not pregnant, but pregnant within 42 days of death ☐ Not pregnant, but pregnant 43 days to 1 year before death ☐ Unknown if pregnant within the past year

37. MANNER OF DEATH ☐ Natural ☐ Homicide ☐ Accident ☐ Pending Investigation ☐ Suicide ☐ Could not be determined

38. DATE OF INJURY (Mo/Day/Yr) (Spell Month) | 39. TIME OF INJURY | 40. PLACE OF INJURY (e.g., Decedent's home; construction site; restaurant; wooded area) | 41. INJURY AT WORK? ☐ Yes ☐ No

42. LOCATION OF INJURY: State: City or Town: Street & Number: Apartment No.: Zip Code:

43. DESCRIBE HOW INJURY OCCURRED: | 44. IF TRANSPORTATION INJURY, SPECIFY: ☐ Driver/Operator ☐ Passenger ☐ Pedestrian ☐ Other (Specify)

45. CERTIFIER (Check only one):
☐ Certifying physician-To the best of my knowledge, death occurred due to the cause(s) and manner stated.
☐ Pronouncing & Certifying physician-To the best of my knowledge, death occurred at the time, date, and place, and due to the cause(s) and manner stated.
☐ Medical Examiner/Coroner-On the basis of examination, and/or investigation, in my opinion, death occurred at the time, date, and place, and due to the cause(s) and manner stated.

Signature of certifier: ____

46. NAME, ADDRESS, AND ZIP CODE OF PERSON COMPLETING CAUSE OF DEATH (Item 32)

47. TITLE OF CERTIFIER | 48. LICENSE NUMBER | 49. DATE CERTIFIED (Mo/Day/Yr) | 50. FOR REGISTRAR ONLY- DATE FILED (Mo/Day/Yr)

To Be Completed By: FUNERAL DIRECTOR

51. DECEDENT'S EDUCATION-Check the box that best describes the highest degree or level of school completed at the time of death.
☐ 8th grade or less
☐ 9th - 12th grade; no diploma
☐ High school graduate or GED completed
☐ Some college credit, but no degree
☐ Associate degree (e.g., AA, AS)
☐ Bachelor's degree (e.g., BA, AB, BS)
☐ Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)
☐ Doctorate (e.g., PhD, EdD) or Professional degree (e.g., MD, DDS, DVM, LLB, JD)

52. DECEDENT OF HISPANIC ORIGIN? Check the box that best describes whether the decedent is Spanish/Hispanic/Latino. Check the "No" box if decedent is not Spanish/Hispanic/Latino.
☐ No, not Spanish/Hispanic/Latino
☐ Yes, Mexican, Mexican American, Chicano
☐ Yes, Puerto Rican
☐ Yes, Cuban
☐ Yes, other Spanish/Hispanic/Latino (Specify) ____

53. DECEDENT'S RACE (Check one or more races to indicate what the decedent considered himself or herself to be)
☐ White
☐ Black or African American
☐ American Indian or Alaska Native (Name of the enrolled or principal tribe) ____
☐ Asian Indian
☐ Chinese
☐ Filipino
☐ Japanese
☐ Korean
☐ Vietnamese
☐ Other Asian (Specify) ____
☐ Native Hawaiian
☐ Guamanian or Chamorro
☐ Samoan
☐ Other Pacific Islander (Specify) ____
☐ Other (Specify) ____

54. DECEDENT'S USUAL OCCUPATION (Indicate type of work done during most of working life. DO NOT USE RETIRED).

55. KIND OF BUSINESS/INDUSTRY

REV. 11/2003

Exhibit 2.2 U.S. Standard Certificate of Death

Cause-of-death – Background, Examples, and Common Problems

Accurate cause of death information is important
•to the public health community in evaluating and improving the health of all citizens, and
•often to the family, now and in the future, and to the person settling the decedent's estate.

The cause-of-death section consists of two parts. **Part I is for reporting a chain of events leading directly to death, with the immediate cause of death (the final disease, injury, or complication directly causing death) on line a and the underlying cause of death (the disease or injury that initiated the chain of events that led directly and inevitably to death) on the lowest used line. Part II is for reporting all other significant diseases, conditions, or injuries that contributed to death but which did not result in the underlying cause of death given in Part I. The cause-of-death information should be YOUR best medical OPINION.** A condition can be listed as "probable" even if it has not been definitively diagnosed.

Examples of properly completed medical certifications

CAUSE OF DEATH (See instructions and examples)

		Approximate interval: Onset to death
32. **PART I.** Enter the chain of events--diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary.		
IMMEDIATE CAUSE (Final disease or condition ——> resulting in death)	a. Rupture of myocardium Due to (or as a consequence of):	Minutes
Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST	b. Acute myocardial infarction Due to (or as a consequence of):	6 days
	c. Coronary artery thrombosis Due to (or as a consequence of):	5 years
	d. Atherosclerotic coronary artery disease	7 years

PART II. Enter other significant conditions contributing to death but not resulting in the underlying cause given in PART I	33. WAS AN AUTOPSY PERFORMED? ■ Yes □ No
Diabetes, Chronic obstructive pulmonary disease, smoking	34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ■ Yes □ No

35. DID TOBACCO USE CONTRIBUTE TO DEATH?	36. IF FEMALE:	37. MANNER OF DEATH
■ Yes □ Probably	■ Not pregnant within past year □ Pregnant at time of death	■ Natural □ Homicide
□ No □ Unknown	□ Not pregnant, but pregnant within 42 days of death □ Not pregnant, but pregnant 43 days to 1 year before death □ Unknown if pregnant within the past year	□ Accident □ Pending Investigation □ Suicide □ Could not be determined

CAUSE OF DEATH (See instructions and examples)

		Approximate interval: Onset to death
32. **PART I.** Enter the chain of events--diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary.		
IMMEDIATE CAUSE (Final disease or condition ——> resulting in death)	a. Aspiration pneumonia Due to (or as a consequence of):	2 Days
Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST	b. Complications of coma Due to (or as a consequence of):	7 weeks
	c. Blunt force injuries Due to (or as a consequence of):	7 weeks
	d. Motor vehicle accident	7 weeks

PART II. Enter other significant conditions contributing to death but not resulting in the underlying cause given in PART I	33. WAS AN AUTOPSY PERFORMED? ■ Yes □ No
	34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ■ Yes □ No

35. DID TOBACCO USE CONTRIBUTE TO DEATH?	36. IF FEMALE:	37. MANNER OF DEATH
□ Yes □ Probably	□ Not pregnant within past year □ Pregnant at time of death	□ Natural □ Homicide
■ No □ Unknown	□ Not pregnant, but pregnant within 42 days of death □ Not pregnant, but pregnant 43 days to 1 year before death □ Unknown if pregnant within the past year	■ Accident □ Pending Investigation □ Suicide □ Could not be determined

38. DATE OF INJURY (Mo/Day/Yr) (Spell Month)	39. TIME OF INJURY	40. PLACE OF INJURY (e.g., Decedent's home; construction site; restaurant; wooded area)	41. INJURY AT WORK?
August 15, 2003	Approx. 2320	road side near state highway	□ Yes ■ No

42. LOCATION OF INJURY: State: Missouri	City or Town: near Alexandria	
Street & Number: mile marker 17 on state route 46a	Apartment No.:	Zip Code:

43. DESCRIBE HOW INJURY OCCURRED:	44. IF TRANSPORTATION INJURY, SPECIFY:
Decedent driver of van, ran off road into tree	■ Driver/Operator □ Passenger □ Pedestrian □ Other (Specify)

Common problems in death certification

The **elderly decedent** should have a clear and distinct etiological sequence for cause of death, if possible. Terms such as senescence, infirmity, old age, and advanced age have little value for public health or medical research. Age is recorded elsewhere on the certificate. When a number of conditions resulted in death, the physician should choose the single sequence that, in his or her opinion, best describes the process leading to death, and place any other pertinent conditions in Part II. If after careful consideration the physician cannot determine a sequence that ends in death, then the medical examiner or coroner should be consulted about conducting an investigation or providing assistance in completing the cause of death.

The **infant decedent** should have a clear and distinct etiological sequence for cause of death, if possible. "Prematurity" should not be entered without explaining the etiology of prematurity. Maternal conditions may have initiated or affected the sequence that resulted in infant death, and such maternal causes should be reported in addition to the infant causes on the infant's death certificate (e.g., Hyaline membrane disease due to prematurity, 28 weeks due to placental abruption due to blunt trauma to mother's abdomen).

When **SIDS** is suspected, a complete investigation should be conducted, typically by a medical examiner or coroner. If the infant is under 1 year of age, no cause of death is determined after scene investigation, clinical history is reviewed, and a complete autopsy is performed, then the death can be reported as Sudden Infant Death Syndrome.

When processes such as the following are reported, additional information about the etiology should be reported:

Abscess
Abdominal hemorrhage
Adhesions
Adult respiratory distress syndrome
Acute myocardial infarction
Altered mental status
Anemia
Anoxia
Anoxic encephalopathy
Arrhythmia
Ascites
Aspiration
Atrial fibrillation
Bacteremia
Bedridden
Biliary obstruction
Bowel obstruction
Brain injury
Brain stem herniation
Carcinogenesis
Carcinomatosis
Cardiac arrest
Cardiac dysrhythmia
Cardiomyopathy
Cardiopulmonary arrest
Cellulitis
Cerebral edema
Cerebrovascular accident
Cerebellar tonsillar herniation
Chronic bedridden state
Cirrhosis
Coagulopathy
Compression fracture
Congestive heart failure
Convulsions
Decubiti
Dehydration
Dementia (when not otherwise specified)
Diarrhea
Disseminated intra vascular coagulopathy
Dysrhythmia
End-stage liver disease
End-stage renal disease
Epidural hematoma
Exsanguination
Failure to thrive
Fracture
Gangrene
Gastrointestinal hemorrhage
Heart failure
Hemothorax
Hepatic failure
Hepatitis
Hepatorenal syndrome
Hyperglycemia
Hypercalcemia
Hypovolemic shock
Hyponatremia
Hypotension
Immunosuppression
Increased intra cranial pressure
Intra cranial hemorrhage
Malnutrition
Metabolic encephalopathy
Multi-organ failure
Multi-system organ failure
Myocardial infarction
Necrotizing soft-tissue infection
Old age
Open (or closed) head injury
Paralysis
Pancytopenia
Perforated gallbladder
Peritonitis
Pleural effusions
Pneumonia
Pulmonary arrest
Pulmonary edema
Pulmonary embolism
Pulmonary insufficiency
Renal failure
Respiratory arrest
Seizures
Sepsis
Septic shock
Shock
Starvation
Subdural hematoma
Subarachnoid hemorrhage
Sudden death
Thrombocytopenia
Uncal herniation
Urinary tract infection
Ventricular fibrillation
Ventricular tachycardia
Volume depletion

If the certifier is unable to determine the etiology of a process such as those shown above, the process must be qualified as being of an unknown, undetermined, probable, presumed, or unspecified etiology so it is clear that a distinct etiology was not inadvertently or carelessly omitted.

The following conditions and types of death might seem to be specific or natural but when the medical history is examined further may be found to be complications of an injury or poisoning (possibly occurring long ago). Such cases should be reported to the medical examiner/coroner.

Asphyxia
Bolus
Choking
Drug or alcohol overdose/drug or alcohol abuse
Epidural hematoma
Exsanguination
Fall
Fracture
Hip fracture
Hypothermia
Hypothermia
Open reduction of fracture
Pulmonary emboli
Seizure disorder
Sepsis
Subarachnoid hemorrhage
Subdural hematoma
Surgery
Thermal burns/chemical burns

REV. 11/2003

Exhibit 2.2 (continued)

Exhibit 2.2a Medical Certifier Instructions

MEDICAL CERTIFIER INSTRUCTIONS for selected items on U.S. Standard Certificate of Death: 2003 Revision
(See Physicians' Handbook or Medical Examiner/Coroner Handbook on Death Registration for instructions on all items)

ITEMS ON WHEN DEATH OCCURRED
Items 24–25 and 29–31 should always be completed. If the facility uses a separate pronouncer or other person to indicate that death has taken place with another person more familiar with the case completing the remainder of the medical portion of the death certificate, the pronouncer completes Items 24–28. If a certifier completes Items 24–25 as well as items 29–49, Items 26–28 may be left blank.

ITEMS 24–25, 29–30 – DATE AND TIME OF DEATH Spell out the name of the month. If the exact date of death is unknown, enter the **approximate** date. If the date cannot be approximated, enter the date the body is found and identify as **date found**. Date pronounced and actual date may be the same. Enter the exact hour and minutes according to a 24-h clock; estimates may be provided with "Approx." placed before the time.

ITEM 32 – CAUSE OF DEATH (See attached examples)
Take care to make the entry legible. Use a computer printer with high resolution, typewriter with good black ribbon and clean keys, or print legibly using permanent **black** ink in completing the CAUSE OF DEATH Section. **Do not abbreviate** conditions entered in section.

Part I (Chain of events leading directly to death)

- Only **one** cause should be entered on each line. Line (a) **MUST ALWAYS** have an entry. **DO NOT** leave blank. Additional lines may be added if necessary.
- If the condition on Line (a) resulted from an underlying condition, put the underlying condition on Line (b), and so on, until the full sequence is reported. **ALWAYS** enter the **underlying cause of death** on the lowest used line in Part I.
- For each cause indicate the best estimate of the interval between the presumed onset and the date of death. The terms "unknown" or "approximately" may be used. General terms, such as minutes, hours, or days, are acceptable, if necessary. **DO NOT** leave blank.
- The terminal event (for example, cardiac arrest or respiratory arrest) should not be used. If a mechanism of death seems most appropriate to you for line (a), then you must always list its cause(s) on the line(s) below it (for example, cardiac arrest **due to** coronary artery atherosclerosis *or* cardiac arrest **due to** blunt impact to chest).
- If an organ system failure such as congestive heart failure, hepatic failure, renal failure, or respiratory failure is listed as a cause of death, always report its etiology on the line(s) beneath it (for example, renal failure **due to** Type I diabetes mellitus).

- When indicating neoplasms as a cause of death, include the following: (1) primary site *or* that the primary site is unknown, (2) benign or malignant, (3) cell type *or* that the cell type is unknown, (4) grade of neoplasm, and (5) part or lobe of organ affected. (For example, a primary well-differentiated squamous cell carcinoma, lung, left upper lobe.)
- Always report the fatal injury (for example, stab wound of chest), the trauma (for example, transection of subclavian vein), and impairment of function (for example, air embolism).

PART II (Other significant conditions)

- Enter all diseases or conditions contributing to death that were not reported in the chain of events in Part I and that did not result in the **underlying cause of death**. See attached examples.
- If two or more possible sequences resulted in death, or if two conditions seem to have added together, report in Part I the one that, in your opinion, most directly caused death. Report in Part II the other conditions or diseases.

CHANGES TO CAUSE OF DEATH

Should additional medical information or autopsy findings become available that would change the cause of death originally reported, the original death certificate should be amended by the certifying physician by **immediately** reporting the revised cause of death to the State Vital Records Office.

ITEMS 33–34 – AUTOPSY

- 33 – Enter "Yes" if either a partial or full autopsy was performed. Otherwise enter "No."
- 34 – Enter "Yes" if autopsy findings were available to complete the cause of death; otherwise enter "No". Leave item blank if no autopsy was performed.

ITEM 35 – DID TOBACCO USE CONTRIBUTE TO DEATH?

Check "yes" if, in your opinion, the use of tobacco contributed to death. Tobacco use may contribute to deaths due to a wide variety of diseases; for example, tobacco use contributes to many deaths due to emphysema or lung cancer and some heart disease and cancers of the head and neck. Check "no" if, in your clinical judgment, tobacco use did not contribute to this particular death.

ITEM 36 – IF FEMALE, WAS DECEDENT PREGNANT AT TIME OF DEATH OR WITHIN PAST YEAR?

This information is important in determining pregnancy-related mortality.

ITEM 37 – MANNER OF DEATH

- Always check Manner of Death, which is important: (1) in determining accurate causes of death; (2) in processing insurance claims; and (3) in statistical studies of injuries and death.
- Indicate "Pending investigation" if the manner of death cannot be determined whether due to an accident, suicide, or homicide within the statutory time limit

for filing the death certificate. This should be changed later to one of the other terms.
- Indicate "Could not be Determined" **ONLY** when it is impossible to determine the manner of death.

ITEMS 38–44 – ACCIDENT OR INJURY – to be filled out in all cases of deaths due to injury or poisoning.

- 38 – Enter the exact month, day, and year of injury. Spell out the name of the month. **DO NOT** use a number for the month. (Remember, the date of injury may differ from the date of death.) Estimates may be provided with "Approx." placed before the date.
- 39 – Enter the exact hour and minutes of injury or use your best estimate. Use a 24-h clock.
- 40 – Enter the general place (such as restaurant, vacant lot, or home) where the injury occurred. **DO NOT** enter firm or organization names.(For example, enter "factory", **not** "Standard Manufacturing, Inc.")
- 41 – Complete if anything other than natural disease is mentioned in Part I or Part II of the medical certification, including homicides, suicides, and accidents. This includes all motor vehicle deaths. The item **must** be completed for decedents ages 14 years or over and may be completed for those less than 14 years of age if warranted. Enter "Yes" if the injury occurred at work. Otherwise enter "No". An injury may occur at work regardless of whether the injury occurred in the course of the decedent's "usual" occupation. Examples of injury at work and injury not at work follow:

Injury at work Injury not at work
Injury while working or in vocational training on job premises Injury while engaged in personal recreational activity on job premises Injury while on break or at lunch or in parking lot on job premises Injury while a visitor (not on official work business) to job premises Injury while working for pay or compensation, including at home Homemaker working at homemaking activities Injury while working as a volunteer law enforcement official etc. Student in school Injury while traveling on business, including to/from business contacts Working for self for no profit (mowing yard, repairing own roof, hobby) Commuting to or from work

- 42 – Enter the complete address where the injury occurred including zip code.
- 43 – Enter a brief but specific and clear description of how the injury occurred. Explain the circumstances or cause of the injury. Specify **type of gun** or **type of vehicle** (e.g., car, bulldozer, train, etc.) when relevant to circumstances. Indicate if more than one vehicle involved; specify type of vehicle decedent was in.
- 44 – Specify role of decedent (e.g. driver, passenger). Driver/operator and passenger should be designated for modes other than motor vehicles such as bicycles. Other applies to watercraft, aircraft, animal, or people attached to outside of vehicles (e.g. surfers).

Rationale: Motor vehicle accidents are a major cause of unintentional deaths; details will help determine effectiveness of current safety features and laws.

REFERENCES
For more information on how to complete the medical certification section of the death certificate, refer to tutorial at http://www.TheNAME.org and resources including instructions and handbooks available by request from NCHS, Room 7318, 3311 Toledo Road, Hyattsville, Maryland 207822003 or at www.cdc.gov/nchs/about/major/dvs/handbk.htm

Exhibit 2.2b Funeral Director Instructions

FUNERAL DIRECTOR INSTRUCTIONS for selected items on U.S. Standard Certificate of Death (For additional information concerning all items on certificate see Funeral Directors' Handbook on Death Registration) 2003 Revision

ITEM 1. DECEDENT'S LEGAL NAME
Include any other names used by decedent, if substantially different from the legal name, after the abbreviation AKA (also known as) e.g. Samuel Langhorne Clemens AKA Mark Twain, **but not** Jonathon Doe AKA John Doe

ITEM 5. DATE OF BIRTH
Enter the full name of the month (January, February, March etc.) Do not use a number or abbreviation to designate the month.

ITEM 7A-G. RESIDENCE OF DECEDENT (information divided into seven categories)
Residence of decedent is the place where the decedent actually resided. The place of residence is not necessarily the same as "home state" or "legal residence". Never enter a temporary residence such as one used during a visit, business trip, or vacation. Place of residence during a tour of military duty or during attendance at college is considered permanent and should be entered as the place of residence. If the decedent had been living in a facility where an individual usually resides for a long period of time, such as a group home, mental institution, nursing home, penitentiary, or hospital for the chronically ill, report the location of that facility in item 7. If the decedent was an infant who never resided at home, the place of residence is that of the parent(s) or legal guardian. **Never** use an acute care hospital's location as the place of residence for any infant. If Canadian residence, please specify Province instead of State.

ITEM 10. SURVIVING SPOUSE'S NAME
If the decedent was married at the time of death, enter the full name of the surviving spouse. If the surviving spouse is the wife, enter her name prior to first marriage. This item is used in establishing proper insurance settlements and other survivor benefits.

ITEM 12. MOTHER'S NAME PRIOR TO FIRST MARRIAGE
Enter the name used prior to first marriage, commonly known as the maiden name. This name is useful because it remains constant throughout life.

ITEM 14. PLACE OF DEATH
The place where death is pronounced should be considered the place where death occurred. If the place of death is unknown but the body is found in your State, the certificate of death should be completed and filed in accordance with the laws of your State. Enter the place where the body is found as the place of death.

ITEM 51. DECEDENT'S EDUCATION (*Check appropriate box on death certificate*) Check the box that corresponds to the highest level of education that the decedent completed. **Information in this section will not appear on the certified copy of the death certificate. This information is used to study the relationship between mortality and education (which roughly corresponds with socioeconomic status). This information is valuable in medical studies of causes of death and in programs to prevent illness and death.**

ITEM 52. WAS DECEDENT OF HISPANIC ORIGIN? (*Check "No" or appropriate "Yes" box*)
Check "No" or check the "Yes" box that best corresponds with the decedent's ethnic Spanish identity as given by the informant. Note that "Hispanic" is not a race and item 53 must also be completed. Do not leave this item blank. With respect to this item, "Hispanic" refers to people whose origins are from Spain, Mexico, or the Spanish-speaking Caribbean Islands or countries of Central or South America. Origin includes ancestry, nationality, and lineage. There is no set rule about how many generations are to be taken into account in determining Hispanic origin; it may be based on the country of origin of a parent, grandparent, or some far-removed ancestor. Although the prompts include the major Hispanic groups, other groups may be specified under "other". "Other" may also be used for decedents of multiple Hispanic origin (e.g. Mexican-Puerto Rican). **Information in this section will not appear on the certified copy of the death certificate. This information is needed to identify health problems in a large minority population in the United States. Identifying health problems will make it possible to target public health resources to this important segment of our population.**

ITEM 53. RACE (*Check appropriate box or boxes on death certificate*)
Enter the race of the decedent as stated by the informant. Hispanic is not a race; information on Hispanic ethnicity is collected separately in item 52. American Indian and Alaska Native refer only to those native to North and South America (including Central America) and does not include Asian Indian. Please specify the name of enrolled or principal tribe (e.g., Navajo, Cheyenne, etc.) for the American Indian or Alaska Native. For Asians check Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, or specify other Asian group; for Pacific Islanders check Guamanian or Chamorro, Samoan, or specify other Pacific Island group. If the decedent was of mixed race, enter each race (e.g., Samoan-Chinese-Filipino or White, American Indian). **Information in this section will not appear**

on the certified copy of the death certificate. Race is essential for identifying specific mortality patterns and leading causes of death among different racial groups. It is also used to determine if specific health programs are needed in particular areas and to make population estimates.

ITEMS 54 AND 55. OCCUPATION AND INDUSTRY
Questions concerning occupation and industry must be completed for all decedents 14 years of age or older. This information is useful in studying deaths related to jobs and in identifying any new risks. For example, the link between lung disease and lung cancer and asbestos exposure in jobs such as shipbuilding or construction was made possible by this sort of information on death certificates. **Information in this section will not appear on the certified copy of the death certificate.**

ITEM 54. DECEDENT'S USUAL OCCUPATION
Enter the usual occupation of the decedent. This is not necessarily the last occupation of the decedent. Never enter "retired". Give kind of work decedent did during most of his or her working life, such as claim adjuster, farmhand, coal miner, janitor, store manager, college professor, or civil engineer. If the decedent was a homemaker at the time of death but had worked outside the household during his or her working life, enter that occupation. If the decedent was a homemaker during most of his or her working life, and never worked outside the household, enter "homemaker". Enter "student" if the decedent was a student at the time of death and was never regularly employed or employed full time during his or her working life. **Information in this section will not appear on the certified copy of the death certificate.**

ITEM 55. KIND OF BUSINESS/INDUSTRY
Kind of business to which occupation in item 54 is related, such as insurance, farming, coal mining, hardware store, retail clothing, university, or government. DO NOT enter firm or organization names. If decedent was a homemaker as indicated in item 54, then enter either "own home" or "someone else's home" as appropriate. If decedent was a student as indicated in item 54, then enter type of school, such as high school or college, in item 55. **Information in this section will not appear on the certified copy of the death certificate.**

NOTE: This recommended standard death certificate is the result of an extensive evaluation process. Information on the process and resulting recommendations as well as plans for future activities is available on the Internet at: http://www.cdc.gov/nchs/vital_certs_rev.htm.

National Mortality Followback Survey. A survey of a sample of the providers of information on death certificates taken annually to secure further information about the characteristics of decedents.

It may be noted that some of these files are derived by linking two other files. Linking files involves matching them on a case-by-case basis to derive a new

integrated file. Linking the records on mortality with other files provides a powerful and efficient means to augment their analytical potential for health research. NCHS has developed a record-linkage program for its mortality data, combining them with its natality (birth) data, population-based surveys, and various administrative databases, such as Medicare enrollments.

Other Administrative Records Systems

Administrative records may be maintained as a surveillance device for monitoring changes in health conditions or health care utilization. Such records include disease registries, twin registries, multigenerational pedigrees and genealogies, Social Security files, records of autopsies, and health-care utilization records, such as records of patients' visits to offices of health-care providers, out-patient clinic services, emergency room visits, hospital inpatient stays, and hospital discharges.

Disease Registries

State and provincial health departments, and public and private research organizations, often maintain files on communicable diseases and selected chronic health conditions. For example, the U.S. National Cancer Institute supports a Cancer Registry (SEER cancer registries) and the U.S. National Heart, Lung, and Blood Institute supports a Pediatric Cardiomyopathy Registry. The SEER (Surveillance of Epidemiological and End Results) program is a comprehensive population-based reporting system for cancer cases. The Pediatric Cardiomyopathy Registry is a national registry for children with different forms of cardiomyopathy, designed to measure the relative frequency of diseases and disorders of the heart muscle, describe the survival experience of the patients, advance knowledge of its causes, and identify new diagnostic and therapeutic procedures (New England Research Institute 1996). The U.S. National Institute of Diabetes and Digestive and Kidney Diseases in conjunction with the U.S. Centers for Medicare and Medicaid Services supports the U.S. Renal Data System, a national data system that collects, analyzes, and distributes information about end-stage renal disease (ESRD) in the United States.

Twin Registries

Several countries of northern Europe (e.g., Denmark, Sweden, and Finland) maintain twin registries, that is, lists of twins and their demographic characteristics, with life-course information regarding their health. One of the largest registries of twins is the Swedish National Twin Registry, which includes records of nearly 25,000 sets of identical twins born in Sweden between 1886 and 1959. Identical twins are

same-sex twins that developed from a single ovum and that share identical genetic "blueprints;" whereas fraternal twins share only 50% of their genes on average. Analysis of the records of identical twins is the "perfect" tool for study of the relative contribution of genes and environment to health outcomes and longevity, especially when the identical twins were separated at birth.

Multigenerational Pedigrees or Genealogies

Data on multigenerational pedigrees or genealogies have become increasingly available in recent decades. Survival data from large pedigrees contain valuable information with which to assess the relative role of genetics and environment in longevity (Garibotti et al. 2006) and the role of selected demographic characteristics in longevity (Gavrilov et al. 2002). An example of such a genealogical database is the Utah Population Database (UPDB). It contains over eight million records, including the genealogies of the founders of Utah and their descendants. The families have been linked across generations, in some cases up to seven generations. As an active genealogy, the UPDB is constantly adding new families and their members. The UPDB has been linked to other sources of data, including birth and death certificates, driver's license registrations, and the Utah Cancer Registry. Consequently, the UPDB now holds data for migrants to Utah and their descendants and includes more than 1.8 million individuals born from the early 1800s to the mid-1900s that are linked into multigenerational pedigrees (Garibotti et al. 2006).

Social Security Records

Before the modern Social Security program, the most widespread form of assistance to the elderly in the United States was the Union Army pension program (Costa 2000). It became a disability and pension program for veterans with the passage of the Act of 1890 and physical examinations were administered to ascertain the degree of disability. This illustrates a type of early historical source of health and mortality data based on administrative records.

Today the Social Security Administration maintains several files of value for health research. Some records are useful mainly for the information that they provide when they are linked with other records. Some records are directly valuable for providing information on deaths, disability, or survival to extreme old age. The files now maintained are the Applications file, Summary Earnings Record File, the Master Beneficiary Record File, the Master Death File, and the Continuous Work History Sample. Social Security files have the considerable advantage that the information on them is, to a large extent, validated, particularly with regard to age and disability status. Moreover, disability and death are uncommon events, and Social Security files alone or merged with other files yield gigantic microdata databases for study of these phenomena, In addition, Social Security files are

longitudinal files so that causality is more easily analyzed with them than with the cross-sectional files from other sources.

Form SS-5, the Application for a Social Security Number, is the form filled out when an individual applies for a Social Security number. It is the original source of data on age, sex, and race (voluntary). It calls for name, date of birth, race, sex, and some geographic information. It is useful mainly for the information it provides to expand what is available on other records.

The Summary Earnings file maintains and accumulates information on earnings. The earnings record contains the demographic information given in the application for the Social Security number, lifetime covered earnings, quarters of coverage, and death information. The earnings record is retained for every person who has ever obtained a social security number. It provides a yearly record since 1950 of individual earnings up to the taxable maximum for all jobs in covered employment, and since 1978 a record of complete earnings for both covered and noncovered employment.

The Master Beneficiary Record File is created when a person applies for benefits. This record also contains demographic information, historical and current information about entitlement to benefits, benefits paid, dependent beneficiaries, disability, and vital status, along with information on entitlement to hospital and medical insurance. When the Social Security Administration is informed of a death, this information is added to the Master Beneficiary Record file as well as the Summary Earnings file. Mortality rates, work disability rates, and the survival history of very long-lived persons can be developed from this file.

The Continuous Work History Sample (CWHS) is a statistical file that carries the individual's information from the administrative files and longitudinal work-related information such as industry and geographic location of the employer.

As noted, deaths are routinely posted to the earnings records and beneficiary records. The Social Security Administration also maintains a Death Master File, which provides information on deaths of SSA beneficiaries. Studies of the completeness of death reporting in these files show that reporting of deaths is virtually complete.

Health-Care Utilization Records

Health-care utilization records are maintained by the agencies that administer health programs. The U.S. Centers for Medicare and Medicaid Services maintains records of health conditions reported for Medicare claims and funds disbursed to patients and health providers. Medicare also maintains records of deaths of Medicare subscribers, which, in combination with counts of subscribers, permits calculation of death rates at the older ages. State departments of health maintain records for Medicaid claims. Medicare is the program of national health insurance for the elderly and Medicaid is the national/state program of public medical care for the indigent. Health-care utilization records – hospital discharge records, outpatient

clinic records, and records of physicians' visits – are the basis for securing data for several of the health-care studies sponsored by the National Center for Health Statistics.

Autopsies

Records of the post-mortem examination of deceased persons have long been an essential tool for quality control of medical care and for evaluating and improving the quality of cause-of-death reporting on death certificates. The most evident value of an autopsy is its use in comparing a physician's prior clinical diagnosis with the precise, anatomical cause of death determined by the autopsy. The autopsy is the "gold standard" for evaluating new and emerging medical technology; the autopsy can identify the potential pitfalls of the new technology with enhanced resolution. The autopsy is the tool of choice in determining whether a patient received the correct treatment for the disease he or she has.

We learn from autopsy studies that the patient does not always receive the correct treatment. For example, Shanks et al. (1990) found that in 213 cases of postoperative death there were major discrepancies in clinical diagnoses for diseases that were treatable; proper diagnoses could have affected survival in 21% of the cases.

An autopsy also enables one to assess comorbidity, that is, the identification of the several other conditions associated with the patient's death in addition to the underlying cause of death. This is important knowledge for epidemiological and medical studies, particularly for understanding the course of various diseases and possible treatments, and for demographic studies analyzing mortality in connection with the preparation of mortality projections. In sum, the autopsy furnishes information for improving patient care, resolving legal contests relating to cause of death, and monitoring the quality of mortality data needed for numerous public health planning purposes.

In spite of such important uses, autopsies have been conducted with decreasing frequency in the United States in the past half century. In particular, autopsy rates in medical institutions have fallen to levels that could be deemed prejudicial to the quality of medical care and of mortality data. The conduct of autopsies declined from 41% of hospital deaths in 1961 to 5%–10% in the mid-1990s (U.S. NCHS 2001). Currently the Residence Review Committee for Internal Medicine requires autopsies in at least 15% of deaths in accredited residency programs but, from 1991 to 1994, less than half of the internal medicine programs reviewed for accreditation met these minimal requirements (Schatz 1995).

Several reasons can be given for this decline in the use of autopsies. Perhaps the most important reason is that physicians do not ask for autopsies, believing that they do not serve any current purpose. Autopsies have become very expensive and the costs are not reimbursed by third-party payers. Finally, there is the elimination of the requirement for autopsies for hospital accreditation in 1971 by the Joint Committee on Accreditation of Hospitals. We consider the use of autopsies in evaluating the data on cause of death on the death certificate further in Chap. 3.

General National Sample Surveys and National Health Surveys

U.S. General National Sample Surveys

The U.S. Bureau of the Census sponsors and conducts four national sample surveys. They are the American Community Survey (ACS), the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), and the American Housing Survey. These surveys occasionally carry questions on health to accompany the principal questions on population, labor force participation, income, and housing.

American Community Survey

The American Community Survey is a nationally representative sample survey of the U.S. population designed to provide annual estimates of the demographic, social, economic, and housing characteristics of the U.S. population and its geographic subdivisions. In 2010 and future decennial census years it will replace the long form (i.e., the sample inquiries) of the 2000 decennial census and so it will become part of the decennial census. The 2000 census used two basic forms, each covering essentially different subjects: (1) the short form, which enumerated the entire population and obtained data on a few basic demographic characteristics, and (2) the long form, which obtained demographic, social, economic, and housing data for a sample of households.

The data are collected by self-enumeration using mail-out/mail-back methods. As fully implemented in 2006, after incremental development over several years, the ACS includes approximately 250,000 households per month, or three million households each year. It will provide estimates every year for all states, as well as for all cities, counties, metropolitan areas, and population groups of 65,000 persons or more. For smaller areas, sample data will be accumulated over 3–5 years, depending on the population of the area. Areas of 20,000–65,000 will require data accumulated over 3 years. For rural areas, city neighborhoods, and population groups of less than 20,000 persons, data will be accumulated over 5 years. Estimates for 3- and 5-year periods for these smaller areas will begin to be released in 2008 and 2010, respectively. The 2005 ACS data were also released for Public Use Microdata Areas (PUMAs), which are special areas of about 100,000 population that partition a state.

As noted earlier, the ACS has been securing data on various types and degrees of disability – sensory disability, mental disability, physical disability, work disability, transportation disability, and self-care disability. The ACS has asked the identical questions as in the 2,000 decennial census, except for the change to a lower limit of age 15 for some of the questions. For the 2010 census, as noted earlier, however, there will be the separation of the sensory disability question to distinguish responses on visual disability and aural disability and the elimination of the work disability question.

Current Population Survey

The Current Population Survey is a national survey of about 55,000 households representing the entire United States, selected states, and selected metropolitan areas. The CPS is the source of many types of demographic, social, and economic data for households and individuals. In addition to the basic questions on labor force participation each month, the CPS annually includes supplemental questions on such topics as income, school enrollment, fertility, migration, marital status, household characteristics, and educational attainment, and occasionally includes questions on health insurance, disability, tobacco usage, and food sufficiency.

Survey of Income and Program Participation

The Survey of Income and Program Participation, another one of the national sample surveys conducted by the U.S. Census Bureau, provides data on income and participation in government-transfer programs of individuals and households in the United States and on a host of related subjects. This survey secures detailed data on cash income, noncash income, taxes, assets, and liabilities, and participation in government transfer programs, including Medicare and Medicaid. Occasionally the survey carries a question on health status such as disability.

American Housing Survey

The U.S. Census Bureau also conducts the American Housing Survey (AHS). It collects data on U.S. housing, including single-family homes, apartments, and mobile homes. Data are secured on vacancies, housing costs, equipment and fuels, size of unit, last housing move, and housing and neighborhood quality. National data are collected in odd-numbered years, and data for each of 47 selected metropolitan areas are collected about every 4 years, averaging 12 metropolitan areas being surveyed each year. The national sample covers an average of 55,000 housing units and the survey for each metropolitan area covers 4,800 or more housing units. The American Housing Survey does not collect any health data per se but some of the data collected can serve as quality-of-life indicators in health studies and as correlates of health-status data secured from other sources.

National Health Surveys and National Health Care Surveys

National health surveys are surveys that are specifically designed to secure health information and hence they carry a large body of health-related questions. In addition, they carry a number of questions on demographic, social, and economic items useful in characterizing the population surveyed and in determining socioeconomic

differentials in health, disability, and longevity. In more recent years, biological and anthropometric measures have been included in an effort to identify the ways in which socioeconomic and demographic characteristics are associated with survival, especially long-term survival (Crimmins and Seeman 2004). An increasing number of countries conduct periodic sample surveys specifically devoted to health and designed to be representative of their national populations.

U.S. National Health Surveys and National Health Care Surveys

The U.S. National Center for Health Statistics sponsors many health surveys and health-care surveys, including the National Health Interview Survey, the National Health Examination Survey, the National Health and Nutrition Examination Survey, the National Hospital Discharge Survey, and the National Nursing Home Survey. These surveys secure their data by interviewing members of the population, conducting physical examinations, examining medical records, or interviewing staff and accessing records. A list of the principal surveys now being conducted, with brief descriptive notes, is given below.

National Health Surveys. The National Health Surveys include the National Health Interview Survey and the National Health and Nutrition Examination Survey.

National Health Interview Survey (NHIS). A nationally representative, continuing sample survey of the civilian noninstitutional population of the United States conducted since 1957 and intended to monitor the health situation in the country. Data are collected in personal household interviews on health care access and insurance, health status, and health outcomes in relation to demographic and socioeconomic variables. Specifically, data are obtained on illnesses (acute and chronic conditions), injuries, impairments, and activity limitations caused by chronic conditions. NHIS is a household, multistage probability sample survey conducted by interviewers of the U.S. Census Bureau. Each year the list of items of inquiry is reviewed and special topics are added or deleted. The sample interviewed in 2004 consisted of 36,579 households, which yielded 94,460 persons in 37,466 families. The household response rate for the survey in 2004 was 86.9%. (www.cdc.gov/nchs).

National Health and Nutrition Examination Survey (NHANES). A nationally representative sample of the civilian noninstitutional population of the United States. NHANES provides information on the health and nutritional status of the population on the basis of household interview data and physical examinations. The survey covers dietary intake, anthropometric measurements, data on body composition, self-reported and measured physical activity, and measures of cardiovascular fitness. NHANES is based on a stratified, multistage, probability cluster sample design. The surveys were conducted periodically from 1960 to 2006. The various waves of NHANES are as follows:

NHANES I, 1960–1962
Epidemiologic Followup Study, 1971–1975

NHANES II, 1976–1980
NHANES III, 1988–1994
NHANES, 1999–2006

Hispanic Health and Nutrition Examination Survey. A special sample for securing examination data for the Hispanic population.

National Health Care Surveys (NHCS). Consists of several separate surveys that provide data on the characteristics of health-care providers and establishments, the utilization of health-care services, and the health-care delivery system. They include the National Hospital Discharge Survey, the National Nursing Home Survey, National Home and Hospice Care Survey, and the National Ambulatory Care Surveys.

National Hospital Discharge Survey (NHDS). Conducted annually since 1965, this survey provides data on the Nation's use of inpatient care on the basis of a sample of inpatient records secured from a national probability sample of the nation's general nonfederal hospitals and other medical-record organizations. Data are obtained on admission and discharge dates, medical diagnoses, procedures performed, and personal characteristics of the patient. Because persons with multiple discharges during a year can be sampled more than once, NHDS produces estimates for discharges, not persons. For 2004 the sample consisted of 500 hospitals, of which 439 were in-scope (eligible) hospitals, and which responded to the survey (92% response rate), providing data for 371,000 discharges.

National Nursing Home Survey (NNHS). A continuing series of national sample surveys of nursing homes, their residents, and their staff. The survey employs self-administered questionnaires and interviews with administrators and staff and includes a sample of 1,500 facilities. Six such surveys were conducted between 1973–1974 and 1999. Nursing homes are defined as facilities with three or more beds that routinely provide nursing care. The facilities may be freestanding or parts of hospitals, retirement centers, or similar institutions, as long as the facility maintains separate residential and financial records. The 1977 survey also included personal care or domiciliary care homes. The NNHS secures information both about the facility and about the residents. For the facility, the data collected include size, ownership, Medicare/Medicaid certification, occupancy rate, days of care provided, and costs. For the residents, the data collected include demographic characteristics, health status, and services received. These data are provided by nurses who can access the medical records and have personal knowledge of the residents.

National Nursing Assistant Survey (NNAS). A national survey of certified nursing assistants working in nursing homes, conducted in conjunction with the National Nursing Home Survey.

National Home and Hospice Care Survey (NHHCS).

National Ambulatory Care Surveys. Records-based surveys that provide national estimates of the use of health care in the out-patient setting, that is, visits to office-based physicians and to hospital emergency and out-patient departments, in the previous 12 months.

National Survey of Ambulatory Surgery. Conducted during 1994–1996 and 2006; compiles data on ambulatory surgery performed in hospitals and free-standing surgery centers. The survey provides data on diagnoses, surgical and nonsurgical procedures, and length of stay.

National Ambulatory Medical Care Survey (NAMCS). Provides national estimates of visits to physicians' offices.

National Hospital Ambulatory Medical Care Survey (NHAMCS). Provides national estimates of visits to hospital outpatient and emergency departments.

The U.S. National Center for Health Statistics sponsors some other specialized health surveys, such as:

National Survey of Family Growth (NSFG). A national representative sample survey of the household population, covering men and women of childbearing ages. Originally the survey interviewed only women (during the first five surveys, 1973–1995), but in 2002 NSFG interviewed a national sample of over 12,500 men and women 15–44 years of age. The survey collects data on various topics related to pregnancy, childbearing, family formation, and reproductive health.

National Immunization Survey (NIS). Data from the NIS have been published in the *Mortality and Morbidity Weekly Report* series of NCHS since 1995.

Surveys of Health-Care Expenditures

Surveys sponsored by two other federal agencies secure data on health-care expenditures. They are the National Medical Expenditure Survey, the National Long-Term Care Survey, and the Medicare Current Beneficiary Survey.

National Medical Expenditure Survey (NMEPS). A nationally representative annual survey of the U.S., civilian noninstitutionalized population, sponsored by the Agency for Healthcare Research and Quality. It consists of four components: Household component, medical-provider component, insurance component, and nursing-home component. In 2002 the sample size of the household component was approximately 15,000 households, representing 37,000 people. NMEPS collects data on health services' use by respondents: source of payment, the cost, scope, access to health services, and use of private health insurance held by the respondents. (www.meps.ahrq.gov/mepsweb).

National Long-Term Care Survey (NLTCS) is a longitudinal sample survey covering about 20,000 persons pulled from the file of Medicare enrollees and maintained by the Centers for Medicare and Medicaid Services (CMS). The survey was initiated in 1982 and follow-up surveys were taken in 1984, 1989, 1994, 1999, and 2004. The NLTCS provides both cross-sectional and longitudinal data on the health and functional status of Americans aged 65 years and over. It provides data on medical conditions, the prevalence and patterns of both physical and cognitive functional limitations, use of health-care services, kinds and amounts of formal

and informal long-term-care services received, caregivers, and for those who have died, data on next-of-kin. It seeks to study changes in the health, functional status, health expenditures, service use, and the availability of personal, family, and community resources for caregiving.

Medicare Current Beneficiary Survey (MCBS). The MCBS is a longitudinal panel survey of Medicare beneficiaries 65 years old and over and disabled Medicare beneficiaries under 65 years old. 1998, 1999, 2000, and 2001 surveys: The survey includes the institutional population as well as community residents, covering about 12,000 completed interviews each year. Three survey rounds are conducted each year. The survey employs a rotating panel design, one-third of the sample being replaced each year. The survey collects data on demographic characteristics as well as data on health and functioning status.

Use of Combinations of Surveys, and Vital Statistics and Surveys

To study the health characteristics of certain groups or events, such as diabetes and obesity, injuries, women, role of religion in health, or children, data from more than a single NCHS survey may have to be taken into account. For example, to analyze national trends on injuries, one may refer to the National Health Interview Survey, the National Health Care Survey, and vital statistics. To analyze the state of women's health, all of the NCHS data systems may be examined, although some systems have specific sections and/or periodic supplements that might be of special interest. For analyzing trends in the health status and health-care use of older Americans, most of the data would come from NCHS sources, especially the National Health Interview Survey, but other data would come from the Centers for Medicare and Medicaid Services, particularly the Medicare Current Beneficiary Survey. Information on religious or spiritual concerns and health can be found in NHANES 2005–2006 and NHANES III (attendance at religious services), NHIS 2002 (complementary and alternative medicine), and NSFG (religious practices). Similar combinations of NCHS survey data provide a wealth of data on mental health, pain, medication use, quality of health care, and other specialized areas.

U.S. Data Sets on Health and Mortality

The National Center for Health Statistics, the Centers for Medicare and Medicaid Services, and other U.S. government agencies have individually or collaboratively prepared data sets on specialized health topics. A selected list is as follows:

National Warehouse on Trends in Health and Aging. This is a web-based comprehensive source of detailed information on health-related behavior, health conditions, health-care utilization, health insurance, and health expenditures of older Americans from the data systems of NCHS, U.S. Census Bureau, Centers for Medicare and Medicaid Services, and other U.S. government agencies. The

data can be accessed on the NCHS Warehouse web site, and Beyond 20/20 software can be used to manipulate, extract, and download the data. For example, one can extract data on trends in the prevalence and treatment of coronary artery disease and diabetes and on the mortality from these diseases. (www.cdc.gov/nchs/agingact.htm.)

Data Warehouse on Health Data for All Ages. This is a web-based source of information that includes data on pregnancy and birth, health conditions, risk factors, health-care access and use, and mortality.

Linkage of National Health Interview Survey with the National Death Index. The most recent linkage is between NHIS survey years 1986–1999 and deaths through December, 2001. This linkage provides a longitudinal component to the NHIS that allows for ascertaining current vital status and cause of death for decedents. The linkage makes possible the use of NHIS data for analysis of survival patterns, mortality rates, and life expectancy while providing data from the NHIS for use as covariates.

Healthy Women: State Trends in Health and Mortality.
Minimum Data Set of the Centers for Medicare and Medicaid Services.
The State and Local Area Integrated Telephone Survey (SLAITS).

Other U.S. National Surveys

Several national surveys sponsored by U.S. agencies focus on health but are restricted to specific age segments of the population, or focus on subjects other than health but include health items. Here is a list of the more important ones:

First Longitudinal Study of Aging (LSOA I) and *Second Longitudinal Study of Aging* (LSOA II). A study of older Americans sponsored jointly by the National Center for Health Statistics and the National Institute on Aging. The study population is an area probability sample of individuals identified from census records. LSOA I was taken in 1984, 1986, 1988, and 1990. The initial interviews of LSOA II were conducted between 1994 and 1996 as part of the National Health Interview Survey on Disability, Phase II. LSOA II was designed (1) to provide information on (a) the sequence of health events among the elderly and their consequences, (b) the utilization of medical-care services, and (c) the causes and correlates of changes in functioning, and (2) to replicate LSOA I, in order to determine whether there have been changes in the extent of disability and impairments between the 1980s and 1990s. The LSOA II sample includes 9,447 persons and is a nationally representative sample of noninstitutionalized persons 70 years of age and over in 1995. Two waves of interviews were conducted between 1994 and 2000 in addition to the initial interview, the first in 1997 and 1998 and the second in 1998 and 1999.

Health and Retirement Study (HRS). A nationally representative longitudinal survey of the elderly population that focuses on retirement and health. It is conducted by the Survey Research Center, University of Michigan, under the sponsorship

of the National Institute on Aging. The first cohort under study was the 1992 cohort. It consisted of persons born between 1931 and 1941 (aged 51–61) and their spouses. They were reinterviewed every 2 years. Another cohort study, the Survey of Assets and Health Dynamics among the Oldest Old (AHEAD), consisting of persons 70 years and over (born in 1923 and before), was instituted in 1993. The studies were merged in 1998 with one another and two other cohorts, so that the study now represents the entire population born in 1947 and before (51 and older). Through 2002 114,000 interviews have been conducted. In 2004 the birth cohorts of 1948–1953 were added. The study collects information on family structure, health conditions, disability, nursing home use, health insurance and use, work, income, wealth, pensions, and participation in government programs. (*http://hrsonline.isr.umich.edu*)

Children's Well Being. Covers children in three age groups, under 6 years, 6–11 years, and 12–17 years, and includes questions on the health of children in the household, care of the children by nonfamily members, and the quality of the neighborhood.

National Health and Social Life Survey
National Longitudinal Survey of Adolescent Health
National Survey of Adolescent Males
National Survey of Children with Special Health Care Needs.
National Survey of Children's Health
Youth Risk Behavior Survey.

Some Tools for Accessing and Analyzing U.S. Data on Health

Some tools for accessing and analyzing U.S. data on health and mortality include the following:

Public use microdata samples (PUMS): PUMS files are available for most of the surveys listed above. These can be accessed on the Internet or CD-ROM.

Special files such as Beyond 20/20, CDC WONDER, WISCARS, and others.

Statistical Export and Tabulation System (SETS) software. SETS can be used to view, search, tabulate, manipulate, present, and export/extract data through the internet. This tool can be employed in connection with analysis of data from many of the NCHS surveys.

ANDRE. NCHS's remote access system.

SAS, SPSS, STATA, SUDAAN software for computing variances and analyzing the data. For example, SUDAAN is a statistical software package that provides procedures for analyzing survey data and calculating variances of surveys with complex sample designs, such as multistage and cluster designs.

National Health Surveys in Other Countries

Most other industrialized countries take national health surveys like the U.S. National Health Interview Survey. Examples of population-based surveys in other countries are the National Population Health Survey of Canada, the National Health Interview Survey in the Netherlands, the Göteborg Study in Sweden, and the Health Survey for England. Canada and the United States collaborated on the 2002–2003 Joint Canada/United States Survey of Health (JCUSH), with questions from the U.S. and Canadian health surveys. Canada has a new Health Measures Survey (CHMS), patterned after U.S. NHANES. Some specialized surveys of other industrialized countries, mainly of the elderly population, are the English Longitudinal Study on Aging, the Nihon University Japanese Longitudinal Survey of Aging, and the Longitudinal Aging Study of Amsterdam. An interesting example of the use of the national health survey to measure comparative health of important segments of a national population is the study of the differences in health and longevity between the Swedish-speaking and Finnish-speaking populations in Finland (various studies of Fjalar Finnäs and Jan Saarela).

The European Union (Eurostat) collects much demographic and socioeconomic data on a uniform basis for the countries of the European Union, including data on disability. The European Union Community Household Panel (ECHP) compiled data on the age profile of disability for the EU-15 for the years 1994 and 1996. Separate data were secured for men and women for two categories of disability (moderately disabled and severely disabled), for all member countries of the European Union. The variation in the incidence rate of disability was found to be very high in spite of the fact that exactly the same question was asked (in some cases even in the same language, as in the United Kingdom and Ireland). This experience suggests how difficult it is to measure disability and to secure comparable data on its incidence.

Numerous health surveys have been conducted in the less developed countries. The U.S. Agency for International Development and other organizations have funded a number of Demographic and Health Surveys (DHS), including special surveys on AIDS (AIS) and malaria (MIS), that were carried out by ICF Macro, of New Carrolton, Maryland. Between 1985 and 2010 190 surveys were conducted in 81 countries of Africa, Asia, the Middle East, Latin America, and the Caribbean (Table 2.1). Typically a DHS secures demographic and health data from a nationally representative sample of between 5,000 and 30,000 households, with a focus on women between the ages of 15 and 49. The surveys deal with such subjects as contraceptive practice, breastfeeding, lifetime reproductive behavior, health characteristics of children (e.g., height, weight, immunization, diarrhea, fever), women's work history, and background information about the husband. These data are used to provide information on the state of health of women of reproductive age and their children in these countries.

Table 2.1 Demographic and health surveys conducted and completed by ICF Macro and National Agencies: 1985–2010

	Number of countries	Number of surveys
All areas	81	190
Africa	44	119
Western Africa	14[a]	43
Middle Africa	7	11
Eastern Africa	14	42
Southern Africa	5	8
Northern Africa	4	15
Asia	20	29
West Asia	5	12
South Central Asia	9	19
Southeast Asia	6	16
Latin America and Caribbean	15	40
Caribbean	3	10
Central America	5	8
South America[b]	7	22
Eastern Europe	2	2

Source: ICF Macro website www.measuredhs.com
Includes AIDS Indicator Surveys (AIS) and Malaria Indicators Surveys (MIS) in addition to Demographic and Health Surveys (DHS)
[a]One survey was taken in Ondo State, 1986
[b]Includes Northeast Region of Brazil, 1991

The World Health Organization, the United Nations, the U.S. National Institutes of Health, and the U.S. Centers for Disease Control and Prevention have also sponsored health surveys in less developed countries. These surveys are cross-sectional, secure data on health by self-reports of respondents, and usually relate to specific diseases and the elderly population.

The national health survey in China was taken as a joint project of several universities, both in the United States and China. The Chinese Health and Family Life Survey (CHFLS) is a collaborative project of the University of Chicago/NORC, China's Renmin University, Peking Union Medical College, and the University of North Carolina, with primary funding support by the U.S. National Institute of Child Health and Human Development. This survey was carried out in 1999–2000 in 18 provinces of China (excluding Tibet and Hong Kong), and is representative of the population of China aged 20–64 years of age. Over 3,800 completed interviews were obtained, with a response rate of 76%. The survey topics relate largely to sexual behavior (Population Association of America 2003).[1] China also conducts a Longevity Healthy Survey.

[1] A public dataset is available from the Population Research Center, University of Chicago/NORC and is downloadable from their website (src.uchicago.edu/prc/chfls.php).

Epidemiological Studies

Nature and Types of Epidemiological Studies

The National Institutes of Health sponsor several community-based observational studies relating to health and the associated risk factors that as a class are denominated epidemiological studies. An epidemiological study is a type of sample survey concerned with the distribution of diseases, injuries, and impairments in human populations and the possible risk factors associated with them. The goal is to identify the determinants of the diseases and to devise programs of disease prevention and control. Epidemiological studies are a principal tool of community medicine and public health, fields of medicine which have the community as their primary object of concern and treatment. Unlike the health surveys described earlier, the samples used in epidemiological studies are not always representative of the total population.

There are several different types of epidemiological studies. In case control studies, a group of patients with a disease under study is matched by a group of healthy persons (the "controls") on the basis of a variety of criteria including age, sex, race, and socioeconomic characteristics. Such studies try to identify other characteristics that distinguish the two groups. Case-control studies may be less stringent in their design by comparing a group with the disease and a group free of the disease, without initial matching of the characteristics of the participants, and by seeking differences between the groups only subsequently.

In longitudinal studies, a group that does not have the disease under study is selected for the study of the disease, base data on demographic, socioeconomic, and health characteristics are secured, and the group is followed over time to see who develops the disease and what their distinguishing characteristics are. One can also conduct a longitudinal study of people who already have a disease as a way of determining what existing population characteristics contribute to subsequent changes in health status, the expression of related diseases, and the age-specific risk of death. When the same individuals are canvassed over time, the study is designated a panel study. When the same population group (e.g., a birth cohort) is tracked, the study is designated a cohort study. One could also employ a panel study to learn about the timing and intensity of physical changes associated with aging. Cross-sectional studies intended to provide information on changes associated with aging could give misleading indications of what health changes occur in real cohorts as they age because cross-sectional observations combine numerous birth cohorts in the same year that may have different characteristics and lifetime experiences. Illustrations of such problems are given in Chap. 5.

U.S. Epidemiological Studies

The National Cancer Institute conducts epidemiological surveillance studies of health and disease outcomes for various population groups. NCI's surveillance activities include studies of the access to cancer care, treatment options and survival, and the role of socioeconomic status in the incidence and survival outcomes of cancer. The National Heart, Lung, and Blood Institute conducts longitudinal epidemiological studies of the incidence of cardiovascular disease and associated factors.

Early Studies

Some of the important epidemiological studies were initiated several decades ago and continue to provide data on health and longevity. These include the Framingham Heart Study, the Baltimore Longitudinal Study of Aging, the Duke Longitudinal Study of Aging, the Alameda Study, and the Nurses Health Study.

Framingham Heart Study. Inaugurated in 1948 by the U.S. Public Health Service, this study was the first longitudinal study of its kind. It followed 5,209 healthy Framingham, MA residents between the ages of 30 and 60 over two decades to determine who succumbed to cardiovascular diseases. The volunteers reported to the study center every 2 years for 20 years. At each visit, they had extensive physical tests and filled out lengthy questionnaires. The research then worked backward to analyze the factors, e.g., diet, exercise, and smoking, that might be correlated with the risk of cardiovascular disease and death. The purpose of the study was to determine the then unknown factors that caused build-up of plaque in the arteries. The results showed, for the first time, that the higher the participants' blood pressure or cholesterol, the more fatty their diet, and the more sedentary their lifestyle, the greater the chance of their incurring cardiovascular disease. The Framingham Heart Study is now following the third-generation descendants of the original 5,209 volunteers. The new phase of the work looks at the interaction of genetics and known risk factors, such as obesity, in the development of heart disease.[2]

Baltimore Longitudinal Study of Aging (BLSA). This longitudinal study was begun in 1958 at the Gerontology Research Center, National Institute on Aging, in Baltimore, and has continued for half a century. Since its initiation, more than 1,500 males, ranging in age from 17 to 96, have volunteered to join the study; and since 1978, 700 women have volunteered. Every 2 years all the surviving participants take a large battery of tests in Baltimore. The main focus of the BLSA is to determine the trajectories of change in physiological biomarkers with increasing age, the physical, mental, and emotional correlates

[2]For a detailed history of the Framingham Heart Study, see Daniel Levy and Susan Brink, *A Change of Heart: How the People of Framingham, Massachusetts Helped Unravel the Mysteries of Cardiovascular Disease*, NewYork: Alfred A. Knopf, 2005.

of aging in healthy persons living in the community, the relation between cross-sectional and cohort changes in aging, variations in the timing of age changes among different individuals, and the differences between aging and disease. Some general findings of the study are that older persons vary more widely in many physiological and psychological measurements than do younger adults, vulnerability to disease increases and performance speed declines with advancing age, our bodily organs lose their efficiency at different rates as we grow older, and chronological age alone is a poor indicator of aging or performance because of the great individual variation (Hayflick 1994:140–148).

Duke Longitudinal Study of Aging. This longitudinal study was conducted in 1955–1976 and was designed to distinguish normal physiological changes due to aging from age-associated pathological changes. Aiming to describe changes in individuals with age over time, this study gave repeated (11) physiological and psychological tests to a small (267) selected sample of elderly persons during the 21-year study period. Analysis of these data reveals a process of mortality selection as the population ages, removing a relatively young, unhealthy subgroup through death and leaving a highly selected group of survivors at the advanced ages (Manton and Soldo 1992).

Alameda County Study. This study was conducted by the State of California and was designed to identify the risk factors associated with longevity. It was launched in 1965 with a representative sample of 6,928 adult residents of Alameda County aged 20 years and over, who completed an extensive questionnaire about the behavioral, social, and psychological aspects of their lives. Analysis of these data showed that 45-year-old men who practice seven healthful habits (i.e., exercising regularly, maintaining moderate weight, not eating snacks, eating breakfast, not smoking, not drinking excessively, sleeping at least 7 h a day) would gain several years of life over those practicing three or fewer of these habits (Belloc and Breslow 1972; Wiley and Camacho 1980; Berkman and Breslow 1983). A later analysis of the 17-year mortality of participants who were 60 years and over at the start of the study found that increased risk of death is associated with several of the same factors and being male (Kaplan et al. 1987).

Nurses' Health Study. The Nurses' Health Study is the largest and longest-running epidemiological study of women's health. The NHS is directed by researchers in the Cambridge/Boston medical community. It was started in 1976 to study the possible role of the use of oral contraceptives and cigarette smoking in the risk of major chronic diseases of women. The prevention of cancer has been the primary focus, but the study has produced new information on risks of cardiovascular diseases, diabetes, and many other conditions. The initial results represent the responses of 127,000 female registered nurses between the ages of 30 and 55. Dietary assessments, aspirin use, and colon examinations were added in 1980. The scope of the NHS was extended in 1989 to cover 238,000 nurse-participants for evaluating the effect of various lifestyle factors, such as exercise and diet, on women's health. Participants are sent lengthy questionnaires every 2 years calling for medical histories and reports of dietary practices and major life events that occurred during the preceding 2 years. The average response rate is 90%.

More Recent Studies

Several newer epidemiological studies have been launched. Here is a list of and brief notes on some leading studies.

Established Populations for Epidemiological Studies for the Elderly (EPESE). Longitudinal surveys of three community-based cohorts in Durham, NC, East Boston, MA, and New Haven, CT. These surveys focused on persons 65 years and over, included larger samples than previous surveys of the elderly (e.g., Baltimore study, Duke study), and were representative of the local populations covered in the sample.

McArthur Studies of Successful Aging. A longitudinal study of relatively high functioning men and women aged 70–79. Participants were subsampled, on the basis of age and physical and cognitive functioning, from the three community-based cohorts in Durham, N.C., East Boston, MA, and New Haven, CT that are part of the Established Populations for Epidemiological Studies of the Elderly (EPESE). Age-eligible men and women were screened on the basis of four criteria of physical functioning and two criteria of cognitive functioning, to identify those who functioned in the top third of their age group. Collection of baseline data was completed between May 1988 and December 1989. A cohort of 1,313 subjects met the screening criteria. The study included an interview that secured data on physical and cognitive performance, health status, and social and psychological characteristics. The cohort was reinterviewed in 1991 and again in 1996.

Coronary Artery Risk Development in Young Adults (CARDIA). Long-term studies of cardiovascular risk for young adults as they age, covering a sample of 5,000 persons. The CARDIA study is sponsored by the National Heart, Lung, and Blood Institute (NHLBI) and is carried out in four U.S. areas, Birmingham, AL, Chicago, IL, Minneapolis, MN, and Oakland, CA. A recent analysis reviewed data on 2,909 of the CARDIA participants over the 10 years between 1985–1986 and 1995–1996, with the purpose of examining the role of fiber as compared to fat and other dietary elements in producing high levels of insulin in the blood, obesity, high blood pressure, and high serum cholesterol. This particular analysis showed that high fiber diets may protect against obesity and cardiovascular disease in healthy young adults by lowering insulin levels (Lenfant 1999).

Randomized Clinical Trials

Studies to evaluate the efficacy of a treatment protocol compared with an alternative treatment or none at all are called clinical trials. Clinical trials provide the treatment guidelines used in the practice of clinical medicine. To test the efficacy of a treatment protocol, clinical trials must be employed for humans, inasmuch as animal-type experiments are not appropriate for them. Typically the reporting units in clinical

trials are individuals from whom certain information is secured and who are subjected either to some treatment or no treatment.[3]

In designing these as in designing other health surveys, the investigator usually secures data on a basic set of descriptive variables, such as age, sex, race, and socioeconomic status, and a more specialized set of variables relevant to the specific study.

In the research design called randomized clinical trials (RCT), an experimental group receives the treatment being tested (e.g., an estrogen tablet) and another group, the so-called control group, receives a placebo or dummy treatment (e.g., a sugar tablet). Placebos may be defined as harmless inactive substances or procedures resembling the real medical treatment used in the clinical trial that allow the investigators to learn whether the medical treatment being applied works better or no better than a preexisting or alternative medical treatment, or no medical treatment. The placebo can cause a real or apparent improvement in the patient's condition due to the expectation of a beneficial effect by the investigator or the patient. Randomized clinical trials are designed as double-blind studies so that neither the subjects nor the researchers know who is in the experimental group and who is in the control group. In double-blind studies when the treatment is a medication, only the pharmacist knows "who gets what," and the other members of the health care and research staff are not informed. Double-blind studies are designed to prevent anyone from influencing the results.

In the type of clinical trial called an interventional trial with sequential design, data are secured for pairs of patients, each person in the pair receives one of the two treatments being compared, and the responses have a binary (i.e., yes or no) outcome. Discordant pairs, where the treatment effect differs for the members of the pair, provide the basis for preferring one treatment over the other. In this protocol a small number of examinations of the data are made during the course of the trial and repeated significance testing is carried out. The investigator may discontinue the use of the treatment during the trial, even giving it to all subjects if a sufficiently large favorable difference is observed during the course of the trial.

A randomized clinical trial is represented as the gold-standard of research methods for evaluating treatments, but it is still a complex, costly, and time-consuming undertaking that must be carefully managed for the results to be valid. As stated, clinical trials use double-blind randomized samples to compare medical treatments (including a placebo-administered control group). Randomization is the

[3] Sometimes population groups are used as units in the clinical trial for testing treatment protocols. In community studies random samples of population groups may be studied as paired units (one being a control). Schools or even cities may be the population units sampled. The sample in one city may be subjected to the health program being tested while the sample in another city is not. A sample of schools in a city may be divided into schools subject to the treatment program and those used as controls. Alternatively, all schools in a city may be included in a study, but only a sample of students in each school is subject to the program and another sample in each school is carried as controls. In another variation, a single sample of the population may be followed longitudinally and examined before and after a health program is put into effect.

assignment of alternative treatments to the participants in the study on a random basis. Randomization is also intended to identify two groups that are as similar as possible to one another in their demographic, social, and economic characteristics. Randomization and blinding are two methods in the treatment, accompanying the use of a placebo in a control group, that are designed to deal with the problem of possible selection bias in the results of clinical trials. Research with randomized clinical trials is clouded by the difficulties of avoiding selection bias and achieving optimum randomization. Selection bias is particularly difficult to avoid when enrollment in a trial is sequential and previous assignments are revealed, given the interest of the recommending physician to opt for the treatment rather than the placebo for his/her patient. Under these circumstances, and particularly because of the physician's role in contributing to selection bias, special management of the selection process in the design of the clinical trial is necessary. Glantz (2005) argues that the only way to assure randomization is by use of a table of random numbers or related method. Enrolling patients by date of entry into the study or alternating the assignment to the two groups would bias the results.

There are other limitations of randomized clinical trials. The trial may not last a sufficient number of years for the effects of the intervention to appear, as for some chronic diseases. Next, such trials usually enroll selected categories of participants, say older men, so that the results are not properly generalizable to the general population. The converse is also true: The results of a trial that enrolls a representative sample of the general population may not apply to a special segment of the population, such as older men. In practice, clinical trials enroll special or volunteer patients who choose to go to medical centers for special treatment. They are, therefore, hardly representative of the general population. Finally, the trial may involve too few patients to have sufficient power to detect a true difference.

Meta-Analysis

Meta-analysis is the statistical analysis of the results of a group of independent studies, all designed to test a given hypothesis or treatment, or address a set of related research hypotheses, for the purpose of arriving at an integrated conclusion or to determine the effect of an experimental protocol more conclusively. Meta-analysis is intended to serve important research needs. The individual studies composing the meta-analysis may show apparent contradictions in the research findings. Meta-analysis aims to interpret or reconcile conflicting conclusions even where the primary results do not permit firm or consistent conclusions. The individual studies may suffer from large sampling errors, wide confidence limits, and reduced statistical power because of small sample sizes. Combined analysis of the primary results adds to the effective size of the sample, and hence narrows the overall confidence limits and increases the power of the findings.

Meta-analysis is now being widely employed in the health sciences to supplement traditional methods of research and analysis. For example, the New England

Research Institutes (1994) conducted a meta-analysis of the protective effects of hormone replacement against coronary heart disease in women and found an insignificant association between estrogen therapy and sickness or death from heart disease, once major methodological differences in the various studies were controlled.

Meta-analysis is a collection of systematic techniques that include the conversion of the metric of each study to a common metric. Because different studies may be investigating somewhat different variables or concepts, have different research designs, or employ different measurement scales, meta-analysis must translate the results into a common metric with similar concepts or variables before statistically exploring the differences in the findings. Meta-analysis also allows for appropriate weighting of each study to allow for differences in the number of cases and for other characteristics of the study. Common measure of the size-effect of a variable is the relative risk or the log-odds ratio.

Meta-analysis raises many questions and there are those who are strong critics of the method. One set of questions is concerned with the studies included. Only published studies are included but many studies with negative results and reduced risks are not submitted for review by journals or, if submitted, are not accepted for publication. Studies that report no association are of little reader interest and are passed over in favor of those that do. Moreover, the studies published vary widely in quality. The question arises as to how similar do studies have to be to be included in the same meta-analysis. Sources of bias are not controlled or removed by meta-analysis. The investigators tend to be biased in favor of having the study come out the way they hope and this will appear in the slant of their interpretations. Even a good meta-analysis cannot eliminate the effects of badly designed studies. Yet some researchers include weaker studies in their meta-analysis while adding a measure of the quality of each of the studies included in it. Next, meta-analysis may hide important differences among individual studies so that each study must be carefully examined for its contribution to the whole analysis. Finally, it is usually very difficult to reproduce the results of a meta-analysis.

Qualitative Sources

Quantitative methods provide most of the data and analytic information about health, but qualitative methods are used in a variety of ways to “back up” quantitative research. Specifically, they are used to plan quantitative studies, either to develop the questionnaire or to develop hypotheses about the relevant factors, and to aid in the analysis of the results of the quantitative studies. Case studies involving in-depth interviews, ethnographic research involving participatory methods, and focus-group discussions testing possible survey questions and responses represent three widely used qualitative methods employed in health studies.

Focus groups have been widely accepted as an effective way to understand people’s attitudes, beliefs, and experiences. Accordingly, survey researchers use

focus groups to develop and improve the design of their questionnaires, such as in fashioning questions on disability, and political analysts use focus groups to explore voters' reactions to the messages of political candidates or government officials, such as those relating to health reform.

Ethnographic research, in particular, can be used to construct models of "regional" (i.e., geographic, ethnic, class) illnesses, including their symptoms and etiology, by gathering data about the terminology used for different conditions and about folk methods of prevention, diagnosis, and treatment. Measures of self-assessed health can be greatly improved by developing questionnaires and instructional material that take into account regional views as to good health, by educating health providers in the most accurate ways of reporting illness, and by educating individuals in the community about health matters (Obermeyer 1996).

Some analysts define the role of qualitative methods as more important than suggested above. They maintain that quantitative, i.e., objective, measurements are inadequate to comprehend the complexity of the human experience and that qualitative methods have made their own independent contributions to the study of human health and aging.

Quality of Data

The measurement of illness is a very complex task, mainly because the concept of illness is ambiguous and fuzzy. There is no clear dividing line between wellness and illness, and as we have seen, the determination of illness is often made on the basis of subjective statements of respondents in surveys. Morbidity data cannot be readily evaluated for completeness or accuracy, but evidence suggests that they are subject to considerable reporting error. Inaccuracies in reporting health conditions may be measured by the discrepancy between objective and subjective reports of health. The uncertainty regarding the estimates of the health status of the population of an area is amplified when the estimates are based on small samples and/or selected populations rather than large samples that are representative of the general population. Some of the sources employed (e.g., clinical records and disease registries) are quite limited in scope and, therefore, of limited utility for making inferences regarding the state of health of the general population of an area. Moreover, the sample may be biased and not representative of the universe from which it was drawn (e.g., samples based only on listed telephone subscribers), or it may not represent the larger population (e.g., samples based only on adult white males).

By contrast, mortality data are both easier to collect and easier to interpret. Vital registration systems in the more developed countries are generally complete and dependable. While the death statistics for the less developed countries are subject to much greater error, there are ways of estimating the extent of incompleteness or for exploiting alternative sources of data.

Census and Survey Data, Vital Statistics, and Administrative Records

General Problems

Aggregate census and survey data, as is true of all reported data, are subject to errors of coverage (i.e., omission, duplicate reporting of persons or households), errors of classification (i.e., reporting an incorrect category of a classification system), and nonreporting (i.e., failure to report a subject item). Entire households may be omitted, some members of households in included households may be omitted, responses for a range of items may not be secured for entire households, some items may not be reported for some individuals, and items may be misreported. Data entries may be inconsistent, the entries may be nondescript or vague, and dual answers may be entered for some items. Census and survey data for certain groups, such as some age groups, males, some racial minorities, socioeconomic status, and certain types of households, are particularly affected by undercounting. Errors of classification (e.g., of mortality and morbidity data) also vary for gender, race, socioeconomic status, and marital status. Evaluation studies suggest that the health and mortality data for whites, females, and groups of higher socioeconomic status (i.e., the more educated and affluent residents) are of better quality than those for some nonwhite racial groups, males, and groups of lower socioeconomic status. Vital statistics and administrative records are also subject to these types of errors.

As a result of the various types of errors enumerated, the raw aggregate data may suffer from many types of biases. Bias refers to the deviation of aggregate data from the true values because of reporting errors, for example, the net tendency to overreport good health at the expense of fair health. The biases result mainly from problems of recall, but some biases result from deliberate misreporting and nonreporting and some are due to processing errors. Many individuals are reluctant to respond to questions about certain items, such as income, so that, although multiple studies emphasize the importance of socioeconomic status in health and health care, the data are often biased in identifying the population at risk. Moreover, persons of different socioeconomic statuses or races may interpret a given question on health differently.

The responses relating to health conditions in social surveys are based on either subjective or objective information. Subjective information is the respondent's own perception of his or her health, whether it is the overall health condition, the presence or absence of chronic diseases and disabilities, or life expectation ("subjective survival"). The percent of respondents assessing their health as poor, fair, good, very good, or excellent is a commonly used scale. Objective information is the information on the health of respondents secured from medical records, including physicians' medical records, hospital discharge records, and medical examinations. Much study of the merits of subjective data has been carried out. It has been discovered, for example, that subjective reporting of health condition is a good predictor of the respondent's survival period.

Shortcomings in the quality of the health and mortality data are especially characteristic of the data secured in the less developed countries. Instead of accepting seriously inadequate records, data from these sources should be adjusted, to the extent possible, to allow for their statistical deficiencies, or data for only selected sample enumeration or registration areas should be used as proxies for data for the area under study.

Effect of Age of Respondents

Several health surveys collect data only about old persons, (e.g., LSOA and HRS). We can be reasonably concerned about the quality of these data. Such a concern arises from the realization that often cognitive functioning declines in old age and old persons usually provide the information about themselves. Comparisons of the quality of responses of old persons with actual records or the results of other surveys for such persons, and comparisons of measures of error in the responses of old persons with those of middle-aged persons and youth have demonstrated, however, that the errors are typically not much different among younger and older respondents.

On the other hand, old persons tend to show more nonresponses. They appear to be less likely to complete the survey interview than middle-aged and younger adults. The differences are greater when the contact is by telephone than when it is by face-to-face interview. Nonresponse can lead to biased results because the nonrespondents and respondents are likely to have different characteristics. Nonrespondents are, on average, less healthy than respondents mainly because health problems are an important reason that older persons do not participate in some surveys. Health problems may explain why age differences in response errors are not great: Older persons who would have the greatest difficulty in responding to the questionnaire are underrepresented in the survey. Old persons have less tolerance for long questionnaires, need more assistance in interpreting the questions, and require more time for the interview than younger respondents (Herzog and Rodgers 1988; Carp 1989; Herzog and Kulka 1989; Magaziner et al. 1988). Accordingly, survey planners should try to use simple questionnaires, proxy respondents when indicated, specially trained interviewers, and extended periods for interviewing disabled or frail respondents. Research shows a mixed record on the use of proxies, but for very old persons the balance, considering the issues of response error and nonresponse, may weigh in favor of their use. In sum, there is little question about the viability of surveys of older adults, but such surveys do entail special problems in dealing with them.

Combination and Comparability of Survey Data

Changes in methods of data collection (e.g., introduction of CATI, i.e., computer-assisted telephone interviewing), survey redesign (e.g., change in the sample frame,

revisions of major code sets), and revisions of reporting standards or concepts (e.g., new concepts of race, poverty, and disability) contribute to making databases discontinuous over time. Many of the NCHS surveys were redesigned with the availability of the 2000 census results, and many changes have been made in the U.S. health system (e.g., new types of providers; addition of drug benefits). For this reason, the data on health in time series may not be comparable. The need to redesign surveys and adjust content as time passes, so as to serve current needs for policy planning and program development, must be balanced against, and preferred to, the need to maintain comparability over time. A statistical bridging device should be designed to aid the user in this transition where possible.

Problems arise not only with the combination of data for a given survey for two or more years but also with respect to the combination of data from different surveys. Various adjustments in the data may have to be made, such as shifts in blocks of data for reconciling concepts, modification of sampling weights, etc.

Sampling Bias and Variance

In health surveys, as in all sample surveys, one has to confront the problems of sampling bias (i.e., the extent to which the average sample value would differ from the true value if an infinite number of samples were taken), and sampling variance (i.e., the variability of the sample data if an infinite number of samples were taken). One must balance off bias against variance in designing the sample, and occasionally a biased method is employed to secure a reduced variance. One must also reconcile the competing goals of achieving maximum possible power (i.e., identifying significant differences) and the least possible cost. This is a trial and error process, balancing off competing goals. Thus, one can minimize costs, subject to variance constraints, or minimize variance, subject to cost constraints.

Adjusting and Correcting the Sample Data

A number of devices are implemented to allow and correct for deficiencies in the raw survey and other record data. Among these are editing, imputation, record linkage, and adjustment of the data to independent (population or facilities) "controls" (Herzog et al. 2007). Editing is required to correct for inconsistent data entries, nondescript or extremely vague entries, and double entries where only one entry is sought. Imputation covers a range of devices for filling in entries that are missing, whether for selected questions, all entries for a person, or all entries for a whole household. Editing for misreported items and imputation for missing items and persons may themselves introduce some bias in the data but usually they improve the accuracy of the data.

After data editing and imputation, the sample data must be adjusted for the omission of households (or facilities, such as nursing homes) and of persons within enumerated households (or facilities), inflated for the sample size, and then adjusted

to independent estimates of population in various categories, such as age-sex, race, and Hispanic-origin groups ("ratio adjustment"). Omitted households are those in the sample which failed to respond to the survey and omitted individuals are those who are in sample counted households but who were not counted. The sampling weights represent the inverse of the combined sampling ratio, which may be derived as the product of two or more sampling ratios, where the sample survey has a complex design. The final ratio adjustment is designed to force agreement of the sample-inflated, edited data with independently derived census-based postcensal population estimates (Siegel 2002). The ratio adjustment allows for some sampling error, but it mainly adjusts for net census errors in the survey (i.e., both net coverage errors and net reporting errors) for each age- sex-race-Hispanic group (but only up to the census level of coverage).

Epidemiological Studies and Clinical Trials

Epidemiological Studies

Epidemiological studies have similar and additional limitations. They require a heavy investment of money, and many years must pass before useful information is obtained. They cannot solidly prove a cause-effect relation between risk factors and a disease; at best, they show a probable cause-effect relation; Risk assessment postulating a cause-effect relation in epidemiological studies is more likely to be valid if the association is strong than when it is weak. The demands on the participants in these studies are considerable and there will inevitably be misreports, missing data, dropouts, and deaths. Since the participants in the survey are likely to be self-selected with respect to education, socioeconomic status, and health, as will be the dropouts (the former being positively selected and the latter being negatively selected), there is considerable risk of bias in the study sample, particularly as time passes.

Misreporting. A serious potential problem is inaccurate reporting by respondents, who may give over-favorable reports of their current health status or who may not recall previous health conditions accurately or at all. With increasing education, there may be an "inflation of morbidity." Increasing education is associated with greater awareness of health problems, higher utilization of health services, and, hence, higher reporting of morbidity. Moreover, maintains Johansson (1991), especially in the less developed countries, as the level of mortality falls, rising health expectations on the part of ordinary people will lead them to report their illnesses more fully or to exaggerate them. Riley (1992), on the other hand, argues that any such tendency should not be overemphasized at the expense of the reality of increasing morbidity resulting from increased duration of chronic illnesses when mortality falls.

A non-life threatening disease may be considered of little importance from a medical point of view but considered quite disabling by the individual in a self-assessment of health. Respondents tend to interpret their health status in relation to their own view of what normal health and well-being are. They may be too willing to accept pain and functional limitations, or the opposite, depending on age, sex, education, and racial/ethnic background. With advancing age, conditioning to various discomforts and disorders occurs and they may be accepted as normal for one's age, not as diseases to be reported.

In some countries it is not uncommon for local views of an illness and the biomedical categories for the illness to be at odds. This difference of view exacerbates the problem of inconsistency between objective indicators and subjective reporting of health. These problems are particularly evident in the area of reproductive morbidity in the less developed countries.

Missing data. Missing data present a special problem in addition to the fact that, uncorrected, they add to the error of the count. They may bias the results. They are not well represented by the cases for which data are reported. Analysis of longitudinal data with incomplete data is problematic. Unless appropriate measures are taken, missing data can result in seriously biased results and interpretations, especially in longitudinal data. Persons who do not report their health status are more likely to be suffering from a health condition, or a more severe health condition, than those who do report their health status. For example, persons who do not report their disability status are more likely to be disabled, or more severely disabled, than those who do report their disability status. It is better, therefore, to impute the missing data by statistical methods than to regard them as insignificant, or to assume that the missing data are randomly distributed or distributed like the reported data.

A variety of methods have been devised to impute missing data, including mathematical interpolation, substitution of the mean (i.e., replacing missing data for a variable with the average of the reported data for the variable), simple regression, and multiple imputation (i.e., applying several regressions that produce multiple results). Multiple imputation is now viewed as superior to the other procedures (Wayman 2003). There are cases where the imputation of missing cases does not improve the data. For example, in a study of the effect of three common approaches to handling missing data on the results of a predictive model of hospital admissions, Gorelick (2006) performed a series of simulations in which data were deleted for varying proportions of patients and a logistic regression making various types of imputations of missing data was carried out. His results showed that, whether the analysis was restricted only to cases with complete data, whether the missing data were assumed to be normal, or whether values were imputed for the missing cells, the results were biased estimates of the predictive model of hospital admissions.

Deaths and dropouts. The results of a study may be biased, and the analyst's interpretation of the results affected, by the frequency and timing of deaths and dropouts and the failure to take account of deaths and dropouts in the study (Rabbitt et al. 2005). If the study is concerned with health status or cognitive functioning,

the deaths are definite indicators of poor health and hence the decedents are likely to have characteristics correlated with the subject of the study. Dropouts are of two kinds. Most people who drop out of a longitudinal study on health do so because of poor health or difficulties in mobility (Rabbitt et al. 1994) and these conditions are also likely to be correlated with the subject of the study. Dropouts and deaths tend to bias studies of health status "downward."

A segment of the dropouts are not unhealthy, frail, or immobile, however, but are making important social transitions in their lives, such as getting a new job, marrying, or moving. Their omission biases the study in the opposite direction. In spite of the exclusion of this group of dropouts from an analysis, the net effect of the neglect of all dropouts is to overestimate the true health status of a population and to underestimate the extent of declines in health with time. It can usually be shown that, if no corrective action is taken, the omission of data can produce serious biases in the results and difficulties in interpreting them.

Censoring. Censoring refers to the fact that information on the timing of an event under study is not known for all participants for the period of the study. The actual date of the onset of a health condition cannot be known at the first examination date – only the health status of the person at the time of the examination – and the survey may be terminated before some of the subjects have incurred the health condition or died. For example, an epidemiological study may seek to identify risk factors for infection from HIV and the progression from infection to clinical disease. In HIV, the exact time of infection is not known, but it can be set as falling in the interval between the last negative blood test and the first positive blood test. The data for this period are described as interval-censored. If the infection occurred before the study, the data are described as left-censored. The exact date of the infection is not known, only that it occurred some time earlier.

Usually a study does not run long enough to reveal the full effects of the event or condition under study on all participants. Some of the subjects who will get infected are not diagnosed with it at the time it occurred or at the time the study is concluded. The data for them are right-censored. Similarly, the period between the actual time of infection and the actual time of onset of disease cannot be fixed closely since neither the first date nor the last date is known. This period is named the latency period. Data for the latency period are censored both on the left and the right, or doubly-censored. One way to deal with the censoring problem is to take more frequent observations of the study population, but this may be a costly choice. Another is to design the analysis of the results specifically to deal with the censoring problem.

Multicollinearity. Multicollinearity is the correlation of risk factors in epidemiological studies. The problem with multicollinearity is that it may confound the results of the analysis. If the factor being tested is correlated with another factor that may truly be exerting a cause-effect relation, the effect may erroneously be attributed to the factor being tested. For example, a study may be interpreted to show that eating green, leafy vegetables reduces the risk of heart disease, but people who eat

green, leafy vegetables may also exercise a lot and be more health conscious. Hence, the cause of the reduction in heart disease may be eating green, leafy vegetables, exercising frequently, or being health conscious, or all in combination. The problem of multicollinearity requires explicit attention in multiple regression analysis and hence in the analysis of epidemiological results.

Clinical Trials

Randomized clinical trials (RCT) are considered the "gold standard" for evaluating the effectiveness of medical treatments. However, clinical trials are subject to some of the same problems as health surveys in general, including response errors, and epidemiological surveys in particular, including selection bias, deaths, dropouts, and censoring. These potentially bias the results, In clinical trials, as in epidemiological studies, the sample is often restricted to a particular segment of the population (e.g., middle-aged white males; nurses), and under these conditions, the results cannot safely be generalized to the whole population. In the past, such trials have failed to cover women and elderly persons. The planned randomization may be affected by the initial self-selection of participants, and the placebo effect cannot be eliminated entirely because the outcome can be affected by the subjects' reaction to the study and investigators' tendency to believe that one treatment is better than another. Physicians and investigators have opposing positions with respect to the selection of participants in clinical trials. The physician would like to see all his patient/selectees given the medication or procedure and the investigator seeks to achieve a balance between selectees' receiving active treatment and selectees' receiving the placebo. The physician may hold back a possible candidate for a protocol if he/she believes the candidate will receive the placebo and encourage the selection of a candidate if he/she believes the candidate will be given the active treatment. Hence, a "game" is played between two non-cooperative players in accordance with Nash equilibrium game theory.

Meta-Analysis

Meta-analysis cannot overcome the defects of poorly designed studies or eliminate response-recall bias, and if the individual studies show marginally significant results, meta-analysis will not necessarily provide statistically significant results. There are other more serious problems with meta-analysis. The investigator may select the studies included for the meta-analysis on the basis of certain criteria and the selection process may influence the outcome of the meta-analysis. Moreover, given that studies with successful outcomes are more likely to be published than those with negative outcomes, the former are more likely to be overrepresented among the published studies. This biased representation among the published

studies may influence the results of the meta-analysis. Ravnskov (1992), for example, concluded that the studies selected for the meta-analysis of cholesterol studies are biased samples of those available.

Given these limitations, Thompson and Pocock (1991) question whether meta-analysis can be depended upon to provide reliable results. Many scholars continue to have confidence in meta-analysis, however, since studies based on it continue to be published.

Methods of Survey Collection of Health Data

Methods of Interviewing

To appreciate the difficulties of securing valid and reliable health data economically, it is useful to take note of the variety of methods by which data on health are obtained in health surveys. Historically the personal interview, that is, an oral interview of a respondent by an interviewer in the respondent's home, was the common method employed. The respondent provided information about their own health or that of members of the household either on the basis of subjective judgments ("self-rated health") or on the basis of medical reports obtained from physicians or other health providers. With technological developments, particularly the introduction of computers, and the need to contain survey costs, the interviews were directly recorded by the interviewer on the computer in the respondent's home, a process referred to as Computer-Assisted Personal Interviewing (CAPI).

More recently, partly to restrain costs further, the interviews have been conducted by telephone and recorded on computers, a process designated as Computer-Assisted Telephone Interviewing (CATI). Other interview protocols have also been introduced, in particular, the direct recording of the answers of the respondent on the computer, designated as Audio Computer-Assisted Self-Interviewing (ACASI) and Computer Audio-Recorded Interviewing (CARI). The use of sampling/online interviewing has also been introduced and has grown rapidly. Many survey takers have embraced web research because of its speed, video capability, flexibility, and usually lower costs. The quality of the data remains to be evaluated.

Methods of Collecting Sensitive Data

A number of alternative interview methods have been proposed to secure information in the developed countries on health topics that are sensitive, such as use of illegal drugs, alcoholism, mental health problems, and sexually transmitted diseases. Among these methods are adaptive sampling, the randomized-response technique, and the three-card method. Direct interviewing – the traditional interviewer/respondent method – has also been used. Brief descriptions of these methods are given below.

Direct Interview

The sensitive questions may be asked directly in in-person interviews or the respondent may answer the same questions with the aid of computers. For example, the most sensitive questions of the National Survey of Family Growth are asked by using Audio Computer-Assisted Self-Interviewing (ACASI).

Adaptive Sampling

In adaptive sampling, the procedure for selecting the sample depends on values of variables of interest observed during the survey. New respondents may be selected on the basis of reports by earlier respondents. In health studies surveying hard-to-reach or hidden populations, such as injection drug users or persons infected with HIV/AIDS, hepatitis C, or genital herpes, links from qualified respondents are followed to other qualified respondents to expand the size of the sample. This device resembles network or multiplicity sampling, described below, as a way of expanding the effective size of the sample for covering hard-to-reach or rare populations.

Randomized Response Technique

Another procedure designed to assist in securing more complete and accurate responses on sensitive subjects is the randomized response technique. The technique has more than one variation. Commonly, the sensitive question is asked in conjunction with another, innocuous question, and the respondent determines which question to answer, using some probability device under his or her control. For example, if the probability device is the toss of a coin, heads could mean that he or she should respond to the sensitive question. The enumerator would not know the choice in the individual case made by the respondent, but the probability of picking the sensitive or the innocuous question over all respondents is known because the probability device established these chances (Shimizu and Bonham 1978).

The Three-Card Method

Another procedure for collecting sensitive data that assures privacy of response is the three-card method. This is also a survey-based indirect estimation- technique but it seeks to avoid the complex design of the randomized response technique. The method was originally devised to provide estimates of the proportion of illegal aliens among the foreign-born population in the United States, but it has the potential for wider use, including use in securing information on various sensitive health problems (Droitcour et al. 2001; U.S. GAO 1999).

In this method, three representative samples of the population are selected, and each of the sample populations answers the key sensitive question and other questions regarding the respondent's health, but with a slightly different

arrangement of the mutually exclusive categories for responding to the questions. A set of three cards is designed to contain all the mutually exclusive categories for the question but arranged in three different groups – one arrangement for each card – for use with each sample. The respondent is asked to select one of the three groups on the card for his or her answer. The sensitive category is listed on each of the cards but it is always grouped with one or more of the nonsensitive categories. One of the nonsensitive groups is listed separately on one of the cards, so that the percentage of the population with this condition can be determined for its sample population. Another nonsensitive category is listed separately on the second card for the second sample population, so that the percentage with this condition can be determined for the second sample; and the same for the third card and the third sample population. The investigator can estimate from the responses in the three samples the proportions of persons who fall in each of the nonsensitive response categories. Because each sample yields directly the separate proportion in one of the nonsensitive groups, the proportion for the sensitive group can be obtained by subtraction of all these percentages from 100.0.

The method is designed to apply to questions with a limited number of response categories (six or less) in order to determine the proportion of persons in only one of these – e.g., illegal aliens among a brief list of citizenship statuses (e.g., citizen, legal resident alien, refugee or asylee, illegal alien, other). With respect to health, the application of the three-card method is more limited because it cannot be used with the numerous health categories that are possible. The health categories may be broadly grouped, however, so as to reduce their number: Suppose, for example, the survey is trying to secure information on sexually transmitted diseases. The categories could be sexually transmitted diseases (i.e., the sensitive condition), musculoskeletal diseases, cardiovascular diseases, cancer, other conditions, and none of the others. The task could be measuring the proportion of all pregnancies terminating in abortion and the questions can be structured to inquire about a few pregnancy outcomes – miscarriage, abortion, fetal death, and birth. The results could be summarized as follows:

$$\text{All pregnancies} - \text{births} - \text{miscarriages} - \text{fetal deaths} = \text{abortions}$$

Some hypothetical proportions obtained might be: $1.00 - .35 - .25 - .15 = .25$. Hence, the three sample surveys resulted in an estimate of 25% of births terminated by abortion.

Methods of Collecting Data on Rare Events

The efficacy of conventional surveys to locate or estimate rare or special populations (e.g., persons with cancer, hearing impairment) can be substantially improved with the adoption of adaptive sampling or multiplicity counting rules. Unlike adaptive sampling, multiplicity specifies the categories of additional persons to be covered, without actually interviewing them. Multiplicity or network counting rules usually

include close relatives of the members of the primary household who may reside in other households. This technique enlarges sample size and reduces the variance of the estimate. Network sampling does not necessarily alleviate the problem of response bias and may in fact increase reporting error. Moreover, to develop the multiplicity estimators (i.e., the formulas for weighting the sample returns), additional questions are needed to secure data about the number and characteristics of the relatives who are eligible for "inclusion" in the survey and who may be target individuals. These additional questions increase the complexity and cost of the survey (Sirken 1970, 1975).

Even though the multiplicity counting rule always yields more cases than the conventional counting rule, some conditions are so rare that even using a multiplicity counting rule does not provide enough cases. The strategy in the choice of a counting rule is to minimize the combined effects of net reporting error (i.e., bias) and sampling error (i.e., variance) within the constraints of the survey budget, while avoiding undercounting or overcounting of the variable in question. Multiplicity counting may have higher reporting biases but lower mean squared errors than does the conventional counting rule unless the sample sizes are very large (Czaja et al. 1986).

Confidentiality and Disclosure Issues

There is another important problem to be resolved in completing sample surveys. Almost of all data collected by the U.S. Census Bureau, the National Center for Health Statistics, the National Institutes of Health, and other government agencies are obtained under a pledge to the respondent (whether an individual, household, or establishment), mandated by Federal legislation, that the data will be used only for statistical and research purposes, and will not be released in such form that an individual, household, or establishment can be identified. The requirement that the data for reporting units be kept confidential calls for some modification of the original data to avoid revealing the identity of the reporting unit.

This issue arises with respect to all forms of release of health data – published tables, summary tape files, public-use microdata files (PUMF), and computer-accessed data – as well as all forms of population coverage – 100% data, sample data, and microdata. Since individual records are available in public use microdata files, information identifying a particular respondent is more likely to appear. Even aggregate data may reveal the unique records of individuals in certain cells, so that the issue of disclosure arises with such data as well as with microdata. The problem is attenuated somewhat with aggregate weighted sample data. If the data are inflated for the sample ratio and the sampling rate is not revealed, the data may easily pass confidentiality inspection. The need of data-collecting organizations to protect the confidentiality of their records is complicated by the developments in technology (e.g., online computer access to data), the increased demand for highly disaggregated data, the availability of microdata files, and the linkage of various data sets.

Confidentiality Edits

Many different devices have been developed for purposes of disclosure limitation, i.e., to make confidentiality edits. Data suppression and data modification are types of disclosure limitation. With cell suppression certain data are removed from the internal record and are not made available in the release of the data. Cell suppression has been a common way in the past of limiting disclosure of census, survey, and other record data. It requires the complementary suppression of other cells, such as marginal totals. The information to be suppressed depends on the size of the geographic area and the level of geographic detail provided. Cell suppression has serious limitations, in particular the reduction of the volume of data published and modification of the original totals for areas and their basic demographic characteristics.

With cell modification, new information is substituted for the information that has to be concealed. One type of cell modification is called swapping or switching. Data for some households are swapped or switched with the data for other households residing in different geographic locations but having identical characteristics on a certain set of key variables. Data may also be blanked out and new entries imputed using multiple imputation. Another way to reduce the risk of disclosure is to impose a threshold rule, for example, that cells have to have at least five cases. Then random rounding, swapping, suppression and imputation, and "blurring" (i.e., aggregating across small groups of respondents and replacing individual values with the average) may be used to fill the gap. Disclosure-limitation devices are under continuing review with the goal of minimizing the possibility of revealing an individual's identity and maximizing the amount and quality of the data made available to the public (Siegel 2002).

Appendix 2.1 Compendia on Health and Mortality Data

United States Data

The Federal Agency Forum on Aging-Related Statistics publishes *Older Americans: Key Indicators of Well-Being*, a chartbook that presents statistical charts and data collected and compiled by the federal agencies that are members of the forum. The chartbook is currently updated every year.

The U.S. National Center for Health Statistics, Centers for Disease Control and Prevention, publishes *Health, United States* annually, a compilation of tables on the health situation in the United States. The 2010 report contains tables with data on the trends in mortality and health as well as a chart book with special feature on death and dying.

The Population Reference Bureau publishes *United States Population Data Sheet* annually. It provides data on both basic and specialized demographic characteristics of the United States and each state.

International Data

The U.S. Department of Health and Human Services periodically publishes an *International Health Data Reference Guide*, which presents an overview of international health data including vital statistics, hospital statistics, health personnel statistics, and population-based surveys. The purpose of the guide is to provide information about sources of health data around the world useful for supporting international studies and comparisons of population health.

The Population Reference Bureau is the source of a few periodic international compilations, including its annual *World Population Data Sheet* and its special data sheets on *Women of Our World*, *Africa Population Data Sheet*, and *The Wealth Gap in Health*. The latter sheet presents data on differences in various health indicators for three economic status groups in 53 Less Developed Countries.

International Population Data Center, U.S. Census Bureau, publishes The *World Population Profile* every few years, a report providing population estimates and projections for the countries of the world.

The United Nations and its component and member organizations publish several reports periodically that carry health data, including principally the *Demographic Yearbook* and the *World Health Report*. The *Demographic Yearbook* has been released annually by the U.N. Statistical Office since 1947. Each annual volume includes basic demographic data for each country and, in addition, specialized data on which the particular volume focuses. *The World Health Report*, first published in 1995 by the World Health Organization, is WHO's leading publication. Each year the report combines an expert assessment of global health, including statistics relating to the countries of the world, with a focus on a specific subject. The 2006 report focus on the global health workforce.

References and Suggested Readings

Censuses and Surveys

Belloc, N. B., & Breslow, L. (1972). Relationship of physical health status and health practices. *Preventive Medicine, 1*(3), 409–421.

Berkman, L. F., & Breslow, L. (1983). *Health and ways of living*. New York: Oxford University Press.

Costa, D. L. (2000). Understanding the twentieth-century decline in chronic conditions among older men. *Demography, 37*(1), 53–72.

Crimmins, E. M., & Seeman, T. E. (2004). Integrating biology into the study of health disparities In L. J. Waite (Ed.), Aging, health, and public policy: Demographic and economic perspectives: *A supplement to Vol. 30. Population and development review* (pp. 89–107). New York: Population Council.

Elman, C., & Myers, G. (1997). Age and sex differentials in morbidity at the start of an epidemiologic transition. *Social Science and Medicine, 45*, 943–956.

Elman, C., & Myers, G. (1999). Geographic morbidity differentials in late nineteenth-century America. *Demography, 36*(4), 429–443.

Geronimus, A. T., Bound, J., Waidmann, T. A., Colen, C. G., & Steffick, D. (2001). Inequality in life expectancy, functional status, and active life expectancy across selected black and white populations in the United States. *Demography, 38*(2), 227–251.

Hayflick, L. (1994). *How and why we age*. New York: Ballantine Books.

Hayward, M. D., & Heron, M. (1999). Racial inequality in active life among adult Americans. *Demography, 36*(1), 77–91.

ICF Macro (Macro International). www.measuredhs.com.

Kaplan, G. A., Seeman, T. E., Cohen, R. D., Knudsen, L. P., & Guralnik, J. (1987). Mortality among elderly in the Alameda County study: Behavioral and demographic risk factors. *American Journal of Public Health, 77*(3), 307–312.

Kroeger, A. (1983). Health interview surveys in developing countries: A review of the methods and results. *International Journal of Epidemiology, 12*, 465–481.

Lenfant, C. (1999, October 27). Dietary fiber, weight gain, and cardiovascular disease risk factors in young adults. *Journal of the American Medical Association, 282*(16), 1539–1546.

Lloyd, C. B., & Marquette, C. M. (1992). *Directory of surveys in developing countries: Data on families and households, 1975–1992. A population and development review* (Supplement, Vol. 27). New York: Population Council.

Manton, K., & Soldo, B. (1992). Disability and mortality among the oldest-old: Implications for current and future health and long-term care service needs. In R. M. Suzman, D. P. Willis, & K. G. Manton (Eds.), *The oldest-old* (pp. 199–250). New York: Oxford University Press.

Population Association of America. (2003). *PAA affairs*. Winter 2003:7–8.

Preston, S. H., & Haines, M. R. (1991). *Fatal years: Child mortality in late nineteenth century America*. Princeton, NJ: Princeton University Press.

Wiley, J. A., & Camacho, T. C. (1980). Life style and future health: Evidence from the Alameda Country Study. *Preventive Medicine, 9*(1), 1–21.

Vital Statistics and Other Administrative Sources

Garibotti, G., Smith, K. R., Kerber, R. A., Boucher, K. M. (2006). Longevity and correlated frailty in multigenerational families. *Journal of Gerontology: Biological Sciences, 61A*(12): 1253–1261.

Gavrilov, L. A., Garvrilova, N. S., Olshansky, S. J., & Carnes, B. A. (2002). Genealogical data and the biodemography of human longevity. *Social Biology, 49*, 160–173.

Hoyert, D. L., & Martin, J. A. (2002, February). Vital statistics as a data source. *Seminars in Perinatology, 26*(1), 12–16.

North American Association of Central Cancer Registries. (2001). *CiNA + [Cancer in North America]*. Online at www.naaccr.org/cinap/index.htm>

Schatz, I. J. (1995). Internal medicine residency program violations of the minimum autopsy rate. *Journal of the American Medical Association, 273*, 1092.

Shanks, J. H., McCluggage, G., Anderson, N. H., & Toner, P. G. (1990). Value of the necropsy in perioperative deaths. *Journal of Clinical Pathology, 43*, 193–195.

U.S. National Center for Health Statistics. (2001). The autopsy, medicine, and mortality statistics. By D. Hoyert. *Vital and Health Statistics, 3*(32) (DHHS Publication No. (PHS) 2001-1416).
U.S. National Center for Health Statistics. (2006). Summary health statistics for the U.S. population: National Health Interview Survey, 2004. By P. F. Adams & P. M. Barnes. *Vital and Health Statistics, 10*(229).
U.S. National Office of Vital Statistics. (1950). *International recommendations on definitions of live birth and fetal death*. Washington, DC: U.S. Public Health Service.

Secondary Sources

New England Research Institute. (1994). *Network*. Summer/Fall.
New England Research Institute. (1996). *Network*. Spring/summer.
U.S. Bureau of the Census. (2001). *International database*. census.gov/ipc/www/idbnew.html.
U.S. Department of Health and Human Services. (2000). *International health data reference guide, 1999* (DHHS Publication No. (PHS) 2000-1007).
U.S. National Center for Health Statistics. (2002). *National center for health statistics: programs and activities* (DHHS Publication No. (PHS) 2002-1200).
United Nations. (1994). *Demographic yearbook 1993*. New York: United Nations.
World Health Organization. (2005). *World health report, 2004*. Geneva, Switzerland: World Health Organization.

Quality of Data

Andrews, F. M., & Herzog, A. R. (1986). The quality of survey data as related to age of respondent *Journal of the American Statistical Association, 81*, 403–410.
Carp, F. M. (1989). Maximizing data quality in community studies of older people. pp. 93–122 In M. P. Lawton & A. R. Herzog (Eds.), *Special research methods for gerontology*. Amityville, NJ: Baywood.
Czaja, R. F., Snowden, C. B., & Casady, R. J. (1986). Reporting bias and sampling errors in a survey of a rare population using multiplicity counting rules. *Journal of the American Statistical Association, 81*(394), 411–419.
Droitcour, J., Larson, E. M., & Scheuren, F. J. (2001). The three card method: Estimating sensitive survey items–with permanent anonymity of response. *Proceedings of the American Statistical Association, August 5–9, 2001*, Alexandria, VA.
Glantz, S. A. (2005). *Primer of biostatistics*. (6th ed.). New York: McGraw-Hill.
Gorelick, M. H. (2006). Bias arising from missing data in predictive models. *Journal of Clinical Epidemiology, 59*(10), 1115–1123.
Herzog, A. R., & Kulka, R. A. (1989). Telephone and mail surveys with older populations. A methodological overview. In M. P. Lawton & A. R. Herzog (Eds.), *Special research methods for gerontology* (pp. 63–89). Amityville, NJ: Baywood.
Herzog, A. R., & Rodgers, W. L. (1988). Age and response rate to interview sample surveys. *Journal of Gerontology: Social Sciences, 43*, S200–S205.
Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). *Data quality and record linkage techniques*. New York: Springer Publishing Company.
Johansson, S. R. (1991). The health transition: The cultural inflation of morbidity during the decline of mortality. *Health Transition Review, 1*(1), 39–65.

Kalter, H. (1992). The validation of interviews for estimating morbidity. *Health Policy and Planning, 7*, 30–39.

Magaziner, J., Simonsick, E. M., Kashner, T. M., & Hebel, J. R. (1988). Patient-proxy response comparability on measures of patient health and functional status. *Journal of Clinical Epidemiology, 41*, 1065–1074.

Marquis, K. H., Marquis, M. S., & Polich, J. M. (1986). Response bias and reliability in sensitive topic surveys. *Journal of the American Statistical Association, 81*(394), 381–389.

Obermeyer, C. M. (1996). A research agenda for reproductive health. *Newsletter 54*, Geneva, Switzerland: International Union for the Scientific Study of Population.

Rabbitt, P., Watson, P., Donlan, C., Bent, N., & McInnes, L. (1994). Subject attrition in a longitudinal study of cognitive performance in community resident elderly people. In B. J. Vellas, J. L. Abarende, & P. J. Gary (Eds.), *Facts and research in gerontology: Epidemiology and aging* (pp. 29–34). Paris: Serdi.

Rabbitt, P., Lunn, M., & Wong, D. (2005). Neglect of dropout underestimates effects of death in longitudinal studies. *Journal of Gerontology: Psychological Sciences, 60B*(2), P106–P109.

Ravnskov, U. (1992). Cholesterol lowering trials on coronary heart disease: Frequency of citation and outcome. *British Medical Journal, 131*, 579–588.

Riley, J. C. (1992). From a high mortality regime to a high morbidity regime: Is culture everything in sickness? *Health Transition Review, 2*(1), 71–77.

Shimizu, I., & Bonham, G. (1978). Randomized response in a national survey. *Journal of the American Statistical Association, 73*, 35–39.

Siegel, J. S. (2002). *Applied demography: Applications to business, government, law, and public policy*. San Diego, CA: Academic Press.

Sirken, M. G. (1970, March). Household surveys with multiplicity. *Journal of the American Statistical Association, 65*, 257–266.

Sirken, M. G. (1975). Network surveys of rare and sensitive conditions. *Advances in health survey research methods*. Hyatsville, MD: National Center for Health Statistics.

Thompson, S. G., & Pocock, S. J. (1991). Can meta-analysis be trusted? Lancet, 338, 1127–1130.

U.S. General Accounting Office. (1999, November). *Survey methodology: An innovative technique for estimating sensitive survey items*. GAO/GGD-00-30.

Wayman, J. C. (2003). *Multiple imputation for missing data: What is it and how can i use it?*. Paper presented at the 2003 annual meeting of the American Educational Research Association, Chicago, IL. Baltimore, MD: Center for Social Organization of Schools, Johns Hopkins University.

Part II
Concepts and Measures of Mortality and Morbidity

In Part II we begin the examination of the substantive material of health demography. Chapter 3 sets forth the concepts and basic measures of mortality through a description and analysis of death statistics. The chapter covers both observed summary and observed specific measures of mortality, as well as adjusted measures designed to provide greater analytical information about mortality levels and greater comparability between measures. The treatment of mortality is extended in Chap. 4 with a discussion of the anatomy, construction, and uses of the life table, a specialized tool for measuring mortality that describes the survival history of a cohort of births over its lifetime and produces the well-known summary measure of population longevity, life expectancy. A presentation of a variety of parameters of the age distribution of deaths in the life table in Chap. 4 supplements the tools for analysis of the age distributions of observed deaths given in Chap. 3. Chapter 5 presents a nearly parallel discussion of the concepts and measures of health status and functioning to that given for deaths in Chap. 3. The analysis of health status presents some additional complexities because a health condition involves special measurement problems, such as those relating to quality of data, duration, degree of severity, and degree of incapacitation.

Chapter 3
Concepts and Basic Measures of Mortality

Uses and Definition

Uses of Death Statistics

Death statistics are a basic element in measuring progress toward improved health and increased longevity of a population. They are needed both for demographic studies and for public health administration. Death statistics are used in the analysis of the past and present demographic status of a population as well as its prospective growth; in serving the administrative and research needs of public health agencies in connection with the development, operation, and evaluation of public health programs; and in basic research and analysis of the health, survival, and longevity of a population or some group within it. Death statistics are needed to conduct analyses of past population changes as well as past changes in the health and longevity of the population. These analyses are required to make projections of mortality, population size, and other demographic characteristics, and to prepare, interpret, and evaluate projections of the health status of the population.

Population and mortality projections are employed in developing plans for the financing and administration of health programs such as Medicare, the U.S. federal health insurance program for the elderly; for constructing health facilities; and for training and deploying health manpower. Analyses of mortality statistics are essential to programs of disease control. Local health authorities use mortality statistics to design and implement programs for improving public health in local areas. In addition to all these uses, the death registration system serves the demands of individuals and families for documentary proof of death.

Definition and Concepts

To compile, measure, and interpret data on deaths it is necessary to develop and apply a formal definition of this event, however apparent the occurrence of such an

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_3, © Springer Science+Business Media B.V. 2012

event may seem to be. Continuing developments in medical science make it difficult to draw a sharp clinical distinction between life and death, or more exactly, the exact moment when life ends and death begins. For counting deaths over particular time periods in particular places, government agencies must agree on a clinical definition of birth and death for statistical purposes. The situation is more complicated than may appear because in vital statistics systems a distinction must be drawn between three complementary and mutually exclusive events – fetal or pregnancy losses (also known as "fetal deaths"), birth, and death.

In 1950 the World Health Organization recommended the following definition of death [1]:

> Death is the permanent disappearance of all evidence of life at any time after birth has taken place (post-natal cessation of vital functions without capability of resuscitation).

The order of possible vital events is fetal loss, birth, and death. The definition of death can be understood only in relation to the definition of live birth. A death can occur only after a live birth has occurred. A death must be clearly distinguished from a fetal loss, which can occur only before a prospective live birth. Because of the sharp statistical and factual distinction between deaths and fetal losses, we separate the treatment of these two events, covering deaths in the present chapter and fetal losses in Chap. 9.

As recommended by the World Health Organization in 1950, the statistical definition of a (live) birth is,

> . . . the complete expulsion or extraction from its mother of a product of conception irrespective of the duration of pregnancy, which, after such separation, breathes or shows any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles, whether or not the umbilical cord has been cut or the placenta is attached; each product of such a birth is considered live-born.

According to the above definition of birth, the period of gestation, and the state of life or death at the time of *registration*, are not relevant for defining the event of birth, only the time of *occurrence* of the event of birth. The U.S. Public Health Service has recommended for use in the United States the definitions of birth, death, and fetal death adopted by the World Health Organization.[2] The Statistical Office of the United Nations has also adopted the WHO definitions.

---

[1] World Health Organization. (1950). *Official records* No. 28 (Third World Health Assembly 3.6) (pp. 16–17). Geneva: World Health Organization.

[2] U.S. National Office of Vital Statistics. (1950). *International recommendations on definitions of live birth and fetal deaths*. Washington, DC: Public Health Service. See also U.S. National Center for Health Statistics, *Physicians' Handbook on Medical Certification of Death*, 2003 Revision, and *Hospitals' and Physicians' Handbook on Birth Registration and Fetal Death Reporting*, 2003 Revision.

Sources and Quality of Data

Special Sources of Data for Mortality Studies

In Chap. 2, the vital registration system was described as the principal source of death statistics in the more developed countries. National editing and tabulation of death statistics produce comparable data for the principal geographic subdivisions of countries. Currently the scope of the printed publication of the masses of data available from this source is being curtailed and the data are being made available on compact disks or the Internet. In the United States, for example, detailed death statistics for recent years can be obtained in a CD-ROM labeled *Vital Statistics of the United States: Volume II- Mortality*, or in the form of public use microdata samples (PUMS).

The death certificate contains a considerable amount of demographic and socioeconomic data in addition to data relating to the death itself. The death certificate reports age, sex, race, Hispanic origin, educational level, marital status, place of birth, military service, and usual occupation of the decedent. A copy of the U.S. standard certificate of death, as revised in 1999, was shown as Exhibit 2.2 in Chap. 2.

Registration systems are commonly defective in the less developed countries and other sources have to be used to develop accurate vital statistics. A sample of registration areas may be selected so that superior and increased resources can be focused on fewer areas. A more common basis for national estimates of vital statistics is the household survey in which inquiry is made with respect to deaths in the household in the previous year. The results of such surveys can be biased because of the omission of former one-person or larger households that may have been entirely eliminated by death, or by the omission of households in which a member died and which then moved to another registration area or out of the country.

In addition to the regular vital statistics tabulations, the U.S. National Center for Health Statistics and its parent organization, the Centers for Disease Control and Prevention (CDC), prepares several special files of deaths, including:

National Death Index: A cumulative computer database containing the records of all deaths reported in the United States since 1979. This file is available for research purposes only. The index assists health researchers determine which of their research subjects may have died. In addition, the index can be used to match information provided by researchers to secure additional data on the characteristics of decedents. Death records are added to the NDI annually about 12 months after the end of the year.

Compressed Mortality File: A file, available on line, providing data for 1979–1988 on deaths and population for counties, with age, sex, race, and cause-of-death detail. The database permits the calculation of national, state, and county death rates for race, sex, age, cause-of-death groups of interest (Web site, www.wonder.cdc.gov).

Three leading international sources of mortality data: A Human Mortality Database is maintained by the Demography Department, University of California,

Berkeley, and the Max Planck Institute of Demographic Research (Rostock, Germany). It is a cumulative computer file of deaths in the industrial countries covering the last few centuries. The United Nations' *Demographic Yearbook* contains annual data on deaths and their characteristics for the countries of the world. Data on mortality are provided at irregular intervals in the World Health Organization's *World Health Report*, which is issued annually.

Quality of Mortality Statistics

Death statistics are subject to reporting errors of several kinds, both of coverage and content. They may be underreported, reported for the incorrect year, reported for the incorrect area, and assigned characteristics that are incorrect, especially age and cause of death. Even if the number of deaths for some category is correct, it may be linked with population data that are not comparable. The population data that serve as denominators of death rates, are also subject to error as a result of undercounting and misreporting of age and other characteristics.

Completeness of Registration

In the more developed countries, death registration is complete or virtually complete. In the United States all states have adopted laws requiring the registration of deaths as well as the registration of births and the reporting of fetal losses. NCHS surmises that more than 99% of the deaths occurring in the country are registered. In many less developed countries, however, many deaths go unregistered, especially infant deaths, although some early infant deaths are recorded as births prior to their deaths.

Accuracy of Allocation by Time and Place

In some areas a death may be attributed to the incorrect year because it was registered in that year. This tends to happen with deaths occurring in December, which may be registered in January of the following year and are then tabulated with the deaths of that later year. Whether this misassignment has a serious effect depends on the consistency of this misreporting error from year to year. Deaths may erroneously be tabulated for the place of occurrence rather the place of residence. This tends to happen where the death occurs in a hospital and the hospital is located in another jurisdiction. It may be as serious a problem with the more developed areas as with the less developed areas because of the location of hospitals in the large cities and the tendency for deaths to occur in hospitals.

Comparability of Numerator and Denominator

To derive accurate death rates, the deaths and the population must both be comparable with respect to coverage. If the rate refers to an age group or other subgroup, the requirements for accuracy are even greater. The relative coverage and relative accuracy of reporting of age for population and deaths must be similar. In the United States for a number of decades deaths were more completely registered than the population was enumerated, with the result that the crude death rate was overestimated. Errors of undercounting and age reporting can occur in current postcensal population estimates from errors in the same birth cohorts in the decennial census counts. This is because the current postcensal population estimates, on which the current death rates are based, are derived from decennial census counts. Net undercounts, the balance of coverage errors and age reporting errors, affect both population estimates and the death data in varying degrees, so that the death rates would change with any adjustments of either deaths or population. With adjustment of death rates for population net undercounts, the death rates tend to be reduced. Greater understatement of age of deaths than of the population at the very advanced ages in the United States has the effect of underestimating death rates at these ages.

Factors Important in Analysis

Mortality shows significant variations in relation to certain characteristics of the decedent and certain characteristics of the event. These characteristics of the decedent and the event define the principal characteristics that are important in the demographic and epidemiological analysis of mortality. In view of the very close relation between age and the risk of death, age may be considered the most important demographic variable in the analysis of mortality. More than any other characteristic of the decedent or of the event, age provides a clue as to the risks of mortality in a general population. Other characteristics of the decedent of importance are his or her sex and his or her usual place of residence. Elements of importance characterizing the event are the cause of death, the place of occurrence of the death, and the date of occurrence and of registration of the death.

Other characteristics of the decedent important in the analysis of mortality are marital status, literacy socioeconomic status (i.e., educational level, income, occupation), race, nativity, and ethnicity (e.g., country of birth, religion, language, or citizenship). The standard death certificate in the United States also includes among the characteristics of the decedent the age of the surviving spouse (for married persons), number of children born (for females of childbearing age or older), industry (e.g., hospital, school, trucking company), and marital status of the parents. Mortality also varies with the characteristics of the community and the physical environment, but they are not reported on death certificates. These

characteristics include the climate, the altitude, the availability and quality of health facilities, the types of water supply, degree of air pollution, and the quantity and quality of food available.

Some of the variables mentioned, such as the characteristics of the community and the physical environment, are considered as risk factors for mortality, and are treated in Chap. 6. Many of the demographic and socioeconomic characteristics listed are discussed as correlates of mortality in Chap. 7.

Observed Measures

Crude Death Rate

The simplest and most widely used measure of mortality is the crude death rate. It is defined as the number of deaths during a period divided by the population exposed to the risk of death during the period, typically expressed in person-years. Person years is the total number of years lived by the population collectively during the period.[3] When the calculation refers to a single calendar year, as is usual the midyear population can be used as an approximation to person-years or the population exposed to risk. Conventionally, the rate is computed per 1,000 midyear population and may be represented by the symbol, M.

$$\mathrm{CDR_y \text{ or } M_y} = (\mathrm{D_y} \div \mathrm{P_y}) * 1{,}000 \tag{3.1}$$

Table 3.1 displays the crude death rates for selected countries in 2006. The rate is described as "crude" to indicate that it is not disaggregated by age and that it is affected by the underlying age composition of the population. When the time interval is longer, say 2 years, the population at the midpoint of the interval can be multiplied by the length of the interval to derive an acceptable estimate of the population exposed to risk:

$$\mathrm{CDR_{y \text{ to } (y+1)}} = [(\mathrm{D_y} + \mathrm{D_{y+1}})] \div (2\,\mathrm{P_{y+1/2}}) * 1{,}000 \tag{3.2}$$

Although this measure is useful as an indication of the gross contribution of mortality to population growth or decline, it is of limited analytic usefulness. To serve the latter purpose, it is necessary to disaggregate the crude rate into component rates (e.g., age-specific rates) and/or to calculate various adjusted mortality measures. One way of considering the factors underlying changes in the level of the crude death rate is to separate the compositional factors from the etiological (or causal) factors. Age, sex, race, and cause of death represent compositional

[3] In mathematical terms it is the integral of a specified function of the population (px) for the period, usually a year, or ${}_{\mathrm{X}}\int^{x+1} f\,\mathrm{p_x}dx$, where the subscript x refers to time, as do the limits of the integral.

Table 3.1 Crude death rates of selected countries for urban and rural areas: 2006 (Deaths per 1,000 population)

Country	Total	Urban	Rural
Africa			
Egypt	6.3	7.0	5.8
Zambia	12.9	–	–
Zimbabwe[a]	17.2	12.5	19.7
North America			
Cuba	7.2	7.9	5.0
Mexico	4.7	4.6	4.8
United States	8.1	–	–
South America			
Argentina	7.5	–	–
Chile[b]	5.3	5.2	5.7
Venezuela	4.3	–	–
Asia			
Bangladesh	5.6	4.4	6.0
China	6.8	–	–
India	7.5	6.0	8.1
Israel	5.5	5.6	4.2
Kazakhstan	10.3	11.5	8.7
Mongolia	6.5	6.9	5.8
Turkey	6.2	–	–
Europe			
Bulgaria	14.7	12.2	20.7
Estonia	12.9	11.2	16.7
France	8.5	–	–
Hungary	13.1	12.5	13.9
Netherlands	8.3	8.4	8.1
Russian Federation	15.2	14.4	17.3
Sweden	10.0	–	–
Ukraine	16.2	–	–
Oceania			
Australia	6.5	–	–

Source: United Nations Statistical Office, (2008). *Demographic yearbook 2006*. New York: United Nations
–not available
[a]For 2002
[b]For 2005

factors although, from another view, considering the biological influences in each of these, they all may be considered etiological. Measures taking account of these variables can inform us as to their relative importance in the total (crude) rate. Causal factors are represented by the underlying explanatory variables that cause the death rate to change, e.g., marital status, educational level, income status, and housing conditions. I discuss only age and cause of death as compositional factors in this chapter and the remaining factors in Chap. 7.

Age-Specific Mortality

Age-Specific Rates

The most important factor accounting for variations in mortality is age. For analytic purposes, we need to consider the array of age-specific numbers and rates instead of merely the total number of deaths and the crude death rate. Age-specific rates are basic elements in the further analysis of mortality. There are several possibilities in structuring age-specific rates, and these differ mainly in the selection of the denominator. The two possibilities discussed here are designated central death rates and cohort death rates (also called mortality rates or probabilities of dying). In the next chapter, I add another type of age-specific rate called a hazard rate.

Central death rates. An age-specific (central) death rate is defined as the ratio of deaths at a given age (or limited range of ages) to the midyear population at that age (or limited range of ages) in a given year. Conventionally the rate is expressed per 1,000 midyear population and is represented by the symbol, m_x ($_nm_x$ for a group of n ages). Strictly, for the denominator, the figure desired is the person-years lived during the year, that is, the cumulative exposure of the population during the year. As with the crude rate, this quantity is assumed to be satisfactorily approximated by the midyear population, especially for a period of 1 year. The computation formula for a single age is:

$$\mathrm{m_x} = (\mathrm{d_x} \div \mathrm{p_x}) * 1{,}000 \tag{3.3a}$$

For age groups in a single calendar year, the computation formula is:

$$_{\mathrm{n}}\mathrm{m_x} = (_{\mathrm{n}}\mathrm{d_x} \div {}_{\mathrm{n}}\mathrm{p_x}) * 1{,}000 \tag{3.3b}$$

The measure is described as a central rate because the midyear population is used in the denominator. Note that the calculation of age-specific death rates eliminates the effect of the observed age distribution on the overall pattern of mortality.

Age-specific death rates vary widely with age. A set of age-specific death rates displays a characteristic U-shaped or J-shaped pattern over the age cycle, depending mainly on the level of the infant mortality rate. In the least developed countries, a U-shaped pattern is typical; in the more developed countries, a J-shaped pattern is typical. Whether the age-specific death rate is high in infancy or not, it declines to a minimum in the early teen ages, and then rises steadily until the very advanced ages. See Fig. 3.1a, b for representations of the variation of deaths and age-specific death rates according to age in the United States in 2005, and Table 3.2 for sets of age-specific death rates for selected countries.

Cohort death rates or mortality rates. The second type of age-specific "death" rate is a death rate in cohort form, called a mortality rate, or probability of dying, and is symbolized by q_x. In preparing life tables or the survival rates derived from them, central death rates need to be converted to death rates in cohort form. This is done

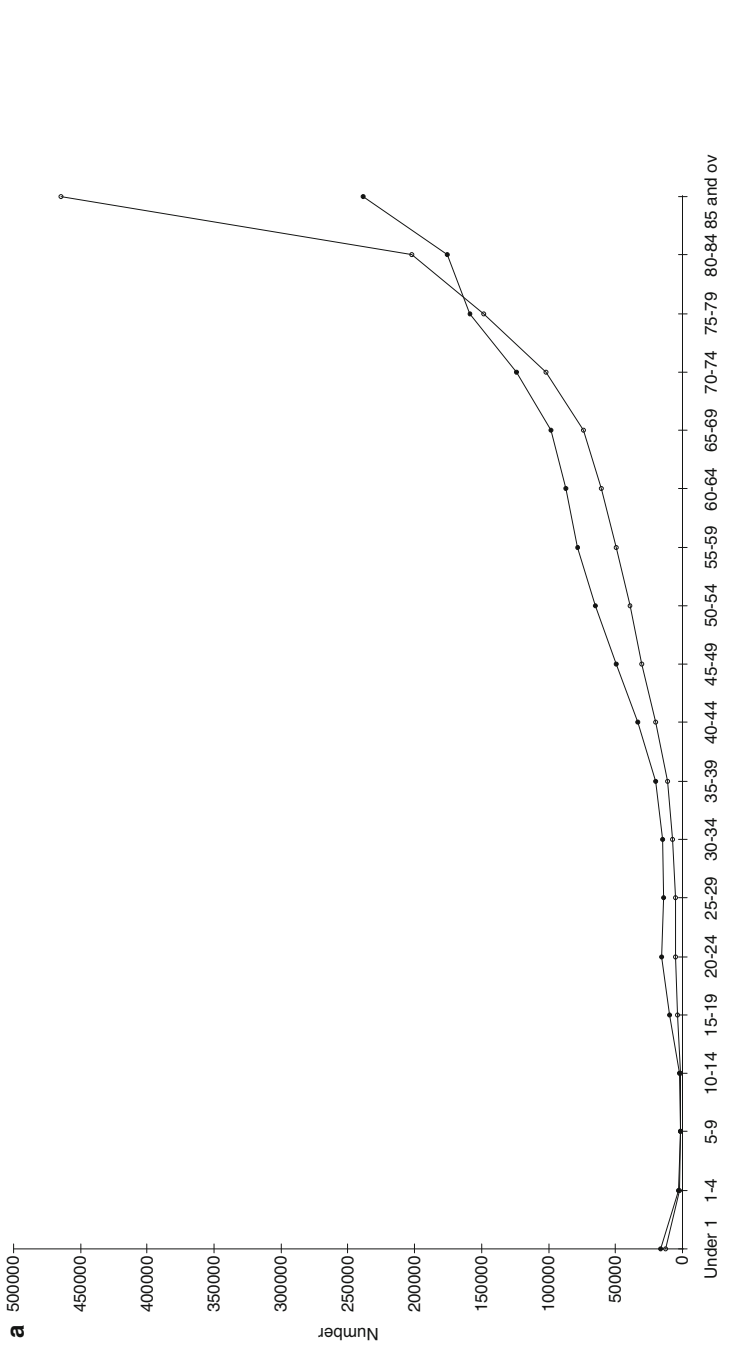


Fig. 3.1 (**a**) Deaths by age: United States, 2005 (Note: The curves are slightly distorted in showing ages 85-and-over as if it were a 5-year age group. The adult modal age group is 80–84years for males and 85–89 years for females; NCHS 2008). (**b**) Age-specific death rates by age: United States, 2005 (Note: The curves are slightly distorted in showing ages 85-and-over as if it were a 5-year age group; Source: U.S. NCHS 2008)

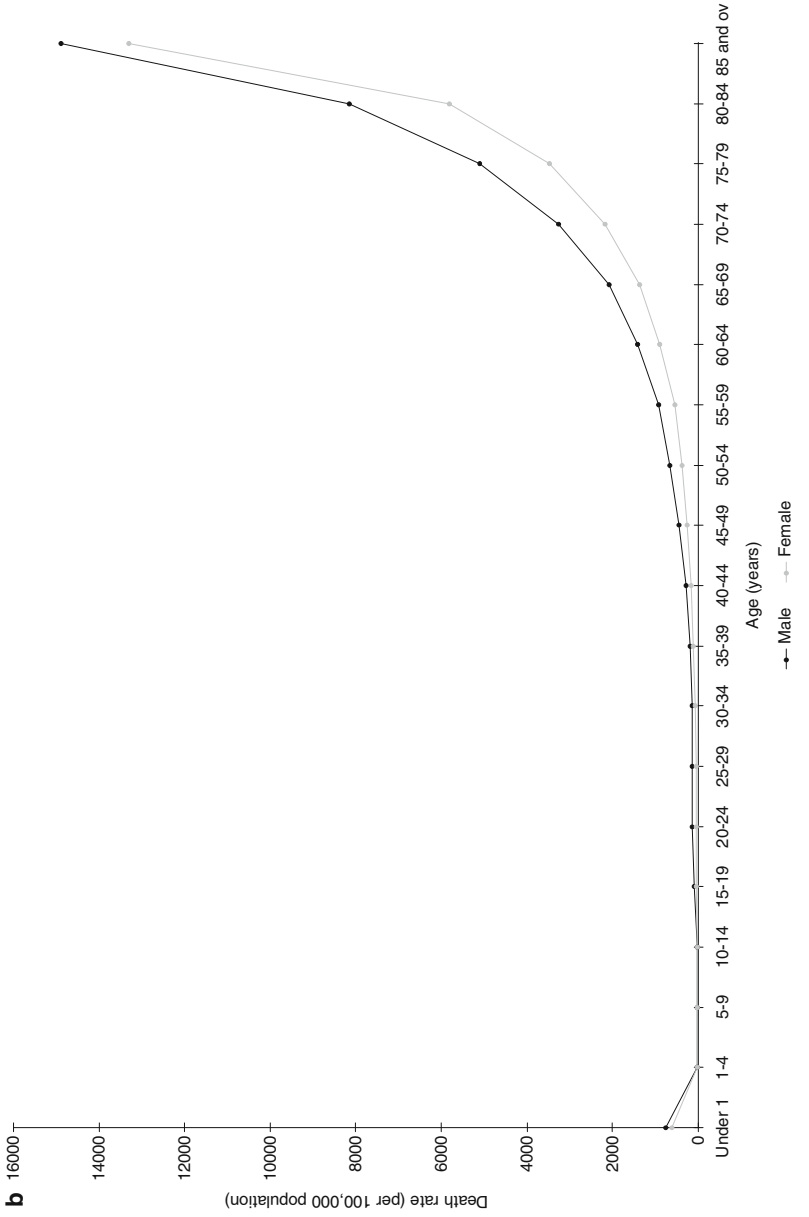


Fig. 3.1 (continued)

Table 3.2 Age-specific death rates for selected countries: 1998–2006 (Deaths per 1,000 population in age group)

Age group	Sweden (2006)	Egypt (1999)	Malawi (1998)	U.S. (2004)	Japan (2006)	Kazakhstan (2006)	Mexico (2006)
Total	10.0	6.4	20.9	8.2	8.5	10.3	4.7
Under 1	2.9	34.3	122.0	6.9	0.7[a]	14.5	15.7
1–4	0.2	2.4	46.4	0.3		1.2	0.7
5–9	0.1	0.7	11.6	0.1	0.1	0.5	0.7
10–14	0.1	0.6	7.8	0.2	0.1	0.5	0.3
15–19	0.3	0.8	6.6	0.7	0.3	1.1	0.7
10–24	0.5	1.2	12.0	0.9	0.4	2.2	1.0
25–29	0.4	1.2	11.7	1.0	0.5	3.6	1.2
30–34	0.5	1.4	14.6	1.1	0.6	4.7	1.5
35–39	0.7	2.1	14.5	1.5	0.8	5.5	1.9
40–44	1.1	3.3	17.6	2.3	1.3	7.4	2.5
45–49	1.7	5.8	16.9	3.5	2.0	10.0	3.8
50–54	3.1	10.3	15.4	5.1	3.1	14.0	5.9
55–59	4.9	15.5	22.1	7.3	4.7	19.3	9.1
60–64	7.9	22.9	19.1	11.4	7.1	26.3	13.5
65–69	12.5	43.1	19.3	17.3	10.5	37.8	20.7
70–74	21.2	71.3	22.7	26.8	17.3	53.4	32.0
75–79	36.4	199.5[b]	24.4	42.0	29.6	80.8	50.4
80–84	66.8		33.2	67.9	49.4	115.1	79.1
85–89	120.3		51.7[c]	137.8[c]	118.2[c]	180.1	117.4
90–94	212.4					299.6	158.9
95–99	339.8					407.9	243.6
100+	499.3					480.4	432.5

Source: United Nations, Statistical Office. (2009). *Demographic Yearbook 2008*. New York: United Nations

[a]Under 5 years

[b]75 year and over

[c]85 years and over

so that the population at some initial date can be efficiently carried forward to the following age or year. If the process is being carried out in single ages for single calendar years, the deaths for the age and year may be assumed to be distributed rectangularly (i.e., evenly) by age and time, and the following formula can be used to derive mortality rates from observed death data or central death rates:

$$q_x = \frac{d_x}{p_x + 1/2d_x} \quad * 1{,}000 \tag{3.4a}$$

This equation can be changed to its more common form by dividing the numerator and denominator by p_x,

$$q_x = \frac{m_x}{2 + m_x} \quad * 1{,}000 \tag{3.4b}$$

where $m_x = d_x/p_x$. For some countries the data on deaths are tabulated by year of birth and these data can be used to calculate mortality rates for birth cohorts directly. If the death statistics are grouped with regard to age, very different formulas are employed since the distribution of deaths cannot be assumed to be rectangular over broader age intervals (see Chap. 4).

Infant mortality rate. Infant mortality rates are often employed in comparative studies of the health and socioeconomic status of nations. They are an important type of age-specific "death" rate in cohort form, or mortality rate. An infant mortality rate relates the deaths of children during the first year of life to the population at risk of experiencing death in infancy. The infant mortality rate cannot be computed by formula (3.3) because infant deaths are not distributed rectangularly, as the above formula assumes, but are concentrated heavily toward the very early days and weeks of life. This means that a disproportionate share of the infant deaths occurred to babies born in the same year and suggests that the infant mortality rate should be based entirely or largely on births of the same year.

The infant mortality rate is conventionally defined as the number of infant deaths during a year per 1,000 births during the *same* year. The formula is, therefore,

$$q_0 = (d_0 \div B) * 1{,}000 \tag{3.5a}$$

(When the midyear population under 1 year of age is used in the denominator of the rate instead of births, the rate is designated a death rate, as are the other central death rates for higher ages.)

If the number of births does not fluctuate much from year to year, the conventional infant mortality rate will represent rather well the probability of an infant dying during a year. If there are sharp fluctuations in the number of births between years, however, the conventional infant mortality rate will give a distorted indication of the level and trend of infant mortality. It is desirable then to adjust the conventional infant mortality rate for the true population exposed to risk. To define an adjusted infant mortality rate that is based on the population actually at risk, it is necessary to take account of both the number of births in the prior year and the number of births in the current year. On the basis of detailed data on infant deaths by age at death in the current year, it is possible to determine the exact proportion of all births in the preceding year that are at risk of dying in the current year. Using these proportions, a formula for the adjusted infant mortality rate for year y can be set down:

$$q_0^y = d_0^y \div \left[f B_{y-1} + (1 - f) B_y \right] \tag{3.5b}$$

where f represents the proportion of births in the preceding year (B_{y-1}) that are at risk of dying in the current year and $1 - f$ represents the proportion of births in the current year (B_y) at risk of dying in this year.

These proportions have been observed to vary on the basis of the level of infant mortality. The lower the level of infant mortality is, the lower the value of

Table 3.3 Separation factors corresponding approximately to selected levels of the infant mortality rate

	Model life tables		U.S. life tables	
Infant mortality rate	**f′**	**f″**	**f′**	**f″**
200	.40	.60	.32	.68
150	.33	.67	.32	.68
100	.25	.75	.25	.75
50	.20	.80	.16	.84
25	.15	.85	.14	.86
15	.05	.95	.14	.86
10	NA	NA	.14	.86
5	NA	NA	.12	.88

Source: Set 1 is based on a regression analysis relating infant mortality rates and separation factors implicit in the series of model life tables shown in Coale, A. J., & Demeny, P. (1966). *Regional model life tables and stable populations*. Princeton, NJ: Princeton University Press

Set 2 is based on the historical series of life tables for the United States published by the U.S. Census Bureau and the U.S. National Center for Health Statistics

f is, and the higher the level of infant mortality the higher the value of f. The separation factor, f, as it is called, varies from about.40 for very high levels of infant mortality to about .10 for very low levels of infant mortality (Table 3.3). Where infant mortality is low, a large proportion of infant deaths is concentrated at the very beginning of life and most of the infant deaths occur to births of the same year, but when infant mortality is high, a substantial proportion of the infant deaths is dispersed among the older ages of infancy and occurs to births of the prior year.

Figure 3.2a shows the relation of time and age for births and deaths (i.e., events) and population (i.e., exposure), and the segments of the deaths used in computing the conventional and adjusted infant mortality rate. Such a diagram relating time and age is called a Lexis diagram. The 2-dimensional Lexis diagram indicates only the relation between age and time and provides no information with respect to the magnitudes of the deaths and population. The diagram shows the relation among births in three successive years ($y-1$, y, and $y+1$), deaths under 1 in these years, and the number of survivors reaching their first birthday in these years (l_1). D'_y and D''_y together make up infant deaths in year y; $D'y$ is the portion occurring to births in year y (B_y) and D''_y is the proportion occurring to births of the previous year (B_{y-1}). To include graphic information regarding the number of infants who died in each year, we need a 3-dimensional Lexis diagram. An abbreviated illustration of such Lexis diagram is presented as Fig. 3.3b. Many more elaborate versions, called population surface charts, are possible.

Consideration of the relation of the infant death rate to the infant mortality rate should aid the reader in understanding the relation between (central) death rates and mortality rates (probabilities) better. This relation is crucial in the interpretation and calculation of life tables to be considered in the next chapter.

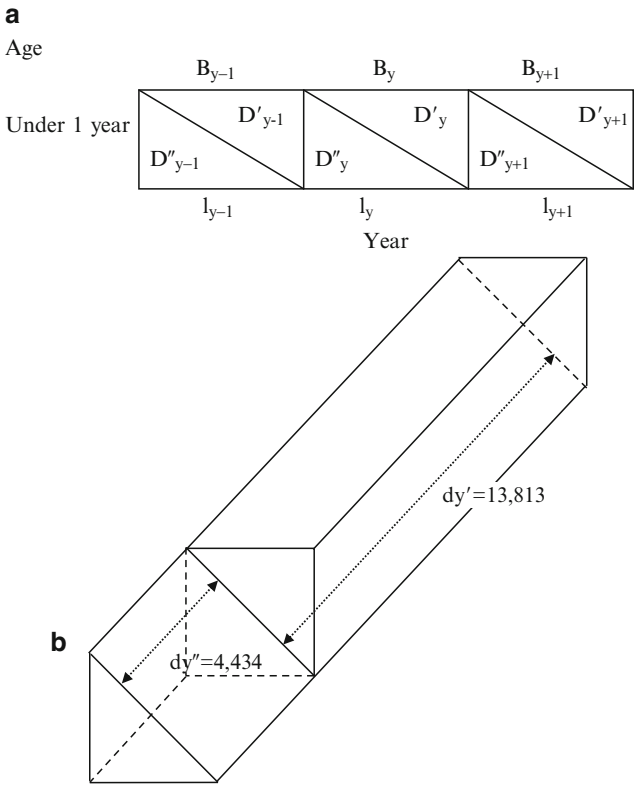


Fig. 3.2 (**a**) Two-dimensional Lexis diagram depicting the relation of age and time in the analysis of infant mortality (Note: D_y – Infant deaths in year y; D'_y – Infant death in year y occurring to births of year y; D''_y – Infant deaths in year y occurring to births of year $y-1$; B_y – Births in year y; l_y – Number reaching first birthday in year y) (**b**) Three-dimensional Lexis diagram depicting the relation of age, time, and number of infant deaths in the analysis of infant mortality (Note: $dy' =$ Infant deaths to births of the current year (partial number) represented as a 3-dimensional triangular extension of the upper section of the basic 2-dimensional square; $dy'' =$ Infant deaths to births of the prior year (partial number) represented as a 3-dimensional inverted triangular extension of the lower section of the basic 2-dimensional square; This is a rough illustration of a three-dimensional figure depicting the relation of age, time, and numbers of infant deaths in a given year. The reflection of the upper triangle of the central two-dimensional square (see Fig. 3.2a) is intended to be 3.1 times in length compared to the reflection of the lower triangle. Inasmuch as infant deaths in a year occurring to births of the same year are not spaced equally within the year, the central square ideally should not be divided into equal triangles and the projections of the two segments of the square should be reshaped accordingly)

Mathematical modeling of human mortality. Many attempts have been made to find a mathematical formula or model describing the pattern of changes in age-specific mortality over the life cycle. Interest in such a model has been motivated by the desire to find a law of mortality, to develop a basis for imputing death rates where mortality data are deficient, to discover what the model suggests as to the pattern of death rates at the advanced ages and for human longevity, and to provide cues as to

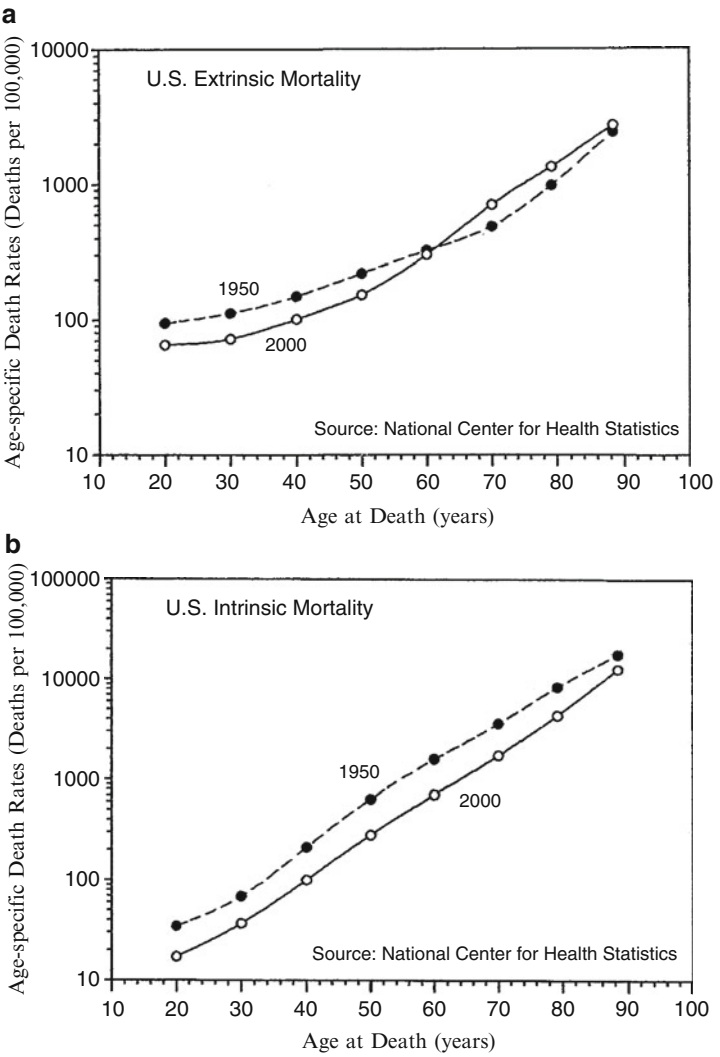


Fig. 3.3 (**a**) Comparison of age-specific exogenous (extrinsic) death rates for ages 15 years and over, by age, for the United States: 1950 and 2000 (Source: Reprinted with kind permission of Springer Science+Business Media B.V. from Carnes et al. (2006), Figure 6. Primary source: U.S. NOVS (1954) and NCHS (2002)). (**b**) Comparison of age-specific endogenous (intrinsic) death rates for ages 15 years and over, by age, for the United States: 1950 and 2000 (Source: Reprinted with kind permission of Springer Science+Business Media B.V. from Carnes et al. (2006), Figure 7. Primary source: U.S. NOVS (1954) and NCHS (2002))

the possibilities for declines in future death rates. Thus, mortality models are used in connection with typing and comparing mortality distributions, interpolating or graduating ("smoothing") observed mortality data, constructing model life tables, and projecting mortality.

A model is a generalized representation of a series of data or the relations between various series of data. A model of the age trajectory of death rates is a concise mathematical description of the change in death rates with age. The goal is to express the level and pattern of age variation in mortality in a way that eliminates random fluctuations, bias, and variance from the actual mortality series – mostly data errors – while essentially preserving the "underlying" level and pattern of mortality. Note that two basic dimensions of the schedule of death rates – level and pattern – are being distinguished; the first refers to the overall magnitude of the rates and the second to their magnitudes relative to one another. One guideline in constructing a mathematical model is to summarize the level and pattern of the rates in terms of the fewest parameters necessary to represent them well. Parameters are mathematical values, usually constants, that determine the characteristics of the equation used to represent the model.

Sometimes the model of the schedule of age-specific mortality is primarily used to eliminate fluctuations in the data (i.e., graduating or smoothing them) or fill in missing "cells" in the series. A guideline in graduating a schedule of observed age-specific death rates is to achieve the optimum balance between describing the true or underlying level and pattern of mortality while retaining real fluctuations in the data, even those that represent deviations from the underlying law.

The most prominent model for a set of age-specific death rates is the Gompertz formula, which assumes that death rates rise at a constant rate between the end of childhood and late adult age. Benjamin Gompertz (1825) attributed death to two groups of causes: The "deterioration of the power to withstand destruction" and "chance." His formula allows only for the first of these factors. In positing that death rates increase by a constant percentage, he is assuming that the logarithms of the death rates increase linearly with increasing age. Specifically, the Gompertz "law" states that, over a significant portion of the age span, the "force" (rate) of mortality rises by a constant factor for successive equal age intervals. This is equivalent to positing that the number of ages it takes the death rate to double is a constant.

The Gompertz formula for the age-pattern of the force of mortality is as follows:

$$\mu_x = Bc^x \tag{3.6}$$

where μ_x is the rate of mortality at age x and B and c are constants. Substituting μ_{20} for B in this equation as the base mortality value (since only adults are covered) and expressing the equation in logarithmic form, we have:

$$\mu_x = \mu_{20}c^x \tag{3.7a}$$

$$\text{In } \mu_x = \text{In } \mu_{20} + x \text{ In } c \tag{3.7b}$$

$$\text{In } \mu_x = \text{In } \mu_{20} + bx \tag{3.7c}$$

where $b = ln\ c$, the rate of increase (a constant) in the force (rate) of mortality at age x. If we replace: μ_{20} by R and express the equation using natural numbers, we have another common formulation of Gompertz's law.

$$\mu_x = Re^{bx} \tag{3.8}$$

This is a simple parametric model with the two parameters, R and b. R varies with the level of mortality but is constant for any age schedule; b is a constant over the age range and gives the rate of increase in mortality with age. As suggested above, the Gompertz formula may be a good fit for the ages up to about age 85, but it overstates death rates at the most advanced ages.

Makeham (1865) proposed a modification in the formula to allow for chance fluctuations in the death rate:

$$\mu_x = A + Re^{bx} \tag{3.9a}$$

$$\mu_x = A + \mu_{20}\, e^{bx} \tag{3.9b}$$

The formula adds a constant A to the estimate of the mortality rate at each age. The additional parameter A allows for the underestimation of mortality at the ages under 40 by the Gompertz formula, but it fails to allow for the overestimation of mortality at the oldest ages by the formula.[4]

Thatcher (1999) and Thatcher et al. (1998) dealt with the overestimation of mortality at the highest ages by proposing a logistic model[5]:

$$\mu_x = \frac{Re^{bx}}{1 + Re^{bx}} + A \tag{3.10}$$

where there are three parameters, R, A, and b, for any age-schedule of mortality. This formula gives about the same results as the Makeham formula at the lower adult ages because the denominator $1 + Re^{bx}$ approximates 1. At the highest adult ages, however, the two models move apart, as the Thatcher model converges to 1. The Thatcher model provides a relatively good fit to the entire adult age range of death rates.

Bongaarts (2005) has proposed a modification of the Thatcher model on the ground that it overestimates mortality between ages 60 and 80. He continues the use of the distinction made by Makeham and Thatcher between intrinsic (internally caused) mortality and extrinsic (externally caused) mortality, describing the mortality schedule in terms of a "senescent" component and a "background" component. In Bongaarts' logistic model at the ages over about 75 the rate of increase with age in the senescent death rate declines and at the very high ages

[4] It is important to note that neither Gompertz nor Makeham intended that their formulas be applied to the advanced ages. Gompertz himself intended that his formula apply only to the population aged 20–60 and Makeham extended this range to age 80.

[5] A logistic model is a special model where the "population" tends toward a specified asymptotic upper limit. The general form of the logistic equation is $f(x) = a \div (1 + be^{-cx})$, where a, b, and c are positive numbers. Typically $f(x)$ denotes the population as a function of time x and the population tends toward a limit represented by a in the equation. The curve has a sigmoid shape, and hence upper and lower asymptotes and a point of inflection at midway.

the senescent death rate approaches 1.0. The background death rate does not vary with age, but does vary over time and place.[6]

Rate of aging and doubling years. One way of describing the pattern of a set of age-specific death rates is in terms of its rate of aging, which is defined as the number of ages required for the death rate to double. This calculation should be confined to the ages where the death rate is rising. It should, preferably, be confined also to the ages where deaths due to external/extrinsic causes (e.g., accidents, infectious diseases) are minimal. Finally, it should exclude the most advanced ages, say the ages over 80 or 85, where mortality may be affected by forces different from those in the intermediate range of ages. Accordingly, I use ages 30–80. The faster the death rate rises, the higher the rate of aging and the fewer the number of ages over which the death rate doubles. This number may be determined by applying the equation for simple exponential growth. To obtain the number of doubling years, first calculate the average rate of increase in the age-specific death rate (or mortality rate) over the age interval 30–80; then, assuming that the death rate doubles, calculate the doubling years using the previously determined average rate of increase in the death rate. Illustrative calculations with U.S. vital statistics data for 2003 are as follows:

1. Solve for r in the equation $m_{80} = m_{30}e^{rx}$, where r = rate of increase and x = number of ages:

a. $m_{80} \div m_{30} = e^{rx}$	a. $.059351 \div .0010.0 = e^{50r}$
b. $\text{In}\,(m_{80} \div m_{30}) = rx$	b. $\text{In}\ 57.6223 = 50r$
c. $\text{In}\,(m_{80} \div m_{30}) \div x = r$	c. $4.05391 \div 50 = .08108 = r$

(3.11)

2. Solve for x, given r in (1c), on the assumption that the rate doubles:

a. $2 = e^{rx}$	a. $2 = e^{.08108x}$
b. $\text{In}\ 2 = rx$	b. $\text{In}2 = .08108x$
c. $\text{In}\ 2 \div r = x$	c. $.69315 \div .08108 = 8.523 = x$

[6]Other logistic formulas with additional parameters have been proposed by Horiuchi and Wilmoth (1998), Thatcher et al. (1998), and Beard (1971). In addition to these formulas, several others have been proposed to model the entire age range – childhood, adulthood, and old age. They employ a component technique, that is, one that describes mortality in terms of particular segments of the age cycle. The formulas usually have three component parts and include several parameters to allow for changes in the parts. The Heligman and Pollard (1980) model is an additive component model with three components and eight parameters. The Thiele (1872) model also has three components: A decreasing Gompertz curve for the childhood ages, a normal curve for accident mortality, and an increasing Gompertz curve for the older ages. Mode and Busby (1982) also proposed a three-component model. Rogers and Little (1994) describe an exponential component model that has four components, each with one or more parameters. (For more information, see Suchindran (2004); Siegel and Swanson (2004)).

The years it takes for the death rate to double using an average rate of aging of 8.1% is 8.5 years. The doubling years may differ for different population groups and at different dates. For example, the male doubling figure was 8.8 years and the female doubling figure was 7.9 years. Even though the female death rates climb more rapidly than the male death rates, females maintain their survival advantage because their death rates start at a much lower level. Over the period 1950–2003, the doubling years for the United States fluctuated around 8.5–8.7, with no discernible trend, but in the period 1900–1950 doubling years fell by one-third. This decline in doubling years may be explained by the fact that the very "high" rates of mortality at the younger ages of adulthood (that rose only slowly with age) in 1900 gave way to much lower rates at the earlier ages (that rose rapidly at the higher rages) in 1950.

The Gompertz/exponential formula is a rough generalization of the age pattern of mortality. The doubling years for a set of age-specific death rates in any year tends to "trend downwards" from age 30 to age 80, typically falling rather gradually until the advanced ages. Other evidence indicates that at the advanced ages the death rate rises more slowly and, as a result, the number of doubling years rises. While age-specific death rates may climb at a relatively steady pace up to about age 85, the rate of increase in death rates tends to fall off beyond this age – roughly in the interval between age 85 and age 105 – and may approximate zero after age 110 (Vaupel et al. 1998; Horiuchi and Wilmoth 1998). The thinness of the data at these very high ages precludes a definitive statement of the pattern. (Information about the pattern of age-specific death rates at the very advanced ages is available only for some low mortality countries.) This general age pattern of death rates, including the deceleration of age-specific rates in later life, has been observed among a large number of vertebrate and invertebrate species as well (Vaupel et al. 1998).

Convergence/crossover of age-specific death rates. It is common for the age-specific death rates of selected pairs of populations to converge and even cross over one another at the older ages. The phenomenon can be demonstrated with U.S. males and females, U.S. blacks and whites, and pairs of geographic areas (e.g., United States and Puerto Rico). In the case of the sexes, ratios of male to female age-specific death rates for the ages 55 years and over in the United States show steady declines both on a cross-sectional and a cohort basis from 1950 to 2000 (Table 3.4). Numerous pairs of states (e.g., Mississippi and Missouri) also show a crossover or at least a convergence in age-specific death rates at the higher ages (U.S. NCHS (2002).

Convergence or crossover of the rates for pairs of populations occur typically when the two populations differ sharply in their socioeconomic characteristics and the different socioeconomic groups have differing health risks. This is particularly applicable to the crossover of the races in the United States. The black and white rates crossed at about age 75 in 1950 and at about age 90 in 2000 (Table 7.6). Part of the large shift in the crossover ages may be due to the use of mortality rates from a more valid source in 2000 – Medicare records – than in 1950. Some argue that the crossover of black and white death rates is a statistical artifact (Coale and Kisker 1986; Coale 1990). The rates for blacks are much more affected than

Table 3.4 Ratio of male to female age-specific death rates 55 years and over, by age, for the United States, 1950–2006

Age group (years)	2006	2000	1990	1970	1950
55–59	1.670	1.647	1.773	2.048	1.752
60–64	1.579	1.594	1.780	2.116	1.661
65–69	1.544	1.571	1.779	2.016	1.576
70–74	1.487	1.598	1.763	1.817	1.403
75–79	1.469	1.467	1.714	1.613	1.287
80–84	1.409	1.390	1.591	1.410	1.205
85 and over	1.121	1.124	1.265	1.148	1.128
All ages	1.011	1.004	1.131	1.350	1.354
Age-adjusted	1.406	1.410	1.741	1.749	1.455

Source: U.S. National Center for Health Statistics, vital statistics reports for indicated years

those for whites by coverage errors and age misreporting in the censuses and by age misreporting on death certificates. The argument that the crossover is a statistical artifact is not mainly empirical but analogical, however. It is based on the fact that many countries with data of good quality do not show the crossover. Both empirical and theoretical considerations support the black/white crossover, however. Death rates based wholly on the Social Security Administration's Master Beneficiary Record still indicate a crossover for the two races at a very high age (Kestenbaum 1992).

Numerous paired population groups within and between countries having data of good quality show the convergence/crossover phenomenon. The crossover may be explained by the effect of mortality selection in a population that is heterogeneous with respect to health risks (Manton et al. 1979; Manton and Stallard 1984; Vaupel et al. 1979). If populations are heterogeneous in their health risks, a crossover, or at least a convergence, can occur even if one population has markedly higher earlier mortality. Those who have survived the environmental stresses of their younger years may be destined to live a relatively long life. Hence, these more robust persons make up a larger proportion of the survivors at the older ages and the older population has relatively lower death rates. As the socioeconomic and health characteristics of two populations become more similar and the quality of the data becomes more accurate, mortality convergence and crossover between them should decline and eventually disappear (Nam 1995). Nam (1995) and others have shown that, for a wide array of comparisons, the general cause-category accounting for much of the convergence is cardiovascular diseases (e.g., heart diseases, stroke).

The concept of heterogeneity in population composition is very important. The role of heterogeneity should always be kept in mind when interpreting temporal changes in demographic and socioeconomic phenomena. The individuals in the birth or other cohort may not change their characteristics (e.g., health); nevertheless, the shares of persons with differing characteristics (e.g., healthy and unhealthy) within the cohort change over time because of differences in survival rates among the different subgroups. This fact affects the overall trend in the characteristics of

the group. The characteristics of survivors are always affected by the fact that the less healthy members of the group drop out at a higher rate than the more healthy members. The characteristics of a so-called homogeneous group are modified by the fact that the group is still heterogeneous with regard to some characteristics that have different health prospects.

Parameters of Age Distributions

The level and age pattern of deaths and death rates vary according to area, reference date, and the demographic and socioeconomic characteristics of the population. In order to compare the age distributions of deaths and death rates for different populations, we need ways of statistically characterizing them with respect to level and pattern. We call these descriptive measures of the age distribution their parameters. The level of the age distribution of deaths and death rates may be summarized in terms of three common "averages." The pattern or shape of the age distribution of deaths and death rates may be summarized in terms of various relative measures of dispersion around the average.

Measures of level: Average age at death. Universally deaths show a characteristic pattern according to age (Fig. 3.1). The number is high in infancy, falls to a minimum at about age 10, then rises gradually until a new peak is reached in late adult life, and finally turns down until the remaining members of the "cohort" are eliminated – about the century mark. The curve of age-specific deaths is bimodal, that is, has two peaks, although the second peak may not be evident in some very low mortality countries until the very advanced ages (requiring detailed data over age 85). The three common measures of central tendency, or "averages," used as summary measures of the level of the age distribution of deaths are the arithmetic mean, median, and mode. The arithmetic mean and the median may be computed for age-specific death rates in the same way as for absolute numbers, with the rates being treated as though they were absolute numbers.

We define the mean age at death in a given year in a given population as the arithmetic mean of the ages of the decedents in that population in the year:

$$\mathrm{a_{dM}} = \sum \mathrm{a_d} \div \mathrm{D} \tag{3.12a}$$

where a refers to age, d to deaths, M to arithmetic mean age, and D to the total number of deaths. For grouped data, the corresponding formula is

$$\mathrm{a_{dM}} = \sum \mathrm{fa_{midd}} \div \mathrm{D} \tag{3.12b}$$

where a_{midd} refers to the midpoint age of the grouped data on deaths and f refers to the number of deaths in the age interval.

The median age at death in a given population in a given year is the midpoint of the ages of decedents in that population. It corresponds to the age below and

above which 50% of the deaths fall when the deaths are ranked by age of decedent. With grouped data, it is necessary to interpolate within the middle age group and by convention this interpolation is usually done on the assumption that the deaths are distributed rectangularly within the interval.[7]

The modal age at death is the adult age at which the maximum number of deaths occurs. Similarly, with grouped data, assuming equal-size class intervals, the age group with the maximum number of deaths is taken as the modal age group. When reference is made to the modal age at death, the reference is normally to the second, or adult, age of death. At this age the diminished number of survivors begins to offset the effect of rising death rates, and the number of deaths begins to fall. (The other modal age is at infancy but is usually disregarded in such analyses.)

While these measures may be used to summarize the changing level of the entire distribution of deaths, for the analysis of longevity it is desirable to restrict the analysis to the deaths at the adult ages, excluding the deaths below age 15, 18, or 20. Hence, in such studies the mean age and median age of deaths may be computed for the age range 20 and over; the determination of the modal age at death normally excludes childhood deaths.

Measures for Comparing Age Patterns

I turn now to measures of the age pattern of deaths and death rates. These are intended to be independent of the level of the mortality data. Use of these measures makes comparisons of the age pattern of deaths and death rates between different populations directly possible – both comparisons among various human populations and comparisons of human populations with animal species. The basic measures described are the ratio of modal deaths to infant deaths, the coefficient of variation, and the relative interquartile range. The coefficient of variation and the relative interquartile range take account of more information on the age distribution of deaths than the ratio of modal deaths to infant deaths, so they are to be preferred for a more exact comparative analysis of disparate populations.

Ratio of Modal Deaths to Infant Deaths

The relative magnitude of deaths at the modal age and infancy varies widely over time and from area to area. This variability is illustrated by the following ratios of modal deaths to infant deaths for population groups in the United States, spanning several decades:

[7] A rectangular distribution of a variable (e.g., deaths) in an age group (e.g., 5-year group) is one in which equal parts (e.g., one-fifth) of the variable fall in equal intervals of the age group (e.g., single ages). For a discussion of interpolation techniques and an illustrative calculation of the median age see Siegel and Swanson (2004).

Population, year	Deaths <1 year	Modal age group, deaths	Ratio
White females, 1993	9,275	80–84 – 158,475	17.09
White females, 1970	23,151	80–84 – 116,567	5.04
White females, 1950	34,423	75–79 – 75,701	2.20
Black males, 1950	11,395	65–69 – 8,892	0.78

Source: U.S. NCHS, various vital statistics publications

When mortality is relatively high, the ratio is low – below one – and the age distribution of deaths is rather symmetrical. When mortality is relatively low, the ratio is high – above 15 and the age distribution of deaths is more J-shaped. Note that this measure is not applicable to death rates because the distribution of death rates has no determinate adult mode.

Coefficient of Variation

The coefficient of variation of the age distribution of deaths represents the ratio of the standard deviation to the mean of the distribution. The mean and standard deviation of the distribution are each affected by the absolute size of the numbers in the original age distributions, but their ratio, the coefficient of variation, eliminates this effect. The standard deviation is computed in the conventional way from the distribution of observed deaths by age:

$$\mathrm{SD}(\text{or } \sigma) = \sqrt{\left[\sum (\mathrm{a_d} - \mathrm{a_{dM}})^2 \div \mathrm{D}\right]} \tag{3.13}$$

where a_d refers to the age of each of the deaths, a_{dM} to the mean age of the deaths, and D to the total number of deaths. A smaller standard deviation corresponds to a smaller dispersion of the deaths with respect to age. To derive the coefficient of variation, divide the standard deviation by the mean age of the distribution:

$$\mathrm{CV} = \mathrm{SD} \div \mathrm{a_{dM}} \tag{3.14}$$

Comparisons between distributions of deaths by age, carried out by use of the coefficient of variation, allows for different levels of the distribution of deaths, populations of different sizes, and species that have different life spans.

Relative Interquartile Range

The interquartile range (IQR) is computed as the difference between the first and third quartiles in the age distribution; that is, it is the age band in which the middle 50% of the deaths fall with respect to age.

$$\mathrm{IQR} = \mathrm{a_{d2}} - \mathrm{a_{d1}} \tag{3.15}$$

where a_{d1} and a_{d2} correspond to the ages where the cumulative percent distribution of deaths equals 0.25 and 0.75. As mortality is reduced to lower and lower levels, typically the age distribution of deaths becomes less dispersed and the interquartile range narrows. For example, in the United States the interquartile range was 25.8 years in 1950 and 20.9 years in 2000:

	1950			2000		
	1st quartile	3rd quartile	IQR	1st quartile	3rd quartile	IQR
Age	51.0	76.8	25.8	64.9	85.8	20.9

The relative interquartile range is derived by dividing the interquartile range by the median age, i.e., the 2nd quartile. For example, between 1950 and 2000 the relative interquartile range declined from 39.0 to 27.0:

	1950			2000		
	IQR	Median	RIQR	IQR	Median	RIQR
Age	25.8	66.2	39.0	20.9	77.3	27.0

The effect of the "standardization" of the IQR between these two dates is to increase the indication of the compression of deaths with respect to age, from 0.81 (on the basis of the IQR) to 0.70 (on the basis of RIQR). Computing the relative interquartile range excludes the effect of the level of the distribution and improves the comparison of the age patterns between 1950 and 2000.[8] Table 3.5 presents a series on the relative interquartile range for the United States in the last half century.

A measure of relative dispersion may be based on the mode also, using the standard deviation of the distribution. Such a measure may be calculated as the ratio of the standard deviation to the mode or, alternatively, the ratio of the ages covered by the mode plus and minus one standard deviation to the mode.

Scaling of Distributions

As we have just seen, when age patterns of deaths for populations having very different levels of mortality are to be compared by a summary measure, the measures of dispersion of mortality can be "standardized" by dividing them by the average

[8]Note that, as the median age of deaths rises, the interquartile range and the relative interquartile range tend to fall. This applies to the coefficient of variation and its components also. As Wilmoth and Horiuchi (1999) note, to combine changes in a measure of variation with changes in average age is to combine components that move in opposite directions. It is difficult to separate these two effects and it is difficult, therefore, to interpret the changes in the composite measure.

Table 3.5 Measures of centricity and dispersion of the age distribution of deaths based on vital statistics: United States, 1950, 1970, and 2000

Measure	1950	1970	2000	Increase 1950–1970	Increase 1970–2000
Mode[a]	72.5	77.5	82.5	5.0	5.0
Median	66.2	70.3	77.3	4.1	7.0
1st quartile	51.0	56.9	64.9	5.9	8.0
3rd quartile	76.8	80.0	85.8	3.2	5.8
IQR[b]	25.8	23.1	20.9	−2.7	−2.2
RIQR[b]	39.0	32.9	27.0	−6.1	−5.9

Source: Calculated from U.S. National Center for Health Statistics vital statistics reports for 1950, 1970, and 2000

[a]Approximation

[b]Interquartile range (IQR) and interquartile range as percent of median age at death (RIQR)

values of the age distribution. For comparing the details of the patterns, numerically or graphically, the data in the distributions can be scaled by use of a "representative" number from each of their distributions. This can be accomplished by dividing the original numbers by their mean, median, or sum. In the case where the median or mean is used to standardize the numbers, the transformed distribution pivots around 1.00; where the total of the distribution is used, the transformed distribution *sums* to 1.00 or 100. The purpose of the scaling is to separate the pattern of the distributions from its level (and its life span) and to improve the visual comparison of different distributions being compared.

Let us call the factor used to scale each distribution f_i. If the scaling factor is based on the median of each distribution, we would divide each number in the distribution by its median. If the scaling factor is the total of the distribution, we would divide each number in the distribution by its total, converting each distribution into a proportionate or percent distribution.

An alternative procedure is to adjust each element in the distribution under study by the ratio of, say, the median age of a standard population to the median age of the population under study. The standard population would be unaffected by applying such an adjustment to itself. The adjustment factors would be Md_s/Md_1 for distribution 1, Md_s/Md_2 for distribution 2, and so on. Suppose the distribution of deaths for one population, taken as the standard population s, is being compared with the distribution for the same population a half century earlier, the ratio of median ages (Md_s/Md_j) would be used to adjust the earlier series upward or downward to the general level of the standard population. A similar technique would be applied if distributions of age-specific death rates were being analyzed.

Although the desirability of scaling the data for comparison of various human populations over short periods may be debated, the need for scaling distributions is especially important when the ranges of the distributions are very different as, for example, when comparing the patterns of deaths or death rates observed over long periods of time, or for different species. The life spans of different species may

be quite different. Biologists comparing the mortality patterns of humans, dogs, and mice must apply scaling devices to their data if they want to measure and graphically depict the differences in the patterns. The scaling calculation transforms the age data for the dogs, say, into "human" years of life and the pattern of the transformed data for dogs can now be compared with that of the human data. Comparability is only approximate, however, because the expansion or contraction of data by a multiplier in this way modifies the measured dispersion of the data.[9] For an example of the scaling of distributions in the comparison of humans, dogs, and mice, see Carnes et al. (2006).

Index of Dissimilarity

The index of dissimilarity (ID) is another scaling measure that can be used to compare the patterns of age distributions of deaths or death rates. Traditionally the index of dissimilarity has been used to measure inequality among groups of areas with respect to the distribution of the races, age groups, household sizes, and incomes. Here I want to apply it to various populations in order to measure the differences in the age patterns of their deaths and death rates. To calculate the index of dissimilarity for the distributions of deaths by age for two populations (Table 3.6):

1. Calculate the percent distribution of the deaths by age for the two populations (equivalent to multiplying the deaths by the reciprocal of their total), thus essentially removing the effect of the level of mortality);
2. Take the absolute differences between the percents in step 1 for the same age groups in the two populations;
3. Sum the differences in step 2 over all ages without regard to sign; and
4. Take one-half of the sum in step 3.

The formula is:

$$\mathrm{ID} = 1/2 \sum |r_1 - r_2| \tag{3.16}$$

where r_1 and r_2 are the percents of deaths at a given age in populations 1 and 2. It is important to note that the ID is affected by the number and width of the class intervals, inasmuch as broader intervals and fewer categories balance out differences within the intervals and reduce the index. Hence, be careful that the distributions of age-specific deaths for the various populations being compared have the same number of intervals and intervals of the same width.

The calculation of the ID is illustrated in Table 3.6 with a comparison of the age distributions of deaths in the United States, Mexico, and Kazakhstan in 2006. The

[9] Specifically, if a distribution is multiplied by a factor f, its variance is multiplied by f^2 and its standard deviation is multiplied by f.

Table 3.6 Calculation of the index of dissimilarity for the age distribution of deaths, for the United States, Mexico, and Kazakhstan: 2006

	Deaths Number			Percent distribution			Difference Diff. from U.S,	
Age group	U.S.A.	Mexico	Kazakhstan	U.S.A.	Mex.	Kazakh.	Mex.	Kazakh.
All ages	2,426,264	493,296	157,210	100.0	100.0	100.0	–	–
Under 1	28,527	30,893	4,154	1.2	6.3	2.7	+5.1	+1.5
1–4	4,631	6,075	1,140	0.2	1.2	0.7	+1.0	+0.5
5–14	6,149	6,654	1,156	0.3	1.4	0.7	+1.1	+0.4
15–24	34,887	17,293	4,879	1.4	3.5	3.1	+2.1	+1.7
25–34	42,952	23,453	9,863	1.6	4.8	6.3	+3.2	+4.7
35–44	83,043	30,972	13,952	3.4	6.3	8.9	+2.9	+5.5
45–54	185,031	46,353	22,085	7.6	9.4	14.1	+1.8	+6.5
55–64	281,401	65,344	21,170	11.6	13.3	13.5	+1.7	+1.9
65–74	390,093	87,695	34,733	16.1	17.8	22.2	+1.7	+6.1
75–84	667,338	99,085	32,065	27.5	20.2	20.5	−7.3	−7.0
85+	701,992	77,620	11,242	28.9	15.8	7.2	−13.1	−21.7
Age n.s.	220	1,865	771					
–zero								
Index of dissimilarity = 1/2 sum of differences disregarding signs							20.4	28.7

Source of death statistics: USA, www.cdc.gov/nchs; Mexico and Kazakhstan, www.un.org

results show that the distributions of deaths in Mexico and Kazakhstan are far less concentrated at the advanced ages than in the United States and that, in general, with indexes of dissimilarity over 20, the distributions of the two former countries are considerably different from that of the United States.[10]

There are several other measures of the patterns of age distributions, such as the Theil index, the squared coefficient of variation, the mean logarithmic deviation, the Gini coefficient, and Keyfitz's H, but because they are primarily used in life table analysis (e.g., inequality of life expectancy, variability of life table deaths) and in measuring socioeconomic differentials in mortality and morbidity, they are discussed in Chaps. 4 and 7, respectively.

Cause-Specific Mortality

Classification of Causes of Death

An international standard for the classification of deaths according to cause has been available for over a century. The *International Statistical Classification of Diseases and Related Health Problems* (ICD) was created at the end of the nineteenth century by an international committee for the classification of deaths and diseases. This classification was based on the traditional biomedical model with a focus on the etiology of the diseases. In 1948 the World Health Organization (WHO) took over responsibility for the international classification. In accordance with WHO regulations, the classification of deaths and diseases used by national governments follows the classification recommended by the WHO. The classification is reviewed and modified every 10–20 years so that the disease classification may be consistent with advances in medical knowledge and changes in diagnostic practice.

Tenth Revision of International Classification of Diseases. The Tenth Revision (ICD-10) of the International Classification of Diseases was promulgated by the

[10]To measure differences in the age patterns of age-specific death rates, we can apply the same formula to the rates as to the absolute numbers. Because of the very high level of death rates at the later ages as compared with the earlier ages, particularly under general conditions of low mortality, the distribution of death rates is dominated by the rates at these higher ages. It is desirable, therefore, to reduce the weight of the rates at the highest ages by reweighting them. The reweighting can be accomplished by multiplying the rates by the population age distribution. The age distribution of the population tapers sharply at the advanced ages and so, by the weighting process, less weight is given to mortality at these ages. The calculation would convert the percent distribution of death rates to a different function of deaths by age, namely the percent distribution of deaths:

$$\text{Weighting age} - \text{specific death rates} = (d_a \div p_a)^*(p_a \div \sum p_a) = (d_a \div \sum p_a) = d_a \div P$$

$$\text{Converting to a percent distribution} = (d_a \div P) \div \sum (d_a \div P) = d_a \div \sum d_a = r$$

World Health Organization in 1992–1994. It replaces the Ninth Revision of the International Classification of Diseases (ICD-9), which remained generally in effect up through 1998. The Ninth Revision was implemented in the United States with 1979 mortality data and the Tenth Revision with deaths for 1999. The ICD sets forth not only a classification of diseases but also provides definitions, tabulation lists, and the rules for coding cause of death.

ICD-10 differs from ICD-9 in a number of ways. First, it is far more detailed than ICD-9; ICD-10 has 8,000 categories while ICD-9 has 5,000. Next, ICD-10 uses alphanumeric codes while ICD-9 uses only numeric codes. A further difference is that some additions and other changes were made to the "chapters" of ICD-9. ICD-10 has 21 chapters compared with 17 for ICD-9 (with two supplementary categories). Finally, some changes were made in the coding rules, including the rules for selecting the underlying cause of death. Some of these changes are evident from Exhibit 3.1, which displays the chapter titles and code ranges for the Tenth Revision of the International Classification of Diseases.

Conventionally cause-of-death statistics are based solely on the underlying cause of death. The *underlying* cause of death is defined by WHO as "the disease or injury which initiated the train of events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury." It is selected from the health conditions entered by the physician in the cause-of-death section of the death certificate. The physician may also list *contributory* causes (that is, causes that resulted from the underlying cause and therefore led to death), and *associated* causes (that is, other health conditions present at the time of death), so that the death certificate may contain more than a dozen "causes" of death. The underlying cause of death is conceptually easy to understand, is a well-accepted measure of mortality, and is most useful to public health officials in developing measures to prevent the onset of the chain of events leading to death (U.S. NCHS 1991). The rules for selecting the underlying cause of death are included in the ICD as a means of standardizing the classification of deaths and improving the comparability of mortality statistics among countries. On the other hand, the underlying cause of death may be an arbitrary choice among the several causes contributing to death and it cannot wholly explain the circumstances leading to death.[11]

Quality of Cause-of-Death Statistics

Cause-of-death data are universally affected by similar types of errors – misreporting or omission of age and cause of death. These errors are moderate in the

[11]Currently NCHS uses an automated scheme called ACME ("Automated Classification for Medical Entities") for coding the underlying cause of death for each certificate in accordance with WHO rules. NCHS has developed two supplementary systems, one to automate coding of multiple causes of death called MICAR, and the other to allow for literal entry of the multiple cause-of-death text as reported by the certifier called SuperMICAR. Records that cannot be automatically processed by MICAR or SuperMICAR are manually coded with respect to cause and then further processed through ACME.

Chapter number	Chapter title	Code range
I	Certain infectious and parasitic diseases	A00-B99
II	Neoplasms	C00-D48
III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	D50-D89
IV	Endocrine, nutritional, and metabolic diseases	E00-E90
V	Mental and behavioral disorders	F00-F99
VI	Diseases of the nervous system	G00-G99
VII	Diseases of the eye and adnexa	H00-H59
VIII	Diseases of the ear and mastoid process	H60-H95
IX	Diseases of the circulatory system	I00-I99
X	Diseases of the respiratory system	J00-J99
XI	Diseases of the digestive system	K00-K93
XII	Diseases of the skin and subcutaneous tissue	L00-L99
XIII	Diseases of the musculoskeletal system and the connective tissue	M00-M99
XIV	Diseases of the genitourinary system	N00-N99
XV	Pregnancy, childbirth, and the puerperium	O00-O99
XVI	Certain conditions originating in the perinatal period	P00-P96
XVII	Congenital malformations, deformations, and chromosomal abnormalities	Q00-Q99
XVIII	Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	R00-R99
XIX	Injury, poisoning, and certain other consequences of external causes	S00-T98
XX	External causes of morbidity and mortality	V01-Y98
XXI	Factors influencing health status and contact with health services	Z00-Z99

Exhibit 3.1 Chapter titles and code ranges for the tenth revision of the international classification of diseases Source: U.S. National Center for Health Statistics (2001b), Table B

more developed countries compared with the degree of error in the less developed countries. Even in the more developed countries, cause-of-death statistics are plagued by problems of misreporting and omission of the underlying cause on the death certificate, misselection of the underlying cause by the coding system, and lack of comparability of cause-of-death recording and coding among sequential international classifications, from one area to another, and from one period to another. The underlying cause may not be identified on the death certificate and yet the coder or automated coding system must make a selection. The misrecording or omission of the underlying cause on the death certificate by the physician may be deliberate (as requested by members of the family), but the list of contributing and associated causes may still give a valuable clue as to the underlying cause.

Comparability of classification systems. The changeover to a new classification system every 10–20 years necessarily introduces certain discontinuities between the tabulations of deaths by cause between the years when the changeover occurs.

List number	Cause of death	Estimated comparability ratio[a]
010	Septicemia	1.1949
016	Human immunodeficiency virus (HIV) disease	1.0637
019	Malignant neoplasms	1.0068
023	Colon, rectum, anus	0.9993
027	Trachea, bronchus, and lung	0.9837
029	Breast	1.0056
033	Prostate	1.0134
037	Lymphoid, hematopoietic, and related tissue	0.0042
046	Diabetes mellitus	1.0082
053	Major cardiovascular diseases	0.9981
054	Diseases of heart	0.9858
070	Cerebrovascular diseases	1.0588
071	Atherosclerosis	0.9637
076	Influenza and pneumonia	0.6982
082	Chronic lower respiratory diseases	1.0478
093	Chronic liver disease and cirrhosis	1.0367
097	Nephritis, nephrotic syndrome, and nephrosis	1.2320
108	Certain conditions originating in the prenatal period	1.0658
109	Congenital malformations, deformations, and chromosomal abnormalities	0.8470
112	Accidents (unintentional injuries)	1.0305
124	Intentional self-harm (suicide)	0.9962
127	Assault (homicide)	0.9983

Exhibit 3.2 Estimated comparability ratios for selected causes of death in the United States: 1996 (Based on ninth and tenth revisions of the international classification of diseases)
Source: U.S. National Center for Health Statistics (2001), Table 3.1
[a]Comparability ratios subject to sampling error

Hence, a measure of the effect of the changeover is critical to the interpretation of mortality trends. Comparability ratios, representing the ratio of the number of deaths from a given cause classified according to the later ICD to the number classified with the same cause according to the previous ICD, are designed to measure the discontinuities in cause-of-death tabulations. The ratios shown in Exhibit 3.2 were derived by coding the same deaths occurring in 1996 by both the Ninth and Tenth Revisions, and measure the net effect of shifting to ICD-10 from ICD-9 on the numbers counted in each cause-class shown. A comparability ratio of 1.00 denotes no net effect of ICD-10 on that cause.

Accuracy of reporting and effect on comparability among areas. The use of a standard classification list, although essential for state, regional, and international comparisons, does not assure strict comparability of the tabulated figures. The U.S. National Center of Health Statistics (1982) notes that all the records of cause of death are not reported with equal accuracy and completeness. The reliability and

Table 3.7 Percent of deaths classified as ill-defined or of unknown cause for selected countries: Around 2003

Country	Year	Percent
Australia	2003	0.7
Cuba	2004	0.8
United States	2002	1.2
Canada	2003	1.4
Kyrgyzstan	2004	2.5
Sweden	2002	2.8
Spain	2004	2.8
Mauritius	2004	2.8
Japan	2004	3.4
Kazakhstan	2004	3.4
Ukraine	2004	4.0
Israel	2003	4.4
Russian Federation	2004	4.6
South Africa	2004	12.1
Brazil	2002	13.6
Paraguay	2003	21.4
Haiti	2003	28.9

Source: United Nations, Statistical Office, *Demographic Yearbook, 2004*, United Nations, New York
Note: Country entries for Chapter XVIII8, R00–R89, in the Tenth Revision of the International Classification of Diseases, "Symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified"

accuracy of cause-of-death statistics are largely determined by the ability of the certifier, usually a physician, medical examiner, or coroner, to make the proper diagnosis and by the care with which he or she records this information on the death certificate. An NCHS compilation of 123 studies indicated no definitive conclusions on the quality of medical certification on the death certificate (U.S. NCHS 1982). No country has a well-defined program for systematically assessing the quality of medical certifications reported on death certificates or for measuring the effects of any errors on the levels and trends of cause-of-death statistics.

The percentage of deaths tabulated as of ill-defined or unknown cause or, in the terminology of the International Classification, "Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified" (R00-R99 in the Xth Revision), is one index of the quality of reporting causes of death. This percentage normally indicates the care given to the certification by the medical examiner or the specificity of the medical diagnosis made by the certifier. For the United States, in 2002 the share of deaths in this category ran a little over 1% (1.2%), but for less developed countries, it is not uncommon for the share of deaths reported in this category to exceed 10%. Table 3.7 provides an indication of the international variation in the percent of ill-defined conditions reported in recent years. The absence of data for the countries of Africa gives a more favorable impression of the international situation than is really the case. The percent varies from these country averages for age groups and other demographic and socioeconomic classes.

Extent of multiple-cause reporting. As mentioned earlier, in addition to the underlying cause, numerous contributory causes and associated conditions may be listed on the death certificate. Studies conducted at NCHS, reviewing samples of death certificates, have revealed that contributory causes are listed often and that there is a considerable degree of misclassification of the underlying cause of death.

Stallard (2002) has analyzed multiple-cause mortality at the higher ages (65 years and over) for the 15 leading causes of death in the United States to determine which causes are usually mentioned in association on death certificates and the frequency with which they are co-mentioned. He found strong pairwise associations in 1998 between:

1. (a) Diseases of the heart and (b) diabetes mellitus, hypertension, atherosclerosis, nephritis/nephrosis
2. (a) Cerebrovascular diseases and (b) hypertension, atherosclerosis, diabetes mellitus
3. (a) Chronic obstructive pulmonary disease and (b) pneumonia/influenza
4. (a) Pneumonia/influenza and (b) septicemia, chronic obstructive pulmonary diseases, Alzheimer's disease
5. (a) Diabetes mellitus and (b) hypertension, nephritis/nephrosis, atherosclerosis,
6. (a) Nephritis/nephrosis and (b) septicemia, diabetes mellitus, chronic liver disease/cirrhosis, atherosclerosis
7. (a) Chronic liver disease/cirrhosis and (b) nephritis/nephrosis, septicemia
8. (a) Septicemia and (b) nephritis/nephrosis, pneumonia/influenza, chronic liver disease/cirrhosis, diabetes mellitus
9. (a) Alzheimer's disease and (b) pneumonia/influenza
10. (a) Atherosclerosis and (b) cerebrovascular diseases, diabetes mellitus. hypertension, nephritis/nephrosis
11. (a) Hypertension and (b) diabetes mellitus, cerebrovascular diseases, aortic aneurysm, atherosclerosis
12. (a) Aortic aneurysm and (b) hypertension

Stallard also found numerous three-way associations of diseases listed on death certificates, such as:

1. diseases of the heart, diabetes mellitus, and hypertension;
2. cerebrovascular diseases, diabetes, and hypertension;
3. cerebrovascular diseases, diabetes mellitus, and atherosclerosis;
4. cerebrovascular diseases, atherosclerosis, and hypertension;
5. diabetes mellitus, nephritis/nephrosis, and septicemia;
6. diabetes mellitus, atherosclerosis, and hypertension; and
7. nephritis/nephrosis, chronic liver disease/cirrhosis, and septicemia.

These associations are only illustrative. There are many others that are not as strong, and often the association was with the residual group of causes other than the leading 15 causes. The listing of these groupings of causes underlines the difficulty of identifying the underlying cause and suggests the rationale for an alternative classification system that takes the groupings of causes directly into account.

Autopsies. Autopsies serve as another way to establish the combination of causes that occur together and the validity of the original clinical diagnosis. Determination of an underlying cause of death is particularly difficult for very old persons, especially elderly persons in nursing homes, because they tend to have multiple chronic conditions. It is not uncommon for physicians to have difficulty assigning the underlying cause of death to elders in the absence of autopsy findings sufficient to explain the death. Because of the low autopsy rate, especially among elderly persons in nursing homes, the cause of death is often incorrectly reported. U.S. NCHS (2001a) reported that in a review of 1,000 autopsies performed between 1983 and 1988, Sarode et al. (1993) found major discrepancies between the autopsy findings and the clinical diagnoses in 317 (32%) of the 1,000 autopsies. Two other studies found major discrepancies in the diagnosis of malignant tumors. Veress and Alafuzoff (1994) reported that 15% of all major cancers were not diagnosed, i.e., missed before autopsy. Manzini et al. (1995) reported that 34% percent of tumors with metastasis were missed before the autopsy (U.S. NCHS (2001a)).

A definitive cause of death may not be identified even after an autopsy is performed. However, conducting an autopsy helps to identify the more probable causes and to exclude the less likely ones. Sufficient evidence may not be available to indicate the cause of death with certainty, so that the pathologist may use nonspecific terminology for the cause or give the cause on the basis of a reasonable probability. In this way overreporting of "popular" illnesses that the autopsy does not support can be avoided.

With the new knowledge of how people die and the prospects offered by the genomic age, a classification of causes of deaths based on the traditional biomedical model may no longer be tenable. Some current theories of aging would now support a major category for cause of death akin to natural death, representing the endpoint of combinations of age-related chronic degenerative lethal diseases of later life. Later chapters, particularly Chaps. 13 and 14, will elucidate this presumably radical point of view.

Measures

Cause-specific rates are an important class of disaggregated death rates and figure importantly in analyses of public health and longevity trends. These rates can be structured in all the same forms as the rates not disaggregated by cause, but I discuss here only the central rates that are commonly the subject of mortality analyses.

Cause-specific death rates. The simplest mortality measure taking account of cause of death is the crude cause-specific death rate, defined as the number of deaths from a specific cause in a year divided by the midyear population. It is generally expressed per 100,00 population. Table 3.8 presents an array of cause-specific death rates for several countries. Note the predominance of chronic degenerative diseases in the more developed countries and the predominance of infectious and parasitic diseases

Table 3.8 Cause-specific death rates for selected countries: 2005 or 2006 (Deaths per 100,000 population)

	Country			
Cause of death	Mexico (2005)	United States (2005)	Japan (2006)	Russian Federation (2006)
Male				
All causes	531.3	827.2	944.3	1740.1
Tuberculosis	3.3	0.3	2.5	35.6
Septicemia	3.5	10.5	6.9	1.3
HIV	7.5	6.3	0.1	2.8
Malignant neoplasms	60.1	198.9	321.7	231.5
Ischemic heart disease	57.9	159.0	67.1	496.4
Cerebrovascular disease	25.1	38.8	99.7	250.3
Diabetes	60.2	25.0	11.8	4.6
Pneumonia	13.6	18.8	91.9	43.9
Chronic lower respiratory disease	23.6	42.5	20.8	39.1
Diseases of liver	43.6	15.9	17.7	45.9
Disorders of kidney and ureter	12.5	15.2	18.0	8.8[a]
Symptoms, signs, and abnormal findings not elsewhere classified	9.2	10.0	20.4	68.8
Accidents	51.6	51.5	37.9	NA
Transport accidents	25.2	23.0	10.2	42.2
Female				
All causes	420.4	824.6	778.8	1, 331.2
Tuberculosis	1.6	0.2	1.2	6.5
Septicemia	3.7	12.5	7.1	0.8
HIV	1.5	2.6	–	0.8
Malignant neoplasms	61.0	178.8	213.2	171.2
Ischemic heart disease	44.6	142.0	52.8	409.4
Cerebrovascular disease	27.5	57.8	103.6	352.7
Diabetes	68.8	25.7	9.9	9.5
Pneumonia	11.7	22.5	78.4	13.0
Chronic lower respiratory disease	19.3	45.5	9.2	15.3
Diseases of liver	15.0	9.0	8.3	27.8
Disorders of kidney and ureter	10.4	15.9	21.4	7.4[a]
Symptoms, signs, and abnormal findings not elsewhere classified	9.0	11.5	38.9	63.8
Accidents	15.6	27.2	23.1	NA
Transport accidents	6.6	9.4	4.3	13.6

Source: United Nations, *Demographic Yearbook 2008*, www.un.org
[NA]Not available, – rounds to zero
[a]Total diseases of the urinary system

in the less developed countries. The death rate for heart disease (I00-I09,I11,I13,I20-I51) in the United States in 2003 is calculated as follows:

$$\begin{aligned} DR_{hd} &= D_{hd} \div Pt^{*}100{,}00 \\ &= 685,089 \div 290,447,644 * 100.000 = 235.6 \end{aligned} \tag{3.17}$$

Source of basic data: U.S NCHS (2007) for death statistics; U.S. Census Bureau (www.census.gov) for population estimate.

Age-sex-cause-specific death rate. Cause-specific death rates can be made specific for age and sex to permit examination of the underlying variation of the rates. The death rate for heart disease for males at ages 55–64 in the United States in 2003 is:

$$
\begin{aligned}
{}_{55-64}^{M}DR_{hd} &= {}_{55-64}^{M}D_{hd} \div^{M} P^{*}_{55-64} 100{,}00 \\
&= 44,522 \div 13,390,503^{*} 100{,}00 = 331.7
\end{aligned}
\tag{3.18}
$$

Source: U.S. NCHS (2007) for death statistics; U.S. Census Bureau (www.census.gov) for population estimate.

Even for countries with inadequate data on deaths by cause, the distribution of age-specific death rates provides strong indications of the distribution of age-specific rates by cause. For example, extremely high death rates at the elderly ages may be attributed to high levels of cardiovascular diseases, cancer, and diabetes; high levels of the rates in young adulthood may be attributed to high levels of accidents and other acts of violence; and high levels of the rates in childhood suggests high levels of infectious illnesses.

Partition into broad categories. Deaths can be partitioned in many ways to serve different analytic uses. For example, commonly the causes of deaths affecting a particular organ, e.g., the heart or the kidneys, are grouped for study. The interest may be in a particular risk factor such as smoking or obesity, and all the diseases associated with or caused by this risk factor may be considered together. Alternatively, the interest may be in a process such as cancer, and the types of deaths attributable to cancer may then be considered in relation to the various systems of the body. Another way of grouping deaths according to process is to distinguish acute conditions, such as those due to an infectious or parasitic agent, and chronic conditions, those having a progressive/degenerative character, such as diabetes. An extension or variation of this approach is to partition deaths into two classes similar to acute and chronic conditions, the so-called exogenous (or extrinsic) causes and endogenous (or intrinsic) causes. This partition is used in the study of aging and longevity; it informs our understanding of the underlying age pattern of mortality, the causes of aging, and the process of senescence, and may be a useful vehicle for making projections of mortality. I consider this classification further below because of its fundamental importance in research on senescence (Carnes et al. 2006).
Endogenous (intrinsic) vs. exogeneous (extrinsic) diseases. Endogenous (intrinsic) causes have a biological character, arise from the genetic makeup of the individual, and are resistant to treatment. This class of diseases encompasses the so-called degenerative diseases of later life, such as heart disease, cancer, cerebrovascular disease, diabetes, Alzheimer's disease, chronic liver disease, renal disorders, and certain diseases peculiar to early infancy. The exogenous (extrinsic) class of diseases encompasses conditions that result essentially from influences external to the body,

such as violence (i.e., accidents, homicide, suicide), infectious and parasitic agents, and abuse of drugs, and other acute conditions (e.g., diseases of the upper respiratory system). For the most part, they are relatively avoidable, preventable, or treatable, have a short-term critical period, and end quickly in cure, permanent impairment, or death.

Two variations, one crude and the other refined, have been developed in keeping with this classification of causes of death into two broad categories. If the crude variation is applied in the selection of the extrinsic causes of deaths, they can be identified immediately from the list of ICD codes for the broad categories of causes suggested by the above illustrative causes. The intrinsic causes are identified as all deaths remaining after the extrinsic causes are eliminated from the total. The exact groups of causes to be included under the intrinsic classes is, therefore, somewhat arbitrary, and the line dividing the two classes varies with the medical knowledge of the times – both with respect to the etiology of various diseases and the possibilities for cure.

An alternative, more refined, selection of extrinsic causes of disease includes those specific causes among the broad categories of "intrinsic" causes of deaths that are externally caused. Carnes et al. (2006) prepared the revised list of extrinsic causes shown in Exhibit 3.3 on the basis of the Ninth Revision of the ICD. Some causes represent microbial assaults on body organs, such as cervical cancer and rheumatic heart disease, and others represent trauma to body organs resulting from a destructive lifestyle, such as lung cancer and cirrhosis of liver. Iatrogenic events (i.e., diseases or injuries caused by physician errors), nosocomial (i.e., hospital-caused infections), and starvation can be added to the list of extrinsic causes in Exhibit 3.3. These and similarly caused diseases can be assigned to the exogenous/extrinsic class rather than the endogenous/intrinsic class because external agents essentially cause the condition.

In the revised classification, the intrinsic causes of death are those remaining after the specified extrinsic causes are removed from the totals. Endogenous/intrinsic mortality is now a new composite of deaths from many causes. Some intrinsic deaths have purely genetic origins (i.e., are germ-line and preordained at conception); others involve gene/environment interactions (i.e., genetic susceptibility) in which genetic influences are dominant; and still others arise from mutations due to biological processes (e.g., "copy error mutation") that accumulate over the life span (Carnes and Olshansky 1997).

Medical developments can complicate or obfuscate the classification of some diseases. Consider the case of AIDS: what was once an acute lethal infectious disease has become a chronic manageable infectious disease because of the development of anti-retroviral drugs. It should remain in the extrinsic class of diseases, however.

From a biodemographic perspective, for the analysis of aging, the extrinsic causes of death are contaminants of all-cause death rates, and the extent of this contamination varies across age, time, space, and other variables (Carnes et al. 1996). Extrinsic deaths obscure the underlying intrinsic mortality level and pattern of all-cause death statistics. A biologically based mortality partition, such as that separating the intrinsic and extrinsic causes, provides a superior conceptual framework for the analysis of mortality and improves mortality comparisons between

ICD code	Cause of death
E800-E999	Injuries and poisoning
001–139	Infectious and parasitic diseases
162	Lung cancer
180	Cervical cancer
260–269	Nutritional deficiences
278	Obesity, adiposity, hyperalimentation
280–281	Iron and other deficiency anemias
283.1–283.2	Acquired hemolytic anemias
291	Alcoholic psychosis
303	Alcohol dependence syndrome
304	Drug dependence
320–322	Meningitis
390–398	Rheumatic fever and rheumatic heart disease
460–519	Diseases of the respiratory system
571.0–571.3	Chronic liver diseases and cirrhosis, alchohol-related deaths
571.4–571.7	Chronic hepatitis and biliary cirrhosis
590	Infections of the kidney
595	Cystitis
597	Urethritis, not sexually transmitted
598	Urethral stricture
599	Other disorders of the urethra and the urinary tract
601	Inflammatory diseases of the female genitourinary system
630–676	Complications of pregnancy, childbirth, and the puerperium
725	Polymyalgia rheumatica
730	Osteomyelitis, periostitis, and other infections involving bone

Exhibit 3.3 List of extrinsic causes of death, based on the ninth revision of the international list of causes Source: Reprinted with kind permission of Springer Science+Business Media B.V. from Carnes et al. (2006), Table 1

populations for different years and geographic units. The curve of intrinsic mortality between the ages of 15 and 80 conforms reasonably well to the Gompertzian model (Carnes et al. 2006). This partition is also important for the comparison of the mortality patterns of different species. Since the underlying disease processes are influenced by the same evolutionary mechanisms, different species share a common "signature" for all intrinsic causes of death combined (Carnes et al. 2006).

Intrinsic mortality patterns are persistent over time and space in the absence of interventions, but they are not immutable. Humans seek medical care (renal dialysis, heart bypass surgery, and so on) when they are ill, and these medical interventions are capable of "manufacturing significant amounts of survival time" that can and do "modify the intrinsic mortality curve" (Olshansky et al. 1998). For example, the overall level of mortality in the United States changed greatly between 1950 and 2000 and, because of medical and social interventions in this period, especially

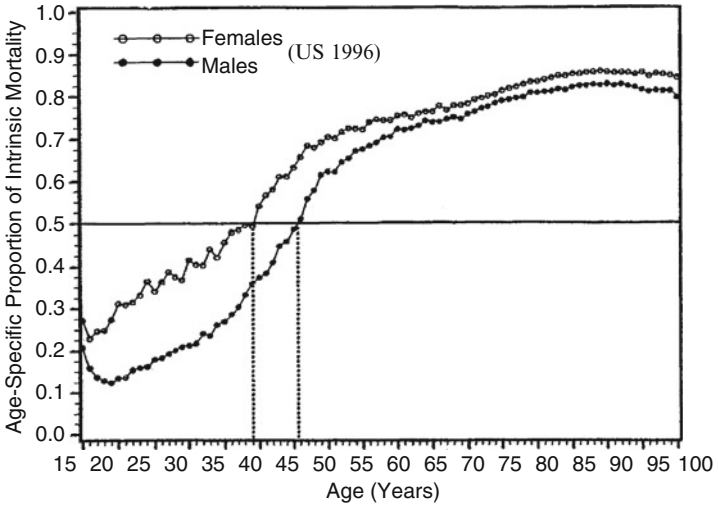


Fig. 3.4 Proportion of total mortality due to endogenous (intrinsic) mortality, for ages 15years and over, by age and sex, for the United States 1996 (Source: Reprinted with kind permission of Springer Science+Business Media B.V. from Carnes et al. (2006), Figure 1. Primary source: U.S. NCHS (1998))

after 1970, the death rates for several endogenous diseases fell and the level of the endogenous/intrinsic mortality curve dropped sharply. Figure 3.3a, b depict the age curves for 1950 and 2000 for both categories of causes in the United States. The two curves have different trajectories as well as different levels. There is a distinctive spike in the extrinsic mortality curve around ages 15–35 that is missing from the intrinsic curve. Until the later part of the reproductive period, extrinsic mortality constitutes a major portion of total mortality, and then intrinsic mortality tends to dominate. Figure 3.4 shows the curve of the percentage of intrinsic mortality in total mortality for the United States in 1996.

It is evident that total mortality and intrinsic mortality are age-dependent. Extrinsic mortality is also generally age-dependent, but its age-dependence has a very different and more limited character. Even in the case of a natural disaster or accidents affecting the general population, some age groups will be more vulnerable than others, typically children and the elderly. See, for example, Table 3.9, which shows the age curve of death rates from motor-vehicle accidents in the United States in 2002. Youth and the very elderly are the more vulnerable groups here.

Adjusted Summary Measures

The rates we have considered so far were described as observed rates. They directly describe the situation as represented by the recorded data for a population. For

Table 3.9 Death rates for motor vehicle fatalities in the United States: 2002

Age (years)	Death rates (per 100,000 population)
All ages	14.85
Under 5	3.11
5–9	3.06
10–15	5.28
16–20	30.88
21–24	27.52
25–34	17.24
35–44	15.14
45–54	14.16
55–64	13.24
65–74	15.43
75+	22.09

Source: National Highway Traffic Safety Administration, National Center for Statistics and Analysis *2002 Traffic Safety Facts*, Table 56
Official data have been adjusted for fatalities with age not reported

example, a crude death rate tells us that, say, 1% of the total population dies in a year, or an infant mortality rate tells us that, say, 1% of the births die in infancy in a year. It is often desirable, even necessary, however, to modify the observed measures to augment their analytic utility. The measures we discuss in this section are hypothetical constructs that have additional analytical utility over the observed measures. The various age-adjusted or age-standardized measures to be described are summary measures of mortality that are based on specified assumptions. Some of the adjusted measures are indexes for comparing the level of different mortality distributions.

Age-Adjusted/Age-Standardized Measures

The size of a crude death rate is affected by the demographic composition of the population for which the rate is calculated. Comparison of two or more crude rates is affected by differences in the populations being compared with respect to such characteristics as age, sex, urban-rural residence, race composition, and so on. For purposes of comparison of death rates over time or from area to area, it is useful to determine the levels of the rates on the assumption that there are no differences between the rates in the composition of the populations being compared, particularly their age compositions.

Inasmuch as age is the key compositional factor affecting the level of the crude death rate, adjustment, or standardization, of crude death rates is usually carried out only with respect to age. Other variables for which death rates may be adjusted or standardized are, as suggested, sex, race, nativity, and urban-rural residence, but

because death rates are often computed separately for these characteristics, it is usually necessary to standardize them only for age composition. Occasionally, crude death rates for the combined sexes are adjusted jointly for age and sex composition.

Age-adjusted, or age-standardized, death rates are then "crude" death rates of two or more populations that have been adjusted for their differences in age composition. They are summary measures of the level of mortality like crude rates, but they have been recomputed from the underlying age-specific death rates on the assumption that the age composition is the same for all populations whose crude rates are being compared. The adjustment of the original crude rates is designed to eliminate from them the effect of the differences in the age composition of the populations being compared. Crude death rates are unsatisfactory if one is trying to compare the health conditions of two populations. The crude death rate of a population may be relatively high because the population has a large proportion of persons at the older ages, where age-specific death rates are relatively high; or the crude death rate may be relatively low because the population has a large proportion of children and young adults, where death rates are relatively low. The crude death rate of a country may actually rise, even though death rates at each age remain stationary, if the population is getting older (that is, if an increasingly larger proportion of the population falls in the older ages).

Several methods have been developed for adjusting crude death rates for differences in age composition, deriving indexes of age-adjusted rates, and comparing mortality levels taking account of differences in age composition. I shall consider the two principle ones and their main variants: The direct method, including the variant called the comparative mortality index, and the indirect method, including the variant called the standardized mortality ratio. Next, I consider two other very different age-adjusted measures: The proportionate mortality ratio and the life table death rate. The measures differ in the kinds of basic data required in their computation, the designs for equating the age composition of the populations being compared, and the extent to which comparability is achieved among the adjusted rates in each comparison.

The direct and indirect methods each employ the same population distributions to weight the age-specific death rates of the pairs of populations being compared, as required, though the weights are different for the two methods. The indirect method and the proportionate mortality ratio directly yield the relative levels of mortality in the populations being compared; the other measures are actual death rates but they may be readily converted to ratios or indexes representing relative mortality levels. The proportionate mortality ratio allows one to determine whether the number of deaths from a particular cause for a particular exposed group is higher or lower than the number expected for the cause on the basis of the cause distribution for all persons, taking its age distribution into account. Unlike all the other measures, the life table death rate involves the use of different population distributions every time it is computed, but it has some distinctive advantages. These general statements should become clearer when the computational algorithms are described.

Like all summary measures, standardized rates mask differences found in the underlying age-specific death rates. Rates for particular population groups are also subject to the limitations of any rates based on small numbers, so that tests of significance should be applied when relevant and possible.

Direct Standardization

The simplest and most straightforward method of age-adjustment of death rates is the direct method. For most comparisons this is the preferred procedure and serves to provide the best basis for determining the relative difference between mortality in two or more areas or at two or more dates. In this method a "standard" population is employed in deriving the age-adjusted rates for two or more populations being compared. If the same standard population is employed, the resulting rates are called standardized and they are directly comparable in the sense that a common population age distribution was employed in their calculation. The formula calls for computing the weighted average of the age-specific death rates for a given area, using as weights the age distribution of the standard population. The formula for direct standardization is:

$$\mathrm{m_1} = \sum \frac{\mathrm{m_a P_a}^*}{\mathrm{P}} 100 \quad \text{or} \quad \sum \mathrm{m_a} \frac{\mathrm{P_a}^*}{\mathrm{P}} 100 \tag{3.19}$$

where $m_a = d_a / p_a$, the age-specific death rate for a given area, P_a represents the standard population at each age, and P or $\sum P_a$ represents the total of the standard population. Each age-specific rate is multiplied, in effect, by the proportion of the standard population in each age group. (In standardizing a death rate for age and sex jointly, each age-sex specific death rate is multiplied by the proportion of the total standard population in that age-sex group.) The age-standardized death rate of the standard population is the same as its own crude death rate, since the age-specific death rates of that population would be weighted by its own population.

$$\mathrm{M} = \sum \frac{\mathrm{M_a P_a}}{\mathrm{P}} = \frac{\mathrm{D}}{\mathrm{P}} \tag{3.20}$$

where M_a represents the age-specific death rates of the standard population, D represents the total number of deaths in the standard population, and so D/P represents both the crude death rate and the standardized death rate of the standard population. The relative mortality of two populations is derived by dividing the age-standardized rate for the study population by the crude death rate of the standard population:

$$\frac{\mathrm{m_1}}{\mathrm{M}} = \frac{\sum \mathrm{m_a P_a}}{\mathrm{P}} \div \frac{\sum \mathrm{M_a P_a}}{\mathrm{P}} = \frac{\sum \mathrm{m_a P_a}}{\sum \mathrm{M_a P_a}} \tag{3.21}$$

Note that, as shown in Eq. 3.21, in direct standardization the age-specific death rates for the area under study are compared with the death rates for the standard population, both being weighted by the age distribution of the standard population.

Table 3.10 Calculation of age-adjusted death rates by the direct method for several countries: Around 2006

	Standard population (in thousands) (P_a)	Age-specific death rates (m_a)		
Age (years)	United States, 2006	Japan, 2006	El Salvador, 2003	Georgia, 2006
Total	298, 755	8.5	4.4	9.6
Under 1	4, 179	0.7[a]	8.1	16.2
1–4	16, 273	–	0.7	0.3
5–9	19, 674	0.1	0.3	0.2
10–14	20, 587	0.1	0.4	0.2
15–19	22, 275	0.3	1.1	0.3
20–24	20, 994	0.4	1.8	0.8
25–29	20, 575	0.5	2.0	1.2
30–34	19, 608	0.6	2.1	1.7
35–39	21, 119	0.8	3.0	2.2
40–44	22, 436	1.3	3.6	3.2
45–49	22, 767	2.0	4.5	4.4
50–54	20, 459	3.1	5.9	6.8
55–59	18, 206	4.7	8.1	8.6
60–64	13, 350	7.1	12.8	13.0
65–69	10, 368	10.5	17.6	23.7
70–74	8, 541	17.3	25.5	36.3
75–79	7, 387	29.6	42.2	57.7
80–84	5, 670	49.4	132.9[b]	98.7
85+	5, 286	118.2		153.0
(1) Total standard population $= \sum P_a = P = 298, 755, 000$				
(2) Expected deaths $= \sum m_a P_a =$	2, 426, 254	1,770,920	3,057,687	3,200,185
(3) Age-adjusted death rate $= \sum m_a P_a \div P =$ (2) $\div$ (1) $=$	8.1	5.9	10.2	10.7
(4) Percent difference, age-adjusted rate from U.S. rate $=$	–	−27.0	+26.0	+31.9
(5) Crude death rate=	8.1	8.5	4.4	9.6
(6) Percent difference, crude rate from U.S. rate $=$	–	+4.9	−46	+19

Source: United States population, U.S. Census Bureau, www.census.gov
Other countries, United Nations, www.un.org
Standard population employed is the United States population in 2006. Rates are per 1,000 population. Calculations were carried out on the basis of rates per person
[a]Rate for under 5 years
[b]Rate for ages 80 and over

Illustrative calculations are shown for a few countries in Table 3.10. The estimated population of the United States in 2006 is employed as the standard population to calculate age-standardized death rates for Japan (2006), El Salvador (2003),

and Georgia (2006). The steps in calculating the age-adjusted death rate by the direct method are as follows:

1. Record the population in each age group for the United States (standard population).
2. Record the age-specific death rates for each country.
3. Compute the cumulative product of the population figures in step 1 and the death rates in step 2, to obtain the "expected deaths" for each country. For El Salvador the result is 3,057,687.
4. Divide the result in step 3 (3,057,687) by the total population of the United States (298,755,000), given in step 1. The result for El Salavador is 10.2 per 1,000. This is the adjusted death rate for El Salvador.

The crude death rate of El Salvador, 4.4 per 1,000 population, falls below the crude death rate of the United States, 8.1, by 3.7% points. The adjustment of the crude rate for El Salvador increased the rate to 10.2, reflecting the fact that the age composition of El Salvador's population is more favorable for a low crude death rate than that of the United States. Equating the age distributions implicit in the crude rates changes the difference from the United States to an excess of 2.1% points. Standardization caused a rise in the death rate in Georgia also. The crude rate gives the impression of lower mortality in El Salvador than in the United States, but the difference in the adjusted rates shows that, in fact, the force of mortality is higher in El Salvador. The adjustment has the opposite effect for Japan; its adjusted rate is lower than its crude rate, reflecting the fact that its age distribution has an unfavorable effect on its crude mortality level.

Choice of standard population. There are endless possible choices with respect to the standard to be selected in computing age-adjusted death rates by the direct method. The standard selected may be the age distribution of one of the areas or dates (e.g., earliest, middle, or latest in a series) being compared or the sum or average of the age distributions for the areas or dates being compared, or an "external" actual or theoretical distribution. Different results for the standardized measures and for the relative differences between standardized rates will be obtained, depending on the age distribution selected as a standard. In fact, the choice of standard may affect the direction of the difference between the rates for two populations being compared. The effect on the relative mortality level of a shift to the use of a different standard age distribution may be moderate or substantial. For example, the adjusted death rate of El Salvador, 10.2, exceeds the United States death rate by 26% when the United States is taken as the standard, but when the population of El Salvador is taken as the standard, its death rate, 4.4, exceeds that of the United States (3.4 as adjusted) by 29%. If the average of the age distributions for the United States and El Salvador is employed as a standard, the relative difference in the adjusted death rates is intermediate to these figures.

Hence, it is desirable to select the standard population carefully. The general rule is to select as a standard an age distribution that is similar to the various populations under study. If the mortality of two populations is being compared, this may best be

Table 3.11 Calculation of the age-adjusted death rate for diabetes mellitus for the United States: 2005 and 1950

Age	Standard population United States, 2000 (P_a)	Death rate, m_a 2005	Death rate, m_a 1950	Percent change in death rate, 1950–2000
All ages	274, 634	25.3	16.2	+56
Under 1	3, 795	–	0.7	−100
1–4	15, 192	–	0.3	−100
5–14	39, 977	0.1	0.6	−83
15–24	38, 077	0.5	1.1	−55
25–34	37, 233	1.5	2.2	−32
35–44	44, 659	4.7	4.2	+12
45–54	37, 030	13.4	12.4	+8
55–64	23, 962	37.2	42.4	−12
65–74	18, 136	86.8	100.3	−13
75–84	12, 315	177.2	166.7	+6
85+	4, 259	312.1	150.3	+108

(1) = Expected deaths = $\sum P_a m_a$	=	67,620 (2005)	63,298 (1950)	
(2) = Total standard population	=	274,634 (2005)	274,634 (1950)	
(3) = Age-adjusted death rate	=	24.6 (2005)	23.0 (1950)	+7
(4) = Crude death rate	=	25.3 (2005)	16.2 (1950)	+56

Source: NCHS 56(10), 2005; and Vital Statistics of the United States, 1950, vol.1, 1954
Population in thousands; death rates per 100,00. Calculations are carried out on a per person basis

achieved by using as a standard the average of the two populations. It is not always wise to follow this rule literally, however, since the populations being compared may have quite different age distributions. Use of the same population or the average of the populations in a time series over a long period is to be avoided since the adjusted rates begin to look very different from the observed crude rates. Yet, the farther apart the age distributions are, the more important it is to make the comparison of their mortality level on the basis of adjusted figures. In some cases it may be desirable to forego comparisons with summary measures and compare the schedules of age-specific death rates.

The need for age-adjustment is particularly important in connection with cause-specific death rates. Certain causes of death are concentrated in one or another part of the age distribution. Hence, the level of the crude death rate from these causes is particularly affected by the age distribution of the population. For example, a population with a large proportion of older persons will tend to have a relatively high death rate from heart disease and a population with a small proportion of older persons will tend to have a relatively low death rate from this cause. The calculation of an age-adjusted death rate for a specific cause by the direct method follows the same form as the calculation of a general age-adjusted death rate by the direct method except that age-cause specific rates are used for the area under study.

In the example in Table 3.11, the population of the United States in 2000 is employed as the standard for computing the age-adjusted death rate for diabetes

mellitus for the United States in 1950 and 2000. As before, the cumulative product of the U.S. population by age and the age-specific death rates for diabetes for 1950 and 2000 is obtained and divided by the total U.S. population. The direction of the adjustment reflects the difference between the age distributions at each date. The rate for 1950, which had a much younger population than now, is sharply increased and the rate for 2005, which has only a slightly older population than in 2000 is only slightly increased. The adjustment process used in this exercise follows exactly the current official practice of the National Center for Health Statistics.

Official age-adjusted rates and standards. The U.N. Statistical Office does not publish age-adjusted rates in the *Demographic Yearbook.* The U.S. National Center for Health Statistics (NCHS) has regularly published age-adjusted death rates for many population groups. They are currently computed by the direct method, employing the (projected) total population of the United States in the year 2000 as the standard. This was a new standard that NCHS introduced in 1999. The new standard replaces the earlier standard, the U.S. total population in 1940. The 1940 standard had been in use for over a half century so that, given the tremendous changes in the age structure of the U.S. population in this period, age-adjusted death rates and the observed death rates for any recent year are very far apart. The introduction of the new standard is intended not only to produce more realistic age-adjusted measures and comparisons, but also to reduce confusion among data users, who have had to deal with rates adjusted by alternative standard populations by other agencies.

The 2000 standard population is much older than the 1940 standard population: It has a higher mean age, a lower proportion of children, and a higher proportion of elderly persons. Inasmuch as such a population would give greater weight to the higher age-specific death rates of later life, use of the new standard produces adjusted rates that are much higher than the old standard. At the same time, rates adjusted by the 2000 standard population are much closer to the observed rates for current years than rates adjusted by the 1940 standard population. All comparisons are affected by the introduction of the new standard, but usually not by very much because the same standard is being applied in each comparison. For example, the percentage decrease in the age-adjusted death rate between 1979 and 1995 based on the year 2000 standard is only moderately smaller than the decrease based on the 1940 standard, mostly as a result of the fact that the base for calculating the percentage change is larger when the rate is adjusted by the 2000 population.[12] The trends in age-adjusted rates for most leading causes of death are nearly parallel with

[12] A fuller explanation of the basis for the introduction of the new standard population and of the effect of introducing it on U.S. mortality trends and sex-race differences is given in the following reports of the U.S. National Center for Health Statistics: Age standardization of death rates: Implementation of the year 2000 standard. In R.N. Anderson & H.M. Rosenberg, *National Vital Statistics Reports, 47(3)*, 1998; and Age-adjusted death rates: Trend data based on the year 2000 standard population. By D.L. Hoyert & R.N. Anderson, *National Vital Statistics Reports*, 49(9), 2001.

the two standards. The percent changes in the age-adjusted rates between 1979 and 1995 using the two standards for some of the leading causes are as follows:

	Standard population	
Cause of death[a]	1940	2000
All causes	−12.7	−9.2
Diseases of heart	−30.7	−26.2
Malignant neoplasms	−0.7	+3.8
Cerebrovascular diseases	−35.8	−34.3
Chronic obstructive pulmonary diseases	+42.8	+58.7
Accidents and adverse effects	−28.8	−24.8
Pneumonia and influenza	+15.6	+29.4
Diabetes mellitus	+36.1	+33.8

[a]Ninth Revision, International Classification of Diseases (ICD-9)

Comparative mortality index. The comparative mortality index (CMI) is a variant of the direct method of computing an age-adjusted death rate. It is an adjusted measure of relative mortality, usually employed for measuring changes over time in the overall level of mortality of an area. The CMI differs from the conventional form of the direct method in using a shifting pattern of population weights, designed to overcome the problems of using a single standard age distribution over a long period. As required, the formula still calls for the same weights in the numerator and denominator for any particular year:

$$\mathrm{CMI_y} = \frac{\sum \mathrm{w_a m_a}}{\sum \mathrm{w_a M_a}} \tag{3.22}$$

where M_a represents the age-specific death rates in the standard or initial year, m_a represents the age-specific death rates in later years, and

$$\mathrm{W_a} = 1/2\left(\frac{\mathrm{P_a}}{\mathrm{P}} + \frac{\mathrm{P_a}}{\mathrm{P}}\right) \tag{3.23}$$

where P_a and P are the age-specific and total populations, respectively, of the standard or initial year and p_a and p are the age-specific and total populations, respectively, of the various later years.

The formula calls for taking a ratio of (1) the weighted sum of age-specific death rates in each year to (2) the similarly weighted sum of age-specific death rates of the initial year. The weights are the average of (a) the proportion of the total population in the age group in the initial year and (b) and the corresponding proportion in each later year. The steps in computing the comparative mortality index for the United States in 2000 in relation to 1990 are as follows:

1. Compute the proportion of the population of the United States in each age group in 1990 and 2000 (cols. 1 and 2).

2. For the 2000 population weights (col. 3), average the proportions for 1990 (col. 1) and the proportions for 2000 (col. 2).
3. Set down the age-specific death rates for 1990 and 2000 in cols. 4 and 5.
4. Calculate the cumulative product of the population weights for 2000 (col. 3) and the age-specific death rates for 2000 (col. 5), to obtain the adjusted number of deaths in 2000. The result is the numerator of the computing formula for the CMI.
5. Derive the denominator of the computing formula as the cumulative product of the age-specific death rates for 1990 (col. 4) and the population weights for the year 2000 (col. 2).
6. Divide the result in step 4 by the result in step 5, to derive the CMI.

The comparative mortality index for 2000 reflects a moderately (?) lower level of mortality than in 1990. Comparative mortality indexes for other years in relation to the mortality level in 1990 may be computed in a similar fashion. Although the indexes directly relate mortality in a given year to the level in the starting year, the ratios of indexes in various later years may be used to determine relative mortality approximately among these later years. Because a different weighting pattern is used for each year, however, the CMIs for the various years are not fully comparable to one another. Nevertheless, the weighting pattern is so similar from year to year that the ratios of CMIs may be considered satisfactory measures of relative mortality over short periods.

Indirect Standardization

As we have seen, calculation of the age-adjusted death rate by the direct method requires age-specific death rates or deaths for age groups for the area under study. Such data are not always available (e.g., for statistically less developed countries and areas not defined as vital statistics tabulation areas). Comparison of the level of mortality independent of the effect of the age distribution is still possible in cases where only the total number of deaths and the population age distribution are known for the area of study because the number of deaths (d) for a population implicitly represents the product of the unknown age-specific death rates and the population in each age, that is, $\mathrm{d} = \sum \mathrm{m_a}\ \mathrm{p_a}$. An indirect method of age-adjustment of the death rate can be applied in this case. The formula for indirect standardization is:

$$\mathrm{m_2} = \left(\mathrm{d} \div \sum \mathrm{M_a p_a}\right) * \mathrm{M} \tag{3.24}$$

where M and M_a represent the crude death rate and the age-specific death rates, respectively, for the "standard" population, and d and p_a represent the total number of deaths and the population at each age, respectively, for the population under study. The formula calls for adjusting the crude death rate of the "standard" population by a factor representing the ratio of the total number of deaths in the study area to the number of deaths that would result if the study area had the death rates of the "standard" population. The use of this method permits a wide expansion of

Table 3.12 Calculation of age-adjusted death rates by the indirect method for selected countries: Around 2006)

Age-specific death rates (M_a)		Population (in thousands) (p_a)		
Age (years)	United States, 2006	Japan, 2006	El Salvador, 2003	Georgia, 2006
All ages	810.4	127, 756	6, 638	4, 398
Under 1	690.7	5, 530[a]	162	46
1–4	28.4	–	645	185
5–14	15.2	11, 954	1, 498	563
15–24	82.2	13, 821	1, 322	726
25–34	160.3	17, 778	1, 149	637
35–44	190.2	17, 125	697	626
45–54	427.5	16, 191	494	602
55–64	890.9	18, 977	332	376
65–74	2069.1	14, 329	220	403
75–84	5115.0	8, 991	120[b]	200
85+	13, 253.1	3, 060	–	34
(1) Expected deaths = $\sum M_a p_a$ =		1,483,334	23,045	32,281
(2) Registered deaths (d) =	2,426,264	1,084,451	29,377	42,255
(3) Ratio, registered deaths ÷ expected deaths = $d \div \sum M_a p_a$ = (2) ÷ (1) =	–	0.731	1.275	1.309
(4) Age-adjusted death rate = 810.4*(3) =	8.1	5.9	10.3	10.6
(5) Percent difference of adjusted rate from U.S. rate =	–	−27	+27	+31

Source: United States death rates: NCHS, www.CDC.gov/NCHS
Other countries population: United Nations, www.UN.org
Standard set of age-specific death rates are for the United States, 2006. Rates per 100,00 population. Calculations were carried out on the basis of deaths per person
[a]Under 5 years
[b]75 years and over

possibilities for computing adjusted death rates over space and time since many areas have figures on the total numbers of deaths and the population age distribution but do not have age distributions of deaths.

The steps in calculating the age-adjusted death rate by the indirect method, using the population of the United States in 2006 as the "standard," are as follows (Table 3.12):

1. Set down the age-specific death rates for the United States in 2006.
2. Set down the population by age for the countries being studied.
3. Compute the cumulative product of the death rates in step 1 and the populations in step 2. This is the number of deaths expected on the basis of the age-specific death rates of the United States in 2006 and the population in the countries under study.

4. Divide the result in step 3 into the total number of deaths registered in each of the countries. The result is the relative mortality in the areas under study and the mortality in the United States.
5. Multiply the result in step 4 by the crude death rate of the United States, to derive the adjusted death rate for each country.

The adjusted death rate for El Salvador calculated by the indirect method (10.3) exceeds the rate for the United States (8.1) by 2.2% points, that is, by the amount derived in step 4 (Table 3.12). The adjusted rates calculated by the direct and indirect methods (10.2 and 10.3) and the relative mortality of the populations being compared (26% and 27%) are very similar. The adjusted rates calculated by the direct and indirect methods for Japan and Georgia are also similar to one another (Tables 3.10 and 3.12). In general, the relative mortality of different populations shown by the indirect method, as compared with the direct method, reflecting the use of different population standards, may be moderately to substantially different and may even show opposite signs if the relative differences are small.

The direct and indirect methods each employ the same population distributions to weight the age-specific death rates of the two populations being compared, as required, though the weights are different for the two methods. We should expect the relative mortality of two areas measured by the direct and indirect formulas to be different. Although the data for the two areas are, in effect, weighted by the same populations in each method, different weights are used in the direct and indirect methods. The standard population is used as weights in the direct method and the population under study is used as weights in the indirect method. We may demonstrate this fact by converting the formulas for the two methods into a common form. The indirect formula may be rewritten as:

$$\mathrm{m_2} = \frac{\mathrm{d}}{(\sum \mathrm{M_a p_a})}{}^{*}\mathrm{M} = \frac{\sum \mathrm{m_a p_a}}{\sum \mathrm{M_a p_a}}{}^{*}\mathrm{M} \tag{3.25}$$

by substituting $\sum m_a p_a$ for its equivalent d. The direct formula may also be put into this form,

$$\mathrm{m_1} = \frac{(\sum \mathrm{m_a P_a})}{\mathrm{P}} = \frac{(\sum \mathrm{m_a P_a})}{\mathrm{P}} * \frac{(\sum \mathrm{m_a P_a})}{(\sum \mathrm{M_a P_a})} * \mathrm{M} \tag{3.26}$$

where $\sum M_a P_a$ replaces its equivalent $P^* M$.

A comparison of the two formulas, (3.25) and (3.26), on the far right, shows that the two sets of age-specific death rates are weighted by different populations, the standard population ($\mathrm{P_a}$) in the direct method and by the population under study ($\mathrm{p_a}$) in the indirect method. A corollary of this difference in the formulas is that a series of age-adjusted rates for several areas standardized by the indirect method would not be comparable from area to area since the implicit set of population weights

varies for each age-adjusted rate. In our example, we may consider the rate for ... and .... comparable to the rate for the United States, but not comparable to one another. This limitation does not apply to a set of age-adjusted rates derived by the direct method, all of which employ the same age distribution as weights. These considerations further support the choice of the direct method when the required data for employing it are available.

Standardized mortality ratio. We turn next to a variant of the indirect method of age adjustment that is often used in applications to socioeconomic data or extensive small area comparisons. It is called the standardized mortality ratio (SMR). The SMR is simply the factor in the formula for the indirect method that reflects the comparative mortality of two populations under the assumption of a common population age distribution, that is,

$$\text{SMR} = \frac{\text{d}}{\sum \text{M}_\text{a}\text{p}_\text{a}} * 100 = \frac{\sum \text{m}_\text{a}\text{p}_\text{a}}{\sum \text{M}_\text{a}\text{p}_\text{a}} * 100 \tag{3.27}$$

In this form the formula has often been employed to measure not only differences in mortality between countries and dates but also between socioeconomic categories and small areas. The formula can be applied to all deaths or to a specific cause of death, such as diabetes or tuberculosis. For example, to measure differences in mortality due to COPD between occupations, we would compare the recorded number of deaths due to COPD in a particular occupation with the number of deaths due to COPD expected on the assumption that the age-specific death rates of all workers due to COPD applied to that occupation. The use of the indirect formula can also be a convenience. The indirect formula is easier to apply in comparing the mortality level of the counties in the United States than the direct formula since age-specific death rates for the counties would not be needed, only the total number of deaths.

The method of calculation is illustrated with data on deaths from cancer for Montgomery County, Maryland, in the year 2000 as compared with deaths from cancer in the United States.

(a) Record the deaths from cancer for Montgomery County = d = 1,299.
(b) Record the age-specific death rates from cancer for the United States, M_a (col. 1).
(c) Record the population age distribution for Montgomery County, p_a(col. 2).
(d) Multiply col. (1) by col. (2), obtaining the overall sum of the products. These are the deaths expected from cancer for Montgomery County = $\sum \text{M}_\text{a}\text{p}_\text{a} = 1,659$..
(e) Divide (a) by (d), to derive the standardized mortality ratio = $(\text{d} \div \sum \text{M}_\text{a}\text{p}_\text{a})*100 = 1{,}299 \div 1{,}659 = 0.7830*100 = 78.30$.

In this calculation the age-specific death rates from cancer for the United States (item b) are assumed to apply to the population of Montgomery County (item c) in order to derive the expected number of deaths from cancer for Montgomery County (item d). The standardized mortality ratio represents the ratio of the number of actual deaths from cancer for the county (a) to the expected number (d), expressed as an index (i.e., ratio per 100). As the above formulas show, the standardized

mortality ratio is equivalent to the relative mortality of an occupation group or a small geographic area computed by the method of indirect standardization.

Deaths from specified occupations can be grouped into deaths for socioeconomic classes for purposes of studying socioeconomic variations in mortality. The broad grouping of the data reduces both the effect of the lack of correspondence between the occupation reported on the census schedule and the occupation reported on the death certificate, and the effect of the discordance in time reference of the reporting of occupation (e.g., current occupation and usual occupation) in the two sources. Where occupation of spouse is reported on the death certificate, the occupation of the husband may be effectively employed as a basis for examining mortality differences according to socioeconomic class among women, who have tended to be less engaged in the paid workforce. Such studies cannot be conducted in the United States by use of vital statistics because occupation of spouse is not asked on the death certificate. Studies of socioeconomic variations in mortality currently are usually based on national longitudinal surveys instead of census data and vital statistics. In this way data on deaths can be related to a variety of socioeconomic indicators, including family income, educational level, household wealth, and occupation.

A very useful indirect method of studying the association of socioeconomic status and mortality does not depend on the availability of data on the socioeconomic status of decedents and deaths distributed by socioeconomic status and age. It depends rather on the availability of statistics on the total number of deaths and the population age distributions for small geographic units that can be grouped into several economic strata. For example, comparisons of mortality according to socioeconomic class can be carried out for cities in the United States by grouping the deaths by census tracts and grouping census tracts into economic strata on the basis of a socioeconomic-status indicator, such as median monthly rent or median income, or density of occupancy of housing units. Standardized mortality ratios for these strata can be calculated in relation to the mortality level for the city. Alternatively, correlation studies of area death rates and measures of socioeconomic status may be employed to derive measures of the relationship between these variables. It is necessary to bear in mind the limitations of ecological correlation in such studies, however (see Appendix 3.1).

Summary Note on Age-Adjustment

We have considered several ways of calculating age-adjusted death rates or age-adjusted relative mortality. It is clear that there is no unique method of eliminating the effect of age composition when mortality levels of different populations are being compared. Although the direct method employs a common standard population for all the areas or dates being compared, the results are affected by the choice of standard population. The standard selected may be unrealistic for widely different populations or over long periods of time. In the indirect method of standardization

the comparability of the results is affected by the use of a different standard population for each area or date for which comparative measures are derived. Efforts have been made to find an ideal standard population for adjusting rates, particularly in the case of direct standardization. These efforts have been linked to studies aimed at decomposing the difference between the crude rates into their component factors, usually population age composition and the age-specific rates.

Decomposition of Difference Between Crude Death Rates

Having two crude death rates at different dates for some population, we seek to know how much of the total difference between them can be attributed to the difference between the underlying age-specific death rates and how much of the difference can be attributed to the difference in the age composition of the populations. Conditions of the ideal method for decomposing the total difference between the two crude rates are that the two parts should be independently determined and should add exactly to the total difference between the crude rates.

Our task in seeking to decompose the total difference between the crude death rates into the compositional effect and the rate effect is complicated by the fact that there is an interaction between the amount of mortality change and the amount of change due to the age composition of the population. The interaction results from the fact that the difference due to mortality depends on the difference in population composition. Decomposition of the total difference between two crude rates based on the results of the direct method using the initial population as standard, or, alternatively, the terminal population as standard, will not estimate the contribution of each factor accurately because it does not take account of the interaction of the components.

We can include the interaction effect in the main effects in two ways, one being simpler and less accurate than the other. The simple way to avoid a separate interaction term is to calculate the rate effect by direct standardization as described earlier and derive the compositional effect plus the interaction effect by subtraction of the rate effect from the total difference in the two crude rates. For example, the difference between the crude death rates for the United States in 2000 and 2005 can be disaggregated as follows:

(1) Change in crude death rate $= 825.9 - 854.0^{a} = -.0281$
(2) Change in ageadjusted death rate (rate effect) $= 798.8 - 869.0^{a} = -.0702$
(3) $=$ (1) $-$ (2) $=$ Composition effect plus interaction effect $= +.0421$

Source: Based on NCHS (2008)

[a]Normally the crude rate and the adjusted rate for the standard population would be identical but standard age distribution (projected population in 2000) and the actual age distribution in 2000 (2000 census) were slightly different

The difference between two crude death rates for population 1 and population 2 can be decomposed more exactly, without any interaction term as follows:

$$\mathrm{CDR}_1 - \mathrm{CDR}_2 = \sum\nolimits_a \frac{p_1 + p_2}{2}(m_1 - m_2) + \sum\nolimits_a \frac{m_1 + m_2}{2}(p_1 - p_2) \qquad (3.28)$$

where m_1 and m_2 refer to the two sets of age-specific death rates and p_1 and p_2 refer to the two sets of age–specific percentages of the total population at the two comparison dates. The rate effect is measured by the first term on the right and the compositional effect is measured by the second term on the right. Alternatively described, each adjusted death rate is calculated as the weighted average of the age-specific death rates of the given year, with the weights being the arithmetic means of populations 1 and 2 at each age in each case. The rate effect is then the difference of the two adjusted death rates and the composition effect is the difference between the total difference between the crude rates and the rate effect. It can easily be shown that the left-hand expression equals the right hand expressions. Simply replace CDR_1–CDR_2 by its equivalent, $\sum_a m_1 p_1 - \sum_a m_2 p_2$. Inasmuch as it is the difference between the standardized rates that is the basis of the contributions of the factors and the choice of standard population affects the results of both standardization and decomposition, standardization and decomposition are inextricably linked.

The techniques used to decompose the difference between the two rates into two components can be extended to include any number of factors or more than two populations. The decomposition can also cover various functional relationships of the factors and rates from cross-classified data. When more than two factors are involved or more than two populations are being compared, the analyst can refer to the formulas provided by Das Gupta (1993). These formulas cover a wide variety of situations, including variations in the number of factors, the number of populations, and the functional form of the difference between the original rates. When more than two populations are involved, adjusted measures and the decomposition of their differences may be accomplished by pairwise comparisons or directly for all populations. Das Gupta cites illustrations of studies made by other analysts that decompose the difference between two crude death rates into three factors (i.e., age composition, race composition, and age-race specific mortality), two total fertility rates into five factors, two indexes of dissimilarity (see below) of occupational sex concentration into two factors (i.e., structural change and sex segregation), two crude birth rates into three factors (general fertility rate, percent of women of childbearing age, percent women in the population), expectation of life at birth into five age groups, and expectation of life into five causes of death at each age.

Issues in time series. Adjusted series may be carried over many years, as the series of adjusted death rates published by the U.S. National Center for Health Statistics demonstrates. As reported, that agency recently shifted its population standard from the U.S. population for 1940 to the U.S. population for 2000. This radically modified the level of the adjusted series, but had much less effect on the relative changes in mortality over time. The choice of the population in year 1 as standard, as

applied by NCHS, may produce a useful time series of standardized death rates, but the decomposition of the changes between them would not be consistent with the change in the crude rates, and over a long period of time the series of adjusted rates tended to diverge greatly from the adjusted rate in the initial year and from the crude rates for each year in the series.

A limited solution to this problem is offered by the comparative mortality index, which uses a shifting population standard. Das Gupta (1993) offered a general solution where the final standardized number is a composite of the standardized rates based on all possible pairwise comparisons of the given populations. With this formula, comparison of the sum of the effects for any two standardized rates in the series would agree with the change in the crude rates.

Other Measures of Age-Adjusted Mortality

I turn now to a discussion of two other measures of age-adjusted mortality that have quite different properties than the age-adjusted death rates we described above. They are the proportionate mortality ratio (PMR) and the life table death rate (LTDR).

Proportionate mortality ratio. The proportionate mortality ratio has applications similar to the standardized mortality ratio so that the basic types of data to which it is applied will seem familiar. A report of the National Institutes of Occupational Safety and Health (U.S. NIOSH/Burnett et al. 1997) showed PMRs for 325 occupation categories and 235 industry categories, for about 190 causes of death. The PMR indicates whether the ratio of the number of deaths from a certain cause in a certain exposed group (e.g., truck drivers and motor vehicle accidents) is higher or lower than the number of deaths expected on the basis of the proportion of deaths due to that cause (i.e., motor vehicle accidents) for all persons (i.e., all occupations). To calculate age-adjusted proportionate mortality ratios, the following contingency table must be set up first (illustrated for cause of death and occupation):

	Cause of death		
Occupation/cause	Cause X	Other causes	All causes
Occupation Y	A_i	B_i	N_{1i}
Other occupations	C_i	D_i	N_{2i}
All occupations	N_{i1}	N_{i2}	T_i

i = ith age group (e.g., 5-year age group)
A_i = Observed number of deaths for a specific occupation and cause, for the ith age group.

The formula is:

$$\text{PMR} = \frac{\sum A_i}{\sum E(A_i)} * 100 \tag{3.29}$$

where $E(A_i)$ = expected number of deaths for an occupation and cause of death for the ith age group. $E(A_i)$ is computed as,

$$E(A_i) = \frac{N_{i1}N_{1i}}{T_i} \tag{3.30}$$

A PMR greater than 100 indicates that the group under study has a higher mortality from a given cause than would be expected for all persons. For example, a PMR of 156 for white male truck drivers aged 20 years and over who died as a result of motor vehicle accidents means that the number of deaths from this cause for white male truck drivers was 56% higher than the number expected on the basis of the proportion of deaths from this cause for all occupations.

One advantage of the PMR is that it does not require the population data needed for measures such as the age-adjusted death rates described earlier. The PMR is not a death rate or a relative measure of death rates because only death statistics are used in its calculation. By the same token, it is not a measure of relative mortality risk. It has some disadvantages, however. The PMR can overstate the risk of mortality if the risk of death for all causes for an occupation is low and it can overstate the risk of mortality if the overall risk of death is high. Abnormally high mortality in any of the major causes of death can distort the PMR. Finally the PMR is affected by the size of the occupation-unknown category and its unknown distribution among the occupations.

Life Table Death Rate

The life table death rate, the last of the age-adjusted rates described here, is the most difficult one to derive. It requires the construction of a life table from the observed age-specific death rates for each population examined. The subject of life tables is covered in the next chapter, but in anticipation, I make certain observations here regarding them. A standard life table assumes a constant annual number of births and a constant set of age-specific death rates corresponding to the observed age-specific death rates for a given year and area. From these data and assumptions, a population is generated with an unchanging total size and age distribution called the life table stationary population. The life table stationary population has its own "crude" death rate, called the life table death rate. In effect, the life table death rate may be viewed as the result of weighting the age-specific death rates in the life table, corresponding to the observed age-specific death rates, by the life table stationary population.

The life table death rate avoids the problem of the arbitrary choice of a standard population since it is generated entirely from its own schedule of age-specific death rates. In effect, each set of age-specific life table death rates is weighted by a different population, derived from its own age-specific rates. Inasmuch as each set of observed age-specific death rates generates its own distinctive life table population,

life table death rates are not strictly comparable from life table to life table in the sense of reflecting a common weighting pattern for the age-specific rates. Hence, life table death rates will not vary in the same way as the corresponding observed death rates or age-adjusted death rates based on them. They will, however, maintain the rank order of the latter. Life table death rates tend to run higher than observed crude death rates because of the low birth rate and resulting high average age of the life table stationary population. The next chapter is concerned with life tables and describes in detail the basis of calculating the life table death rate.

Appendix 3.1 Standard Errors for Random Variability of Mortality Statistics

Even when mortality statistics are obtained for an entire national population, they are subject to random variability. Hence, it is appropriate and useful to calculate standard errors and confidence limits for the numbers of deaths and death rates. To do this, it is necessary to make an assumption about the underlying distribution of the statistic in question. Deaths are relatively infrequent occurrences and the probability of dying is relatively low, so that these statistics can be assumed to have a Poisson probability distribution. The Poisson distribution provides conservative variance estimates for mortality statistics when the probability of dying is relatively low. Standard errors (SE) for mortality data are presented below on this assumption.

1. Number of deaths: $SE(D) = \sqrt{D}$, where D denotes the number of deaths.
2. Crude death rate and age-specific death rates, assuming that the denominator, that is, the population (P), is a constant: $SE(R) = .\sqrt{D}/P^2 = R/\sqrt{D}$, where R denotes the death rate.
3. Relative standard error (RSE), or coefficient of variation, of the number of deaths: This measure is calculated by dividing the number of deaths into its standard error and multiplying by 100: $RSE(D) = 100^*SE(D)/D = 100^*\sqrt{1/D} \div D = 100^*.\theta 1/D$.
4. Relative standard error of crude and age-specific death rates:

$$RSE(R) = 100^*SE(R)/R = 100^*(R/\sqrt{D})/R = 100^*\sqrt{1/D}$$

   The formulas for the relative standard error of the number of deaths and of the death rate are the same.
5. Standard error of the age-adjusted death rate $R.' = [SE(R')] = 100^* \sum_i [w_i^2(R_i^2/D_i)] = 100^*SE(R')/R'$
   where R_i = age-specific rate for the ith age group, w_i = age-specific standard weight for the ith age group from the U.S. standard population such that $\sum_i w_i = 1.0$, and D_i = number of deaths for the ith age group.
6. Relative standard error for the adjusted death rate: $RSE\ (R') = 100SE(R')/R'$

7. Standard error of the infant mortality rate. Calculation of standard errors assumes random variability in both the numerator (D) and denominator (B):

$$\mathrm{SE(IMR)} = \frac{\sqrt{\mathrm{var(D)} + \mathrm{IMR}^{*}\mathrm{var(B)}}}{\sqrt{\mathrm{E(B)}^2}} = \frac{\mathrm{D}}{\sqrt{\mathrm{B}^2}} + \frac{\mathrm{D}^2}{\mathrm{B}^3}$$

   where B denotes the number of births, which is also assumed to be distributed according to the Poisson distribution, and E(B) is the expectation of B.
8. Standard error of the maternal mortality rate: The formula has the same form as the formula for the infant mortality rate, with the substitution of the maternal deaths for infant deaths.

$$\mathrm{SE(MMR)} = \frac{\mathrm{D}}{\sqrt{\mathrm{B}^2}} + \frac{\mathrm{D}^2}{\mathrm{B}^3}$$

9. When the population data are based on a sample (e.g., socioeconomic groups in the Current Population Survey), then the rates are subject to sampling error of the denominator as well as random variability in the numerator. The formulas above have to be modified. For crude and age-specific death rates, the standard error is calculated from:

$$\mathrm{SE(R)} = \mathrm{R}\frac{1}{\sqrt{\mathrm{D}}} + 067\left(\mathrm{a} + \frac{\mathrm{b}}{\mathrm{P}}\right)$$

   where *a* and *b* represent parameters that vary depending on the subgroup of interest (e.g., high school graduates).
10. 95-percent confidence intervals for death rates may be calculated from:

$$\mathrm{L(R)} = \mathrm{R} - 1.96\ \mathrm{SE(R)}$$

$$\mathrm{U(R)} = \mathrm{R} + 1.96\ \mathrm{SE(R)}$$

    where L(R) and U(R) are the lower and upper limits of the confidence interval, respectively. These confidence limits can be interpreted to mean that the chances are 95 in 100 that the limits would contain the "true" death rate.

When the number of deaths is small (less than 100), the Poisson distribution cannot be approximated by the normal distribution. A simple method based on the more general family of gamma distributions, of which the Poisson distribution is a member, can be used to approximate confidence intervals for deaths and death rates when the number of deaths is small. Calculations using the gamma method can be made using commonly available spreadsheet programs or statistical software (e.g., Excel, SAS) that include the inverse gamma function. In Excel the function "gammainv(probability, alpha, beta)" produces values associated with the inverse gamma function for a given probability between 0 and 1. Alpha and beta are parameters associated with the gamma distribution. In Excel the following formulas

can be used to calculate lower and upper 95% confidence limits for the number of deaths and crude and age-specific death rates:

$$\mathrm{L(D)} = \mathrm{Gammainv}\,(.025, \mathrm{D}, 1)$$

$$\mathrm{U(D)} = \mathrm{Gammainv}\,(.975, \mathrm{D} + 1, \mathrm{D})$$

Confidence limits for the death rate are then calculated by dividing L(D) and U(D) by the population at risk of dying (P).

Although the calculations are similar, confidence intervals based on small numbers for age-adjusted death rates, infant and maternal mortality rates, and rates that are subject to sampling variability in the denominator are somewhat more complicated.

For more information regarding the measurement of random variation and its application to mortality statistics, see U.S. National Center for Health Statistics (1995), Technical Appendix. This material is also included on the CD-ROM that carries the same name. See also U.S. National Center for Health Statistics (2008).

References and Suggested Readings

Age Patterns of Mortality

Beard, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. In W. Brass (Ed.), *Biological aspects of demography* (pp. 57–68). New York: Barnes and Noble.

Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. *Demography*, 42(1), 23–49.

Carey, J. R. (2002). The importance of teaching biodemography in the demography curriculum. *Genus*, 58(3–4), 189–200.

Carnes, B. A., Olshansky, S. J., & Grahn, D. (1996). Continuing the search for a law of mortality. *Population and Development Review*, 22(2), 231–264.

Carnes, B. A., Holden, L. R., Olshansky, S. J., Witten, T. M., & Siegel, J. S. (2006). Mortality partitions and their relevance to research on senescence. *Biogerontology*, 7, 183–198.

Galley, C., & Woods, R. (1999). On the distribution of deaths during the first year of life. *Population 5, 1998. Population: An English Selection, 11, 35–60*. Paris: INED.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society of London, 115*, 513–593.

Heligman, L., & Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries, 10*, 49–80.

Horiuchi, S., & Wilmoth, J. R. (1998). Deceleration in the age pattern of mortality at older ages. *Demography, 35*(4), 391–412.

Makeham, W. M. (1865). On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries, 8*, 301–310.

Mode, C. J., & Busby, R. C. (1982) An eight parameter model of human mortality – The single decrement case. *Bulletin of Mathematical Biology, 44*, 647–659.

Olshansky, S. J., Carnes, B. A., & Grahn, B. (1998). Confronting the boundaries of human longevity. *American Scientist, 86*(1), 52–61.

Rogers, A., & Little, J. S. (1994). Parameterizing age patterns of demographic rates with the multiexponential model schedules. *Mathematical Population Studies, 4*, 175–195.
Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *Methods and materials of demography* (2nd ed.). San Diego, CA: Elsevier/Academic.
Suchindran, C. M. (2004). Model life tables and stable population tables. In J. S. Siegel & D. A. Swanson (Eds.), *Methods and materials of demography* (2nd ed., pp. 653–676). San Diego, CA: Elsevier/Academic.
Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society, 162*(Pt. 1), 5–43.
Thatcher, A. R., Kannisto, V., & Vaupel, J. W. (1998). *The force of mortality at ages 80 to 120. Odense*. Denmark: Odense University Press.
Thiele, P. N. (1872). On a mathematical formula to express the rate of mortality throughout the whole of life. *Journal of the Institute of Actuaries, 16*, 213–239.
U.S. National Center for Health Statistics. (2008). Deaths: Final data for 2005. By H. C. Kung, D. L. Hoyert, J. Xu, &S. L. Murphy. *National Vital Statistics Reports, 56*(10).
Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. C., et al. (1998). Biodemographic trajectories of longevity. *Science, 280*, 855–860.

Crossover of Rates and Heterogeneity

Coale, A. J. (1990). Defects in data on old-age mortality in the United States: New procedures for calculating mortality schedules and life tables at the highest ages. *Asian Pacific Population Forum, 4*(1), 1–31.
Coale, A. J., & Kisker, E. E. (1986). Mortality crossovers: Reality or bad data? *Population Studies, 40*, 389–401.
Kestenbaum, B. (1992). A description of the extreme aged population based on improved Medicare enrollment data. *Demography, 29*, 565–580.
Keyfitz, N., & Litman, G. (1979). Mortality in a heterogeneous population. *Population Studies, 33*, 333–334.
Manton, K. G., & Stallard, E. (1984, July 7–10). *Heterogeneity and its effect on mortality measurement*. In Proceedings, seminar on methodology and data collection in mortality studies, International Union for the Scientific Study of Population, Dakar, Senegal.
Manton, K. G., Poss, S. S., & Wing, S. (1979). The black-white crossover: Investigation from the perspective of the components of aging. *Gerontologist, 19*, 291–300.
Nam, C. (1995). Another look at mortality crossovers. *Social Biology, 42*(1–2):133–142.
Thornton, R. G., & Nam, C. B. (1968). The lower mortality rates of nonwhites at the older ages: An enigma in demographic analysis. *Research Reports in Social Science, 811*(1), 1–8.
U.S. National Center for Health Statistics. (1998). *Vital statistics of the United States, 1993: Mortality* (Vol. A). Hyattsville, MD: U.S. National Center for Health Statistics.
U.S. National Center for Health Statistics. (2002). *Vital Statistics of the United States: Mortality* (Vol. II, Parts A and B, 1993). Hyattsville, MD: U.S. National Center for Health Statistics.
Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16*, 439–454.

Rectangularization of Survival Curve and Mortality Compression

Lynch, S. M., & Brown, J. S. (2001). Reconsidering mortality compression and deceleration: An alternative model of mortality rates. *Demography, 38*(1), 79–96.

Myers, G. C., & Manton, K. G. (1984). Compression of mortality: Myth or reality? *The Gerontologist, 24*, 346–353.

Nusselder, W. J., & Mackenbach, J. P. (1996). Rectangularization of the survival curve in the Netherlands, 1950–1992. *Gerontologist, 36*(6), 773–782.

Rothenberg, R., Lentzner, H. R., & Parker, R. A. (1991). Population aging patterns: The expansion of mortality. *Journal of Gerontology: Social Sciences, 46* (2), S66–S80.

Wilmoth, J. R., & Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within populations. *Demography, 36*(4), 475–497.

Cause of Death

Carnes, B. A., & Olshansky, S. J. (1997). A biologically motivated partitioning of mortality. *Experimental Gerontology, 32*(6), 615–631.

Israel, R. A., Rosenberg, H. M., & Curtin, L. R. (1986). Analytical potential for multiple cause-of-death data. *American Journal of Epidemiology, 124*(2), 161–179.

Manzini, V. P., Revignas, M. G., & Brollo, A. (1995). Diagnoses of malignant tumor: Comparison between clinical and autopsy diagnoses. *Human Pathology, 26*, 280–283.

Mosley, W. H., & Chen, L. (1984). An analytical framework for the study of child survival in developing countries. *Population and Development Review*, (A supplement to volume 10), 25–45.

Sarode, V. R., Datta, B. N., Banerjee, A. K., et al. (1993). Autopsy findings and clinical diagnoses: a review of 1000 cases. *Human Pathology, 24*, 194–198.

Stallard, E (2002, January 17–18). *Underlying and multiple cause mortality at advanced ages: United States, 1980–1998*. In CD of proceedings of the society of actuaries, international symposium: living to 100 and beyond: Mortality at advanced ages, Lake Buena Vista, FL.

U.S. National Center for Health Statistics. (1982). Annotated bibliography of cause-of-death validation studies, 1958–80. By A. Gittlesohn & P. N. Royston. *Vital and Health Statistics, 2*(89).

U.S. National Center for Health Statistics. (1986). TRANSAX, the NCHS system for producing multiple cause-of-death statistics, 1968–78. By R. F. Chamblee & M. C. Evans. *Vital and Health Statistics, 1*(20).

U.S. National Center for Health Statistics. (1995). *Vital statistics of the United States: Mortality* (Vol. II, Parts A and B). Hyattsville, MD: U.S. National Center for Health Statistics, 1991.

U.S. National Center for Health Statistics. (2001a). The autopsy, medicine, and mortality statistics. By D. L. Hoyert. *Vital and Health Statistics, 3*(32).

U.S. National Center for Health Statistics. (2001b). Comparability of cause of death between ICD-9 and ICD-10: Preliminary estimates. By R. A. Anderson, A. M. Miniño, D. Hoyert, & H. Rosenberg. *National Vital Statistics Reports, 49*(2).

U.S. National Center for Health Statistics. (2003). *Public-use data set documentation. Mortality data set for ICD-10, 2000*. Hyattsville, MD: National Center for Health Statistics.

U.S. National Center for Health Statistics. (2007). Deaths: Leading causes for 2004. By M. P. Heron. *National Vital Statistics Reports, 56*(5).

U.S. National Center for Health Statistics. (2007). Autopsy patterns in 2003. By D. L. Hoyert, H. C. Kung, & J. Xu. *Vital and Health Statistics, 20*(32).

Veress, B., & Alafuzoff, I. (1994). A retrospective analysis of clinical diagnoses and autopsy findings in 3,042 cases during two different time periods. *Human Pathology, 25*, 140–145.
World Health Organization. (1977). *Manual of the international statistical classification of diseases, injuries, and causes of death, based on the recommendations of the ninth revision conference, 1975*. Geneva, Switzerland: World Health Organization.

Age-Standardization of Rates and Decomposition of Differences Between Rates

Das Gupta, P. (1993). *Standardization and decomposition of rates: A user's manual. Current Population Reports, P23–186. Special Studies*. Washington, DC: U.S. Bureau of the Census.
Horiuchi, H., Wilmoth, J., & Pletcher, S. D. (2008). A decomposition method based on a model of continuous change. *Demography, 45*(4), 785–802.
Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association, 50*(272), 1168–1194.
Liao, T. F. (1989). A flexible approach for the decomposition of rate differences. *Demography, 26*(4), 717–726.
Little, R. J. A., & Pullum, T. W. (1979). The general linear model and direct standardization: A comparison. *Sociological Methods and Research, 7*, 475–501.
U.S. National Center for Health Statistics. (2008). Deaths: Final data for 2005. By H. C. Kung, D. L. Hoyert, J. Xu, & S. L. Murphy. *National Vital Statistics Reports, 56*(10).
Vaupel, J. W., & Canudas Romo, V. (2002). Decomposing demographic change into direct vs. compositional components. *Demographic Research, 7*, 2–14.

Chapter 4
The Life Table

Definitions and Anatomy

Definitions and Types

An important extension of the tools of mortality analysis is the model called the life table. A life table is a statistical table describing the course of mortality and survivorship of a hypothetical birth cohort through the life cycle. The standard life table consists of six functions (i.e., columns of data), including age-specific mortality rates (or the conditional probabilities of dying at each age), survivors of an initial hypothetical cohort of births to each age, deaths of the cohort at each age, years lived in an age interval by the survivors, cumulative years lived in an age interval and later ages by the survivors, and life expectation at each age. The functions of the life table may be presented in single ages, constituting a complete or unabridged life table, or they may be presented in age groups, constituting an abridged life table (Fig. 4.1).

If the age-specific death rates used in constructing these tables are based on data for a single calendar year, the tables are described as period, or calendar-year, life tables. Alternatively, when the death rates for successive ages used in constructing the tables refer to successive calendar years, the tables are described as generation, or cohort, life tables. Such a table represents mortality events and the associated functions more realistically for a given birth cohort than the period table. Sometimes, only an extract of the table is shown, that is, some of the basic functions are omitted. Sometimes, the table is truncated at some age, that is, only a portion of the ages are displayed. The table may refer to a select population, such as the policyholders in a given large insurance company or the foreign-born segment of a national population. Many extensions of the basic life table have been developed, representing not only mortality histories over the life cycle but other health, demographic, and socioeconomic variables as well.

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_4,

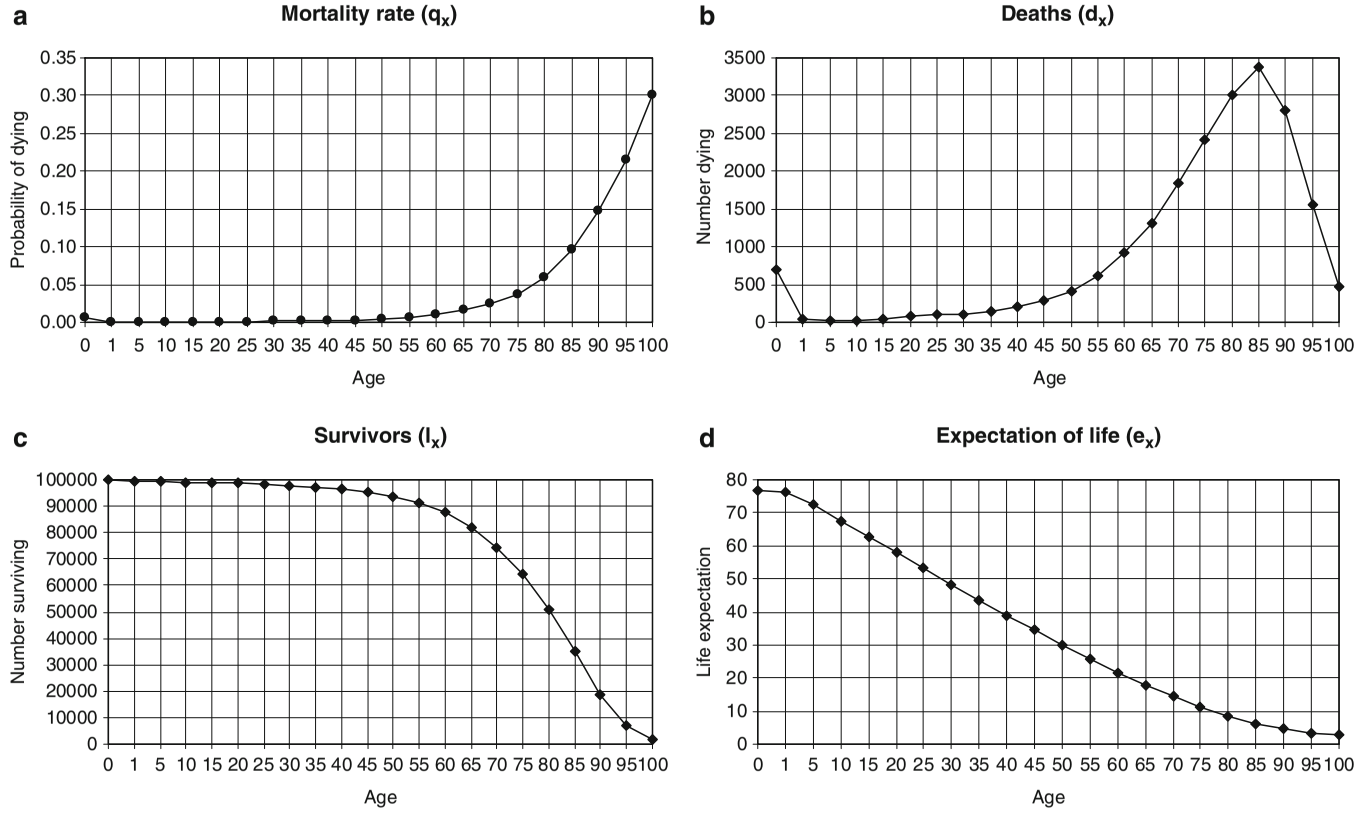


Fig. 4.1 Line graphs depicting four basic functions of the standard life table for the U.S. population in 2000 (Source: Based on Table 4.1 in this volume)

Functions of the Standard Unabridged Life Table

The Life Table as History of a Single Cohort

The anatomy of the life table may best be described in terms of a complete, or unabridged, period life table. For this purpose, let us consider the official life table for the total population of the United States in the year 2000 (Table 4.1). The initial function of the table is the age-specific mortality rate (col. 1). It is the proportion of the survivors to exact age x that die between exact age x and exact age $x + 1$ during a given year. Exact age refers to a specific point on the age scale; for example, exact age 55 means 55.0. Normally age x refers to completed years; for example, age 55 means the age span from 55.0 to 56.0 years. The mortality rate is conventionally symbolized by q_x. In the U.S. life table for 2000, q_{55} is 0.006654. The survival curve or function (col. 2) represents the number of survivors to successive ages (l_x) out of an initial hypothetical cohort of births (l_0), called the radix of the table; here the radix is 100,000. The number of survivors to age 55 in this table is 91,113. The third column in the standard life table is the number of survivors who die between exact age x and exact age $x + 1$ during the year, symbolized by d_x. In this table d_{55} is 606.

Considering the life table as the history of a single hypothetical cohort, the next column, labeled L_x, is the person-years of life lived by the survivors of the cohort as the cohort moves through age x. The person-years of life lived by the cohort at any age x must fall between the number of survivors to the initial age (l_x) and the number of survivors to the following age (l_{x+1}) – that is, for age 55, between 91,113 and 90,507 – because the deaths occurring in the interval x occur gradually through the interval. L_{55} is 90,810.

The next column, T_x, displays the sum of the L_x values from the bottom of the table up through age x. T_x equals $\sum_x^\infty L_x$ symbolically. It describes the cumulative person-years lived from age x forward by the cohort. It is essentially a computational column needed to derive the next and last function, life expectation. In our example, total person-years above exact age 55, or T_{55}, is 2,339,501. Life expectation at exact age x, or the average future lifetime for those surviving to exact age x, is the last of the functions conventionally displayed in a standard life table. This function is symbolized as e_x and is derived by dividing T_x by the survivors to age x, i.e., l_x, Life expectation, or life expectancy,[1] usually refers to the value at birth (e_0, or 76.9 here) although life expectancy is calculated for every age x shown in the table. Life expectancy at ages other than birth is also called average years of remaining life at age x. Here e_{55} is 25.7. Life expectancy should be not confused with life span – a distinction discussed in Chap. 13 – although both are measures of longevity, and the term life span is not an acceptable substitute for life expectancy.

[1]Life expectation, life expectancy, and average remaining lifetime all have the identical meanings. Sometimes the expressions, mean lifetime and the mathematically equivalent expression mean age at death, are used to refer to the same concept.

Table 4.1 Complete standard life table for the total population: United States, 2000

Exact ages, x to $x+n$	Probability of dying between ages x and $x+1$ q_x	Number surviving to age x l_x	Number dying between ages x and $x+1$ d_x	Person-years lived between ages x and $x+1$ L_x	Person-years lived above age x T_x	Expectation of life at age x e_x
0–1	0.006930	100,000	693	99,392	7,686,810	76.9
1–2	0.000517	99,307	51	99,281	7,587,418	76.4
2–3	0.000347	99,256	34	99,238	7,488,137	75.4
3–4	0.000243	99,221	24	99,209	7,388,898	74.5
4–5	0.000202	99,197	20	99,187	7,289,689	73.5
5–6	0.000189	99,177	19	99,168	7,190,502	72.5
6–7	0.000177	99,158	18	99,150	7,091,334	71.5
7–8	0.000167	99,141	17	99,132	6,992,185	70.5
8–9	0.000154	99,124	15	99,117	6,893,052	69.5
9–10	0.000137	99,109	14	99,102	6,793,936	68.6
10–11	0.000125	99,095	12	99,089	6,694,833	67.6
11–12	0.000130	99,083	13	99,077	6,595,744	66.6
12–13	0.000170	99,070	17	99,062	6,496,668	65.6
13–14	0.000253	99,053	25	99,041	6,397,606	64.6
14–15	0.000366	99,028	36	99,010	6,298,565	63.6
15–16	0.000491	98,992	49	98,968	6,199,555	62.6
16–17	0.000607	98,943	60	98,913	6,100,587	61.7
17–18	0.000706	98,883	70	98,848	6,001,674	60.7
18–19	0.000780	98,814	77	98,775	5,902,826	59.7
19–20	0.000833	98,736	82	98,695	5,804,051	58.8
20–21	0.000888	98,654	88	98,610	5,705,355	57.8
21–22	0.000945	98,567	93	98,520	5,606,745	56.9
22–23	0.000983	98,474	97	98,425	5,508,225	55.9
23–24	0.000996	98,377	98	98,328	5,409,800	55.0
24–25	0.000991	98,279	97	98,230	5,311,472	54.0
25–26	0.000981	98,181	96	98,133	5,213,242	53.1
26–27	0.000977	98,085	96	98,037	5,115,109	52.1
27–28	0.000979	97,989	96	97,941	5,017,072	51.2
28–29	0.000993	97,893	97	97,845	4,919,130	50.2
29–30	0.001019	97,796	100	97,746	4,821,286	49.3
30–31	0.001050	97,696	103	97,645	4,723,539	48.3
31–32	0.001087	97,594	106	97,541	4,625,894	47.4
32–33	0.001141	97,488	111	97,432	4,528,353	46.5
33–34	0.001215	97,376	118	97,317	4,430,921	45.5
34–35	0.001302	97,258	127	97,195	4,333,604	44.6
35–36	0.001395	97, 132	135	97,064	4,236,409	43.6
36–37	0.001492	96, 996	145	96,924	4,139,345	42.7
37–38	0.001602	96, 851	155	96,774	4,042,422	41.7
38–39	0.001728	96, 696	167	96,613	3,945,648	40.8
39–40	0.001870	96, 529	180	96,439	3,849,035	39.9
40–41	0.002021	96, 349	195	96,251	3,752,596	38.9

(continued)

Table 4.1 (continued)

Exact ages, x to $x+n$	Probability of dying between ages x and $x+1$ q_x	Number surviving to age x l_x	Number dying between ages x and $x+1$ d_x	Person-years lived between ages x and $x+1$ L_x	Person-years lived above age x T_x	Expectation of life at age x e_x
41–42	0.002181	96,154	210	96,049	3,656,345	38.0
42–43	0.002355	95,944	226	95,831	3,560,296	37.1
43–44	0.002550	95,718	244	95,596	3,464,465	36.2
44–45	0.002768	95,474	264	95,342	3,368,869	35.3
45–46	0.003014	95,210	287	95,066	3,273,527	34.4
46–47	0.003284	94,923	312	94,767	3,178,460	33.5
47–48	0.003567	94,611	337	94,443	3,083,693	32.6
48–49	0.003851	94,274	363	94,092	2,989,250	31.7
49–50	0.004138	93,911	389	93,717	2,895,158	30.8
50–51	0.004443	93,522	415	93,314	2,801,442	30.0
51–52	0.004780	93,107	445	92,884	2,708,127	29.1
52–53	0.005152	92,662	477	92,423	2,615,243	28.2
53–54	0.005579	92,184	514	91,927	2,522,820	27.4
54–55	0.006075	91,670	557	91,392	2,430,893	26.5
55–56	0.006654	91,113	606	90,810	2,339,501	25.7
56–57	0.007309	90,507	661	90,176	2,248,691	24.8
57–58	0.008023	89,845	721	89,485	2,158,515	24.0
58–59	0.008773	89,124	782	88,733	2,069,030	23.2
59–60	0.009563	88,343	845	87,920	1,980,297	22.4
60–61	0.010446	87,498	914	87,041	1,892,377	21.6
61–62	0.011448	86,584	991	86,088	1,805,336	20.9
62–63	0.012521	85,593	1,072	85,057	1,719,248	20.1
63–64	0.013646	84,521	1,153	83,944	1,634,191	19.3
64–65	0.014828	83,368	1,236	82,749	1,550,247	18.6
65–66	0.016058	82,131	1,319	81,472	1,467,498	17.9
66–67	0.017400	80,812	1,406	80,109	1,386,026	17.2
67–68	0.018933	79,406	1,503	78,655	1,305,916	16.4
68–69	0.020701	77,903	1,613	77,097	1,227,262	15.8
69–70	0.022663	76,290	1,729	75,426	1,150,165	15.1
70–71	0.024673	74,561	1,840	73,641	1,074,739	14.4
71–72	0.026741	72,722	1,945	71,749	1,001,098	13.8
72–73	0.029042	70,777	2,056	69,749	929,349	13.1
73–74	0.031663	68,721	2,176	67,633	859,600	12.5
74–75	0.034588	66,545	2,302	65,395	791,966	11.9
75–76	0.037675	64,244	2,420	63,034	726,571	11.3
76–77	0.040886	61,823	2,528	60,560	663,538	10.7
77–78	0.044437	59,296	2,635	57,978	602,978	10.2
78–79	0.048530	56,661	2,750	55,286	545,000	9.6
79–80	0.053313	53,911	2,874	52,474	489,714	9.1
80–81	0.058841	51,037	3,003	49,535	437,240	8.6
81–82	0.065093	48,034	3,127	46,471	387,705	8.1

(continued)

Table 4.1 (continued)

Exact ages, x to x+n	Probability of dying between ages x and $x+1$ q_x	Number surviving to age x l_x	Number dying between ages x and $x+1$ d_x	Person-years lived between ages x and $x+1$ L_x	Person-years lived above age x T_x	Expectation of life at age x e_x
82–83	0.072140	44,907	3,240	43,287	341,234	7.6
83–84	0.079850	41,668	3,327	40,004	297,947	7.2
84–85	0.088195	38,340	3,381	36,650	257,943	6.7
85–86	0.096751	34,959	3,382	33,268	221,293	6.3
86–87	0.105884	31,577	3,343	29,905	188,025	6.0
87–88	0.115605	28,233	3,264	26,601	158,121	5.6
88–89	0.125917	24,969	3,144	23,397	131,519	5.3
89–90	0.136824	21,825	2,986	20,332	108,122	5.0
90–91	0.148322	18,839	2,794	17,442	87,790	4.7
91–92	0.160404	16,045	2,574	14,758	70,348	4.4
92–93	0.173058	13,471	2,331	12,305	55,590	4.1
93–94	0.186266	11,140	2,075	10,102	43,284	3.9
94–95	0.200006	9,065	1,813	8,158	33,182	3.7
95–96	0.214248	7,252	1,554	6,475	25,024	3.5
96–97	0.228960	5,698	1,305	5,046	18,549	3.3
97–98	0.244099	4,394	1,072	3,857	13,503	3.1
98–99	0.259622	3,321	862	2,890	9,646	2.9
99–100	0.275475	2,459	677	2,120	6,756	2.7
100 years and over	1.00000	1,781	1,781	4,636	4,636	2.6

Source: NCHS, **NVSR** 51(3), Dec. 19, 2002.

The Life Table as a Stationary Population

The second interpretation of the life table sees it as the statistical description of a stationary or unchanging population. The life table stationary population (L_x) results from the assumptions that the mortality rates (q_x) used in constructing the table remain the same over an indefinite period and that birth cohorts (l_0) of the same size – e.g., 100,000 – are born each year over an indefinite period. These changes give rise to a population that does not change in size or age distribution, symbolized as L_x. The total has a growth rate of zero, as do each of the age groups. In the illustrative table the stationary population at age 55 (L_{55}) is 90,810 and the number of persons in the stationary population at ages 55 and over (T_{55}) is 2,339,501. The mortality rates and deaths at each age in a year are shown as q_x and d_x. The l_x column represents the number of persons having birthdays at each age in the year.

Some Characteristics of Life Table Functions

To aid the reader's understanding of life table functions, I note here some distinctive characteristics of the functions in the standard life table. Some functions in the life table relate to exact ages and others refer to completed ages or age groups: l_x and e_x are point measures and q_x, d_x, L_x, and T_x refer to age bands. The q_x function is based on the observed age-specific death rates and is independent of the other functions in the table and of the observed population distribution. All other functions of the table are dependent on the q_x function. The l_x function is not affected by the mortality rates at age x and all subsequent ages; it is affected only by the mortality rates prior to age x. The d_x and L_x functions are affected only by the mortality rates from the beginning of the life table through age x. The T_x function depends on all the mortality rates over the age scale since it combines all the L_x's after age x, the level of each of which is determined by all the mortality rates up through age x. Life expectation at any age is a function of the age-specific death rates for the designated age and all subsequent ages. The death rates for prior ages have no effect on life expectation. This seems anomalous because the value of the numerator of the life-expectation calculation, T_x, is dependent on all the mortality rates and the denominator, l_x, is dependent on the prior mortality rates. When the quotient is taken, $\mathrm{T_x} \div \mathrm{l_x}$, the effect of the prior mortality rates is "washed out," and only the effect of the subsequent mortality rates remains. Similarly the ratio of any two l_x's or L_x's is affected only by the mortality rates in the range of the numbers being related.

A comparison of life expectancies at a given age at two dates for the same populations in the context of falling death rates summarizes the mortality declines at all the ages above the designated age. When death rates decline over time, the life expectation that appeared at age x in an earlier table appears at a later age, age $x + t$, in a later table. For example, in the U.S. 1910 life table for females, life expectancy at birth was 53 years. By 1940 53 years was life expectation at about age 16. It appeared as life expectation at about age 23 in 1970 and age 27 in 2000. The age-delay in the appearance of given levels of mortality with the passage of time, occurring in the context of improving mortality, suggests a basis for making projections of mortality. The procedure is discussed further in Chap. 14.

Construction of the Standard Life Table

General Considerations

For the following description of the construction of a standard life table, I assume that reasonably accurate data on births, deaths, and population are available from the vital registration system and census records, at least for grouped data. I illustrate

the method of construction with the life table for the United States in 2000. (Model life tables and indirect techniques of life table construction, employed for the statistically less developed countries, are discussed in Chap. 12.) It is an important step in the preparation of a life table to review the basic data for irregular fluctuations, biases, and other errors. In the selection of the method employed in the construction of life tables, particularly unabridged tables, the analyst must decide on the degree of adherence to the observed mortality and population data that is considered desirable. Full adherence to the observed values for single ages leaves the mortality rates in the life table exhibiting all the fluctuations in the observed data, whether these represent real variations, errors in the data, or random fluctuations. Graduation (i.e., "smoothing") of the observed data could produce a series that better represents the underlying pattern of mortality than the observed data. Such graduation could, however, also eliminate or distort true variations in the observed data. The analyst must decide whether the emphasis should be on the series' adherence to the actual data or on its representation of the theoretical "true" mortality pattern.

As the quality of observed data has improved over time, fewer irregular fluctuations have appeared in them. As a result, there has been an increasing tendency to adhere to the observed data, and a decreasing tendency to modify them and eliminate or reduce any fluctuations. The consequence is that the 5-year totals for deaths and population, and hence the 5-year age-specific death rates, for the ages from about 5 to 84, are accepted without modification in constructing both unabridged and abridged life tables, but the single ages of deaths and population are modified by interpolating these 5-year data. Because of the thinness of the data and greater likelihood of reporting errors, however, the observed death rates at the highest ages – over age 84 – in the United States are modified, or more precisely replaced, for use in constructing official life tables.

Construction of the Complete Standard Life Table

The conventional or standard life table is constructed on the assumption that the age-specific death rates for a particular calendar year apply to each age for the life of the cohort. Under these conditions, the functions all apply, not to a real cohort, but to a synthetic cohort, that is, a hypothetical cohort living its entire life out in one calendar year rather than over a span of calendar years. Consequently, life expectancy at birth represents so-called period, or calendar year, life expectancy and is applicable solely to the calendar year to which the death rates refer.

As suggested, irregularities can appear in the reported deaths and population in single ages, particularly as a result of age misreporting, and these irregularities can affect the probabilities of dying. In order to eliminate or reduce any irregularities and obtain a smooth mortality curve, it is desirable to graduate the reported data. In the historical development of life table studies, different approaches have been taken

to this problem and different mathematical curves have been fit to the data. The leading approaches are summarized here in a few sentences; for more information, the reader should refer to the original life table publications.[2]

Graduation of Mortality Rates

One method of graduating the data employs the given 5-year totals for deaths and population up to about age 84 and interpolates these totals into single ages by an osculatory interpolation formula such as Sprague's fifth-difference osculatory formula or Beers' ordinary fifth-difference osculatory formula.[3] Death rates for single ages are then computed from these graduated deaths and population according to the formula:

$$m_x = D_x \div P_x \tag{4.1}$$

where D_x represents the number of deaths for the year under observation and P_x is the midyear population for that year. The death rates are then converted into mortality rates,

$$q_x = 2m_x \div (2 + m_x) \tag{4.2a}$$

or the mortality rates can be computed directly from the death and population data,

$$q_x = D_x \div (P_x + 1/2D_x) \tag{4.2b}$$

Other techniques are used for the ages under 5 and over age 80, 85, or 95.

---

[2]Some key publications are as follows:

U.S. Bureau of the Census, *United States Life Tables, 1890, 1901, 1910, and 1901–1910*, by J.S. Glover, Washington, DC: U.S. Government Printing Office, 1921. See the section on "Mathematical Theory of Construction of Life Tables."

U.S. Bureau of the Census, *United States Life Tables and Actuarial Tables, 1939–41*, by T.N.E. Greville, Washington, DC: U.S. Government Printing Office, 1946. See Part V, "Method of Construction and Graduation of the Life Tables."

U.S. National Center for Health Statistics, "Methodology of the National and State Life Tables," by R.J. Armstrong, *U.S. Decennial Life Tables for 1989–91*, Vol. 1, No. 2, Hyattsville, MD: U.S. National Center for Health Statistics 1999.

[3]Osculatory interpolation may be defined as a method of interpolation that insures smooth junction between the curves representing the interpolated values in adjacent tabular intervals by requiring that such adjacent curves have the same first derivative (or sometimes the same first and second derivatives) at the point of junction (Greville 1946). Osculatory interpolation formulas are polynomial interpolation formulas of various degrees modified to fit the requirement with respect to a common derivative mentioned. Ordinary osculatory interpolation formulas preserve the original grouped data and modified osculatory formulas modify the original grouped data. The Sprague, Beers, and Karup-King formulas may be designed as ordinary osculatory formulas or modified osculatory formulas. A newer form of interpolation that has not been used in constructing any of the official tables is cubic spline interpolation, a modified type of interpolation that promises even greater smoothness in the final series.

An alternative approach is to accept the 5-year totals of deaths and population, interpolate each series to single "pivotal" ages (e.g., age 7, 12, 17, 22, etc.) by an osculatory formula such as Sprague's, Beers', or Karup-King's, compute the mortality rates for these pivotal ages, and interpolate the rates for the other ages from the rates for the pivotal ages by an osculatory formula such as Sprague's or Beers' formula. A still different approach is to interpolate directly on the 5-year death rates as observed or some variation of them, bypassing the separate interpolation of deaths and population. For example, in its construction of the unabridged life table for 1994, the Office of the Chief Actuary, U.S. Social Security Administration (2005), applied Beers' fifth-difference osculatory formula to the logarithms of the complements of the 5-year probabilities of dying. In all these variations, other techniques are used for the ages under 5 and over age 80, 85, or 95. In the actual application of these interpolation methods, the results are obtained by employing a linear-compound version of the formulas in which sets of multipliers are cumulatively applied to the data.[4]

Calculation of Specific Functions

In the construction of the standard life table, two fundamental conversions of one function into another function requiring a mathematical assumption must be made. The first is the conversion of the single-year-of-age-specific central death rates (m_x) into the corresponding mortality rates, or probabilities of dying (q_x).[5] The central death rates in single ages are derived as described above.

As stated earlier, an age-specific mortality rate is defined as the proportion of a group of individuals that die between exact age x and exact age $x+1$ during a given year among the survivors to the initial exact age x. It expresses the risk of dying in an interval more precisely than a central death rate and, since it is in cohort form, it can be used directly to carry survivors forward from one age to the next. The basic formula for converting a central death rate into a mortality rate (above age 5) in a complete life table was given as formula (4.2a), $q_x = 2m_x \div (2 + m_x)$, where m_x is the central death rate at a given single year of age and q_x is the corresponding probability of dying. The formula is based on the assumption that the deaths between exact ages x and $x+1$ occur, on the average, at age $x + 1/2$, or that deaths at age x

[4]Descriptions of the interpolation formulas and sets of the corresponding interpolation multipliers are given in Shryock et al. (1980), Siegel and Swanson (2004), Appendix C, by D.H. Judson and C.L. Popoff, "Selected General Methods."

[5]In theory a life table should be constructed on the basis of hazard rates $[h(x)]$, or the force of mortality $[\mu(x)]$ – the instantaneous risk of dying at an exact age expressed in the form of an annual rate. In actual practice hazard rates are not used in life table construction since they cannot be determined exactly.

are rectangularly, or evenly, distributed by age and time within the age and calendar year. For example, to derive q_{55}, .008389, in the U.S. life table for 2000, we have,

$$q_{55} = 2(m_{55}) \div (2 + m_{55})$$

$$\text{Given } m_{55} \text{ as .008424}, \quad q_{55} = 2(.008424) \div (2 + .008424) = .008389$$

In general, calculation of the value of q_0 requires the use of distinctive "separation" factors because of the disproportionate concentration of infant deaths toward the beginning of infancy and the fluctuations in the number of births from 1 year to the next. (See the discussion on the infant mortality rate in Chap. 3.) In the event that the number of births has not changed, or changed little, from the previous year to the present year, the conventional infant mortality rate can be used for q_0. One formula for calculating q_0 that allows for the fluctuations in the number of births is as follows:

$$q_0 = (1 - f)D_{0y} \div B_y + f\,D_0 \div B_{y-1} \tag{4.3}$$

where D_{0y} is the number of infant deaths in year y, B_y is the number of births in year y, and the separation factor f represents the proportion of infant deaths in year y occurring to infants born in the previous year $(y - 1)$. In 2000 the separation factor for U.S. deaths approximated 0.12.

The next two functions, l_x and d_x, are obtained by alternating computations, beginning with $l_0 = 100{,}000$ and q_0. Their product is d_0, l_0 minus d_0 equals l_1, and l_1 times q_1 equals d_1:

$$d_0 = l_0 * q_0 \quad 693 = 100{,}000 * .006930 \tag{4.4a}$$

$$l_1 = l_0 - d_0 \quad 99{,}307 = 100{,}000 - 693 \tag{4.5a}$$

$$d_1 = l_1 * q_1 \quad 51 = 99{,}307 * .000317 \tag{4.4b}$$

$$l_2 = l_1 - d_1 \quad 99{,}256 = 99,307 - 51 \tag{4.5b}$$

and so on. In general, then,

$$d_x = l_x q_x \tag{4.6}$$

$$l_{x+1} = l_x - d_x \tag{4.7}$$

$$\text{or} \quad l_{x+1} = l_x(1 - q_x) \tag{4.8}$$

The derivation of L_x involves the second basic conversion in the life table, the conversion from the survivor function to the years-lived function. In a complete life table, the basic formula for this conversion is simply,

$$L_x = 1/2\,l_x + 1/2\,l_{x+1} \tag{4.9}$$

where the years-lived at any age x, 5 years and over, is derived as a simple arithmetic average of the survivors to exact age x and survivors to exact age $x+1$. Accordingly,

$$L_{55} = 1/2\, l_{55} + 1/2\, l_{56}$$

The formula is based on the assumption that the deaths at age x are rectangularly, or evenly, distributed by age and time between exact age x and exact age $x + 1$. For age $x = 0$, the separation factor f is used to calculate L_0.

$$L_0 = fl_0 + (1 - f)l_1 \tag{4.10}$$

For ages 1–4, the separation factors converge quickly from the infant separation factor to 0.5 at age 5 and, as a practical choice, given the small number of deaths at these ages, 0.5 can be used as a separation factor at these ages as well.

The next function of the life table, T_x, is the sum of the L_x values from L_x forward to the end of the table:

$$T_x = \sum_{t=0}^{t=\infty} L_{x+t} \tag{4.11}$$

$$T_{55} = 2{,}339{,}501 = 90{,}810 + 90{,}176 + 89{,}485 + ...2{,}120 + 4{,}636$$

The final function, e_x, or the average years of life remaining at age x, is derived by dividing the total person years to be lived at age x and above (T_x) by the number of survivors to age x (l_x).

$$e_x = T_x \div l_x \tag{4.12}$$

For example,

$$e_{55} = T_{55} \div l_{55}$$

$$e_{55} = 2{,}339{,}501 \div 91{,}113 = 25.7$$

Advanced Ages

Currently in the United States, because of suspected inaccuracies in death rates from the vital statistics system at the upper end of the age scale, the rates at these ages are derived by use of an alternate method with alternate data. For example, in the U.S. life table for 2000, for the ages 85–99, death rates from the Medicare health insurance system are indirectly employed in place of death rates derived from vital statistics and census population data. The q_x and l_x values are derived by special formulas from the Medicare mortality data. The Medicare data are judged to be more accurate than the vital statistics data at these ages. The Medicare death rates are not adopted directly, however. The age-to-age rate of change in the Medicare mortality rates are applied sequentially to the mortality rate at age 84 from the vital

Table 4.2 Rates of mortality change with age, for single ages from age 84 to Age 98, for males and females, based on medicare data: United States, 1997

Exact ages	Both sexes	Male	Female
84–85	0.092590	0.089728	0.103281
85–86	0.090210	0.087018	0.100251
86–87	0.087830	0.084308	0.097221
87–88	0.085450	0.081598	0.094191
88–89	0.083070	0.078888	0.091161
89–90	0.080690	0.076178	0.088131
90–91	0.073810	0.073468	0.085101
91–92	0.075930	0.070758	0.082071
92–93	0.073550	0.068048	0.079041
93–94	0.071170	0.065338	0.076011
94–95	0.068790	0.062628	0.072081
95–96	0.066410	0.059918	0.069951
96–97	0.064030	0.057208	0.066921
97–98	0.061650	0.054498	0.063891
98–99	0.059270	0.051788	0.060861

Source: U.S. National Center for Health Statistics (2002b); primary source, U.S. Centers for Medicare and Medicaid Services

statistics. The formula for deriving the age-to-age rate of change in the Medicare mortality rates is,

$$q_x = q_{x-1} * e^k \tag{4.13}$$

$$k = \text{In}(q_x) - \text{In}(q_{x-1}) \tag{4.14}$$

where k denotes the age-specific rate of mortality change with age. Values of k from Medicare data for 1997 are shown in Table 4.2. These figures show a clearly declining rate of increase in q_x with increasing age. Finally, ${}_\infty q_{100}$ is assigned a value of 1.0 since mortality is certain in this open-ended interval.

The National Center for Health Statistics derived the values of l_{100} and d_{100+} in the official U.S. table for 2000 as follows:

$$l_{100} = l_{99} - d_{99} = 2459 - 677 = 1{,}781$$

and since $l_{100} = d_{100+}$,

$$d_{100+} = 1{,}781$$

To obtain the value for e_{100}, we need values for L_{100+}, or T_{100}, and l_{100}. Individual values for L_x can be derived from the individual values for l_x. ${}_\infty L_{100}$ is calculated in essentially the same way as the other L_x's, but here the individual values of L_x, derived from individual l_x's and carried out to the age when the entire cohort is essentially extinguished, are summed.

Substitute death rates at the more advanced ages, say ages 85 and over, may be estimated in a variety of ways, such as by mathematical extrapolation, mathematical smoothing of the vital rates, or demographic analysis. One device used for securing population bases at the upper ages is the extinct-generation method; that is, the accumulation of the deaths at age x and higher ages occurring to the cohorts that constitute the extreme aged group, until the cohorts have been completely depleted. This method is also known as the method of population reconstruction through deaths. The method assumes no net immigration at these advanced ages. It depends on the quality of the death statistics for the highest ages, both their completeness and age accuracy, and is independent of any reported population data. The formula for estimating the population aged 85 years and over is:

$$ {}^{Y}P_{85+} = \sum_{85}^{\infty} D_c \tag{4.15} $$

where D_c refers to deaths for the cohort from age 85 to the highest age for which deaths are recorded. Additional deaths may have to be projected to wholly deplete the cohort. Then, to derive death rates for age groups:

$$ {}_5m_{85} = {}_5D_{85} \div {}_5P_{85},\ {}_5m_{90} = {}_5D_{90} \div {}_5P_{90},\ \text{and so on} $$

Since an abridged life table is generally sufficient for most purposes of demographic and epidemiological analysis and since it is readily possible, by interpolation, to expand an abridged life table into a complete life table, our further discussion of methods of constructing life tables will be limited to abridged life tables.[6]

Construction of an Abridged Life Table

Abridging the Complete Life Table

Several methods for constructing abridged life tables are described here. The methods differ somewhat in their basic assumptions and in their designs. Some of the methods "piggyback" on existing complete life tables. First, however, it is useful to describe how to convert a complete life table into an abridged life table (Table 4.3). For this purpose life table functions fall into two groups, those that are the same in the unabridged and abridged tables, l_x, T_x, and e_x, and those that are different, q_x, d_x, and L_x. In setting up the abridged table with 5-year age groups, the first group of functions can simply be transcribed from one table to the other for the ages, 0, 1, 5, 10, 15... 95, and 100 and over. The q_x, d_x, and L_x functions must be relabeled, however, as ${}_nq_x$, ${}_nd_x$, and ${}_nL_x$ (or ${}_5q_x$, ${}_5d_x$, and ${}_5L_x$) to represent age

[6] Additional discussion of methods for constructing complete life tables is given in Shryock et al. (1980).

Table 4.3 Abridged standard life table for the total population of the United States: 2000

Exact age	Probability of dying between ages x and $x+n$ ${}_nq_x$	Number surviving to age x l_x	Number dying between ages x and $x+n$ ${}_nd_x$	Persons-years lived between ages x and $x+n$ ${}_nL_x$	Total number of person-years lived after age x T_x	Expectation of life at age x e_x
0–1	.00693	100,000	693	99,392	7,686,810	76.9
1–5	.00131	99,307	130	396,916	7,587,418	76.4
5–10	.00082	99,177	82	495,668	7,190,502	72.5
10–15	.00104	99,075	103	495,278	6,694833	67.6
15–20	.00341	98,882	338	494,200	6,199,555	62.6
20–25	.00479	98,654	473	492,113	5,705,355	57.8
25–30	.00494	98,181	485	489,702	5,213,242	53.1
30–35	.00578	97,696	565	487,130	4,723,539	48.3
35–40	.00806	97,132	783	483,813	4,236,409	43.6
40–45	.01182	96,349	1,139	479,070	3,752,596	38.9
45–50	.01773	95,210	1,688	472,085	3,273,527	34.4
50–55	.02576	93,522	2,409	461,940	2,801,442	30.0
55–60	.03976	91,113	3,615	447,124	2,339,510	25.7
60–65	.06133	87,498	5,366	424,879	1,892,377	21.6
65–70	.09217	82,131	7,570	392,758	1,467,498	17.9
70–75	.13838	74,561	10,317	348,168	1,074,739	14.4
75–80	.20557	62,244	13,307	289,331	726,571	11.3
80–85	.31503	51,037	16,078	215,947	437,240	8.6
85–90	.46111	34,959	16,120	133,503	221,293	6.3
90–95	.61506	18,839	11,587	62,766	87,790	4.7
95–100	.75434	7,252	5,470	20,388	25,024	3.5
100 and over	1.0000	1,781	1,781	4,636	4,636	2.6

Source: NCHS, **NVSR** 51(3), Dec. 19, 2002 (2002b)

groups, and their single-age values must be combined to obtain the required group values. For example,

$$ {}_5d_{55} = d_{55} + d_{56} + d_{57} + d_{58} + d_{59} \tag{4.16} $$

$$ \text{or } 3{,}615 = 606 + 661 + 721 + 782 + 845; $$

$$ {}_5q_{55} = {}_5d_{55} \div l_{55} \tag{4.17} $$

$$ \text{or } 0.03968 = 3{,}615 \div 91113; \text{ and} $$

$$ {}_5L_{55} = L_{55} + L_{56} + L_{57} + L_{58} + L_{59} \tag{4.18} $$

$$ \text{or } 447{,}124 = 90{,}810 + 90{,}176 + 89{,}485 + 88{,}733 + 87{,}920 $$

At the terminal age group, e.g., "100 and over," all functions are the same in both types of tables.

Formal Methods of Construction

Several methods of constructing abridged life tables are described here, partly to familiarize the reader with the design of the various tables that may be found in demographic and actuarial publications and partly as a tool in deepening the reader's understanding of the life table's structure. The Reed-Merrell method (1939) was the first method among the several methods listed here to be proposed, then came the Greville method (1943), the Sirken method (1964), Chiang's method (1972), the Keyfitz and Frauenthal method (1975), and finally, the Schoen method (1978). I describe each of these methods briefly, show illustrative calculations for some of them, and then consider them comparatively.

Reed-Merrell method. In the Reed-Merrell method, the mortality rates are read off from a set of standard conversion tables showing the mortality rates associated with various observed central death rates. The standard tables for ${}_3m_2$, ${}_5m_x$, and ${}_{10}m_x$ have been prepared on the assumption that the following exponential equation applies:

$$ {}_nq_x = 1 - \exp(-n * {}_nm_x - an^3 * n\,{}_nm_x^2), \tag{4.19} $$

where n is the size of the age interval, ${}_nm_x$ is the central death rate, a is a constant, and e is the base of the system of natural logarithms. The conversion of ${}_nm_x$'s to ${}_nq_x$'s by use of the Reed-Merrell tables is usually applied to 5-year or 10-year data, but special age groups are employed at both ends of the life table (e.g., under 1 and 85 and over). For example, the death rate for the age group, ${}_5m_{55} = .008018$, in 2000 in the United States would be converted to the 5-year mortality rate (${}_nq_{55} = .03968$) in the U.S. life table for 2000 by using the Reed-Merrell table of values associating ${}_5q_{55}$ with ${}_5m_{55}$. Once the mortality rates have been obtained, the construction of the abridged life table continues with the computation of each entry in the survivor column, l_x, and the death column, ${}_nd_x$, along standard lines, using the formulas:

$$ {}_nd_x = {}_nq_x l_x $$

$$ l_{x+n} = l_x + {}_nd_x $$

$$ \text{or} \qquad l_{x+n} = (1 - {}_nq_x)l_x $$

All six short-cut methods that are described in this chapter follow the same general procedure in deriving l_x and ${}_nd_x$.

In the calculation of the next life table function, ${}_nL_x$, each of the six methods to be discussed follow a different procedure. In the Reed-Merrell method, T_x values are directly determined from the l_x's for ages 5 and over or 10 and over by use of a parabola passing through four l_x ordinates. An illustration of the computation of an abridged life table for males in rural India, 1957–1958, by the Reed-Merrill method is given in *The Methods and Materials of Demography*, Condensed Edition, published by Academic Press in 1975, Table 15.3.

Greville's method. A method suggested by T.N.E. Greville converts the observed central death rates to mortality rates by use of the formula

$$ {}_{n}q_{x} = \frac{{}_{n}m_{x}}{1/n + {}_{n}m_{x}[1/2 + n/12({}_{n}m_{x} - \ln c)]} \tag{4.20} $$

where c comes from an assumption that the ${}_{n}m_{x}$ values follow an exponential curve. Empirically, the value of c has been found to be between 1.08 and 1.10. Ln c could be assumed to about 0.095 as an intermediate value. Using this method, the observed death rate of .008018 (United States, 2000) would lead to a mortality rate of .03936 as follows:

$$ {}_{5}q_{55} = .008018 \div [0.2 + .008018[0.5 + 5/12(.008018 - 0.095)]] $$

$$ {}_{5}q_{55} = .008018 \div [0.2 + .008018(0.46378)] = .008018 \div .203719 $$

$$ {}_{5}q_{55} = .03936 $$

The derivation of ${}_{5}q_{x}$ by this method requires several columns in a manual calculation, but it may be programmed for direct calculation by computer on the basis of the ${}_{5}m_{x}$'s and the two constants n and ln c.

In the Greville method, the central death rates in the life table and in the actual population are assumed to be the same. Accordingly, the desired values of ${}_{n}L_{x}$ are calculated by the use of

$$ {}_{n}L_{x} = {}_{n}d_{x} \div {}_{n}m_{x} \tag{4.21} $$

For the last open-ended age interval, the approximation of ${}_{\infty}L_{x}$ is given by

$$ {}_{\infty}L_{x} = l_{x} \div {}_{\infty}m_{x} \tag{4.22} $$

An illustration of the application of the Greville method was presented for males in rural India, 1957–1958, in *The Methods and Materials of Demography*, Condensed Edition, published by Academic Press in 1975, Table 15.4. The Greville method was used in constructing the official abridged annual life tables of the United States for 1945–1952.

Sirken's method of reference to a standard table. Another method that is used frequently in routine calculations bases the conversion of the observed ${}_{n}m_{x}$ to the life table ${}_{n}q_{x}$ on the relation that exists in a complete life table between the observed ${}_{n}m_{x}$ and the life table ${}_{n}q_{x}$ (U.S. NCHS/Sirken 1966). Since this method obtains the new table by reference to a standard table, it should only be used when mortality in both tables is of a comparable level. The annual official U.S. life tables were constructed by the method of reference to a standard life table after 1953. Various decennial life tables from 1949–1951 to 1979–1981 were adopted as standard tables. The last official abridged life table using this method was published for 1996 and the standard table employed was that for 1979–1981 (U.S. NCHS 1998).

A simple application of this concept assumes that in each age interval the relation of ${}_nq_x$ to the observed ${}_nm_x$ shown by the standard table applies to the table under construction. First, the value ${}_ng_x$ is computed for the standard life table on the basis of the observed central death rates (${}_nm_x$) and the mortality rates (${}_nq_x$), using the formula,

$$ {}_{n}g_{x} = \frac{n}{{}_{n}q_{x}} - \frac{1}{{}_{n}m_{x}} \tag{4.23} $$

where ${}_ng_x$ represents the average number of years lived in the age interval by those dying in it. The index ${}_ng_x$ lies between 0 and n. This value is assumed to apply to each new life table repeatedly at the same ages. The required mortality rates in the new life table are computed using the formula,

$$ {}_{n}q_{x} = \frac{n * {}_{n}m_{x}}{1 + {}_{n}g_{x} * {}_{n}m_{x}} \tag{4.24} $$

where ${}_nm_x$ refers to the observed ${}_nm_x$ values applicable to the new table under construction. If the 1979–1981 life table is used as the standard, the value of ${}_5g_{55}$ would be equal to 2.369, and the observed death rate of .00852 for ages 55–59 in 1996 would be converted to a mortality rate of .04176 for 1996 as follows:

$$ \begin{aligned} {}_{5}q_{55} &= \frac{5(.00852)}{1 + 2.369(.00852)} \\ {}_{5}q_{55} &= .04260 \div 1.02018 \\ {}_{5}q_{55} &= .04176 \end{aligned} $$

For the calculation of ${}_nL_x$ in the abridged table, one may apply the simple relationship ${}_nL_x \div (l_x + l_{x+n})$ from the standard table. The annual official abridged U.S. tables made between 1954 and 1996 employ another relationship between l_x and ${}_nL_x$, involving a factor designated as ${}_nG_x$. The value of ${}_nG_x$, representing the distribution of deaths in the interval x to $x + n$, is obtained from the standard table from the formula

$$ {}_{n}G_{x} = \frac{n\,{}_{n}l_{x} - {}_{n}L_{x}}{{}_{n}d_{x}} \tag{4.25} $$

This value is assumed to apply to the new table, and it is used in the formula

$$ {}_{n}L_{x} = {}_{n}l_{x} - {}_{n}G_{x} * {}_{n}d_{x} \tag{4.26} $$

to obtain the desired value of ${}_nL_x$ in the new table. The value for the open-ended interval, ${}_\infty L_x$, is determined by a special formula. A factor r_x is computed for the standard table as follows:

$$ r_{x} = ({}_{\infty}m_{x} * {}_{\infty}L_{x}) \div l_{x} \tag{4.27} $$

where ${}_{\infty}m_x$ is the observed central death rate for the open-ended interval. This factor is applied in the new table by using the formula

$$ {}_{\infty}\mathrm{L_x} = \mathrm{l_x\ r_x} \div {}_{\infty}\mathrm{m_x} \tag{4.28} $$

The method of reference to a standard table is particularly useful and convenient in the construction of an annual series of life tables, since the ${}_{n}g_x$ and ${}_{n}G_x$ factors can be calculated once from a complete life table in the initial year and used over and over for each year until a new complete table is prepared. As mentioned, the U.S. abridged life tables for 1996 (U.S. NCHS 1998) represent an illustration of tables prepared by the Method of Reference to a Standard Table.

Chiang's method. Chiang's method (1968) employs the value ${}_{n}a_x$. The ${}_{n}a_x$'s represent the fraction of the interval between the xth and $(x + n)$th birthday lived by those dying within the interval. If deaths within an interval of 5 years following a birthday are evenly distributed in the interval, ${}_{5}a_x$ would be $1/2$. This is rarely the case and ${}_{n}a_x$ has to be determined by an examination of complete life tables. The formula for a 5-year group, with an illustrative calculation for ages 55–60 from the complete life table for 1989–1991 as standard is,

$$ \begin{aligned} {}_{5}\mathrm{a_x} &= ({}_{5}\mathrm{L_x} - 51_{\mathrm{x+5}}) \div 5 * (\mathrm{l_x} - \mathrm{l_{x+5}}) \\ {}_{5}\mathrm{a_{55}} &= (438{,}679 - 427{,}685) \div 5 * (4{,}121) \\ &= 0.5335 \end{aligned} \tag{4.29} $$

The formula for computing the chance of dying within n years after the xth birthday is,

$$ {}_{\mathrm{n}}\mathrm{q_x} = \frac{\mathrm{n} * {}_{\mathrm{n}}\mathrm{m_x}}{1 + \mathrm{n}(1 - {}_{\mathrm{n}}\mathrm{a_x})_{\mathrm{n}}\mathrm{m_x}} \tag{4.30} $$

The number of person-years lived in the interval x to $x + 5$ can be expressed by the following equation on the basis of ${}_{5}a_x$.

$$ {}_{5}\mathrm{L_x} = 5(\mathrm{l_{x+5}}) + {}_{5}\mathrm{a_x} * 5 * {}_{5}\mathrm{d_x} \tag{4.31} $$

The first term on the right indicates that each of the l_{x+5} persons who survive to the end of the interval live 5 person-years and the second term on the right indicates that each of the ${}_{5}\mathrm{d_x}$ persons who died in the interval live, on the average, ${}_{5}\mathrm{a_x}{}^{*}5$ person-years.

Chiang (1968) examined numerous complete life tables to determine the variability in ${}_{n}a_x$ and found that it was relatively invariant from table to table, especially for closely related populations. An age schedule of ${}_{n}a_x$'s can therefore be applied again and again in a series of life tables. The value of a_0 corresponds to the separation factor for infancy and hence varies with the level of the infant mortality rate, as noted in Chap. 3.

Given a set of ${}_na_x$ values and ${}_nm_x$ values, the corresponding values of ${}_nq_x$ can be calculated, on the basis of formula (4.27). For the United States in 2000, ${}_5m_{55} = .008018$, and for the 1989–1991 U.S. decennial life table, ${}_5a_{55} = .5335$; hence,

$$ {}_5q_{55} = \frac{5 * .008018}{1 + 5(1 - .5335).008018} = \frac{.04009}{1.01870} = .03935 $$

(The actual value for ${}_5q_{55}$ shown in the abridged U.S. life table for 2000 is .03968.)

Schoen's method. Schoen (1988) proposes a method, designated the "mean-duration-at-transfer" method, that estimates Chiang's a from the pattern of the observed death rates. In effect, a quadratic curve is fitted through $M(x - n, n)$. $M(x, n)$, and $M(x + n, n)$, that is, three adjacent observed death rates. As we may recall, ${}_na_x$ represents the fraction of the interval between the x and $(x + n)$th birthday lived by those dying within the interval and varies slightly from 0.5. According to Schoen, the relation between l_x and l_{x+n} can be expressed by the following equation (in the author's notation):

$$ \mathrm{l(x + n) = l(x)} \frac{\mathrm{l - u(x, n)M(x, n)}}{\mathrm{l + [n + w(x, n)]M(x, n)}} \tag{4.32} $$

and the number of person-years lived in the interval x to $x + n$ can be expressed by the following equation:

$$ \mathrm{L(x, n) = u(x, n)l(x) + [n + w(x, n)]l(x, n)} \tag{4.33} $$

Where

$$ \mathrm{u(x, n) = (n^2 \div 240)[M(x + n, n) + [n + w(x, n)]\, l(x + n)} \tag{4.34} $$

$$ \text{and } \mathrm{w(x, n) = n^2/240\,[14\,M(x + n, n) + 72\,M(x, n) - 6\,M(x - n, n)} \tag{4.35} $$

The expressions u(x,n) and w (x,n) are weighting factors in the definition of a (x,n):

$$ a(\mathrm{x, n}) = [u(\mathrm{x, n})\mathrm{l(x)} + w(\mathrm{x, n})l(\mathrm{x + n})] \div d(\mathrm{x, n}) \tag{4.36} $$

An illustration of an abridged life table computed by Schoen's mean-duration-at-transfer method is given in Schoen (1988:10).

Keyfitz-Frauenthal method. Keyfitz and Frauenthal (1975) proposed a formula that links survival rates and observed death rates as follows:

$$ \mathrm{l_{x+n} \div l_x = \exp\,[-n({}_nm_x + C)]} \tag{4.37} $$

$$ \text{Then,} \quad \mathrm{{}_nq_x = 1 - (l_{x+n} \div l_x) = 1 - \exp\,[n({}_nm_x + C)} \tag{4.38} $$

$$ \text{where} \quad \mathrm{C = ({}_nP_{x-n} - {}_nP_{x+n})({}_nm_{x+n} - {}_nm_{x-n}) \div 48\,{}_nP_x} \tag{4.39} $$

Hence,

$$ {}_{n}q_{x} = 1 - \exp\left[-n\,{}_{n}m_{x} + \frac{({}_{n}P_{x-n} - {}_{n}P_{x+n})({}_{n}m_{x+n} - {}_{n}m_{x-n})}{48\,{}_{n}P_{x}}\right] \tag{4.40} $$

where P is the observed population and ${}_{n}m_{x}$ is the observed death rate in the interval. For example, the observed ${}_{5}m_{55}$ (US 2000) would be converted into a mortality rate as follows:

$$ \begin{aligned} {}_{5}q_{55} &= 1 - \exp\left[-5(.008018) + \frac{(17{,}585{,}548 - 10{,}805{,}447)(.012579 - .005186)}{48(13{,}469{,}237)}\right] \\ &= 1 - \exp\left[-.04009 + \frac{(6{,}780{,}101)(.007393)}{646{,}523{,}376}\right] \\ &= 1 - \exp[-.04009 + .00008] \\ &=: 1 - \exp[-.04001] = 1 - .96078 = .03922 \end{aligned} $$

The required value of ${}_{n}L_{x}$ is calculated as,

$$ {}_{n}L_{x} = \frac{n(l_{x} - l_{x+n})}{\ln l_{x} - \ln l_{x+n}}[1 + n/24({}_{n}m_{x+n} - {}_{n}m_{x-n})] \tag{4.41} $$

The application of the Keyfitz-Frauenthal method is illustrated in Table 4.4, which shows the calculation of an abridged life table for urban Colombia in 1993. The steps for manual calculation are as follows:

1. Record the age interval (col. 1), the observed death rates (col. 2), and the observed population for these ages (col. 3).
2. Compute the differences between populations for alternate age groups and the differences between the death rates for alternate age groups, and take the product of the differences (col. 4).
3. Multiply 48 by the population in col. 3 and divide the result into col. 4 (col. 5).
4. Obtain the product of $-n$ and ${}_{n}m_{x}$, add the result to col. 5, and take the natural logarithm of the result (col. 6).
5. Derive q_x as the complement of the results in col. 6 (col. 7).
6. Derive ${}_{n}d_{x}$ and l_x in the usual fashion and show the results in cols. 8 and 9.
7. Calculate ${}_{n}L_{x}$ as shown in formula (4.41), employing two or three columns as convenient. For ages under 5 years, the multipliers from the Reed-Merrell calculations have been applied. For infancy, the equation is $L_0 = .276l_0 + .724l_1$; and for ages 1–5, the equation is, ${}_{4}L_{1} = .034l_{0} + 1.184l_{1} + 2.782l_{5}$. The separation factors used in these equations have been used in Table 4.4 for convenience, but 0.15 and 0.85 would have been more realistic choices for deriving L_0.
8. The functions, T_x, and e_x, are calculated in the usual way. Obtain ${}_{\infty}L_{x}$ by dividing l_x by ${}_{\infty}m_{x}$.

Table 4.4 Illustrative calculation of an abridged life table by the Keyfitz-Frauenthal method for males in urban Columbia, 1993

Age interval (exact ages, x to x+n)	${}_nm_x$	${}_nP_x$	$\frac{({}_nP_{x-n} - {}_nP_{x+n})}{({}_nm_{x+n} - {}_n m_{x-n})}$ $48\,{}_nP_x$	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	e_x
0–1	.0275	218,599		.025485	100,000	2,549	98,155	6,560,686	65.6
1–5	.0015	1,045,642		.005848	97,451	570	388,305	6,452,425	66.2
5–10	.0006	1,280,393		.002996	96,881	290	483,982	6,064,156	62.6
10–15	.0007	1,289,824	.0000102	.003545	96,591	342	482,410	5,580.531	57.8
15–20	.0037	1,077,126	.0000238	.018447	96,249	1,775	477,311	5,198,121	52.0
20–25	.0059	1,053,570	.0000017	.029077	94,474	2,747	465.653	4,620,810	48.9
25–30	.0056	1,032,450	−.0000012	.027606	91,727	2,532	452,209	4,155,157	45.3
30–35	.0052	967,247	−.0000006	.025662	89,195	2,289	440,219	3,702,948	41.5
35–40	.0055	769,837	.0000019	.027134	86,906	2,358	428,625	3,262,729	37.5
40–45	.0054	623,693	.0000126	.026700	84,548	2,257	417,176	2,834,104	33.5
45–50	.0067	546,546	.0000376	.033127	82,291	2,726	404,889	2,414,928	29.4
50–55	.0088	381,569	.0000675	.043369	79,565	3,451	389,718	2,012,039	25.3
55–60	.0139	284,877	.0001134	.067669	76,114	5,151	368,499	1,622,321	21.3
60–65	.0213	257,531	.0001748	.101810	70,963	7,225	337,824	1,253,822	17.7
65–70	.0338	176,277	.0004468	.157376	63,738	10,031	294,734	915,998	14.4
70–75	.0514	131,927	.0006518	.229148	53,707	12,307	238,616	621,264	11.6
75–80	.0781	83,108	.0012880	.327626	41,400	13,564	173,041	382,648	9.2
80–85	.1129	48,381	−.0027950	.423353	27,836	11,784	108,841	209,607	7.5
85+	.1593			1.00000	16,052	16,062	100,766	100,766	6.3

Source: Observed age-specific death rates from the United Nations *Demographic Yearbook, 1996*, Table 26; calculations by Kintner (2004). See the text for the formula for converting the central age-specific death rates, ${}_nm_x$, to the life table ${}_nq_x$

Comparison of Methods

If n were only 1 year in width, then the various methods of constructing life tables described would yield essentially the same results since they would use essentially the same assumptions. The methods would differ solely because of differences in the separation factors employed in the ages under 5. When the data are grouped in age intervals of 5 or 10 years, as in the case of abridged life tables, the various methods yield different results and the analyst must choose among the several methods. The methods differ from one another in the specific assumptions used to convert ${}_nm_x$ to ${}_nq_x$ and to obtain ${}_nL_x$ or T_x from l_x. Calculation of 5-year ${}_nq_x$'s from observed 5-year ${}_nm_x$'s by these methods yields very similar results (Table 4.5). It is not possible to state which of the various methods described is to be preferred, whether judged on the basis of greater accuracy, ease in application, or the logic of the assumptions.

In their assumptions the methods differ in the feature of the observed population that they choose to mimic. For example, in Greville's method the life table death rates are made to agree with the observed death rates. This is not the case in the

Table 4.5 Comparison of mortality rates computed by four abridged life table methods, for the total U. S. population, 1991

		Computed ${}_nq_x$			
Age interval (exact ages, x to x+n)	${}_nm_x$,[a] observed	Reed-Merrell	Greville	Reference to standard table	Keyfitz-Frauenthal
1–5	0.000474	0.001856	0.001894	0.001975	
5–10	0.000215	0.001074	0.001074	0.001099	
10–15	0.000258	0.001289	0.001289	0.001305	0.001293
15–20	0.000890	0.004441	0.004441	0.004416	0.004432
20–25	0.001101	0.005491	0.005491	0.005485	0.005483
25–30	0.001230	0.006133	0.006133	0.006118	0.006125
30–35	0.001541	0.007678	0.007678	0.007724	0.007676
35–40	0.001977	0.009840	0.009840	0.009879	0.009853
40–45	0.002536	0.012606	0.012606	0.012566	0.012664
45–50	0.003805	0.018859	0.018859	0.018884	0.019011
50–55	0.005758	0.028412	0.028411	0.028422	0.028554
55–60	0.009263	0.045341	0.045340	0.045392	0.045527
60–65	0.014319	0.069283	0.069282	0.069333	0.069135
65–70	0.021368	0.101741	0.101739	0.101846	0.101717
70–74	0.030251	0.148948	0.148946	0.149017	0.149153
75–80	0.048068	0.215454	0.215461	0.215286	0.216034
80–84	0.075754	0.319215	0.319274	0.318507	0.320953

Source: Calculations by Kintner (2004).

See the text for the formula for converting the central age-specific death rates, ${}_nm_x$, to the life table ${}_nq_x$. for each method; observed age- specific death rates are taken from the U.S. National Center for Health Statistics (1993)

[a]Ratio of deaths to population observed in the United States in 1991

other methods, and there is no generally agreed upon requirement that these two sets of death rates should agree. On the contrary, it could be maintained that, because the distribution of the life table population is different from the distribution of the actual population in each age interval, the two sets of rates should be different. The Keyfitz-Frauenthal method adopts the curvature of the observed population age distribution and of the observed age-specific death rates on which to base the curvature of the l_x function within age groups. Chiang, Schoen, and Sirken express their assumptions in terms of the curvature of l_x implicit in corresponding unabridged life tables.

Caution needs to be employed in the calculation of q_0 and ${}_4q_1$. Note that it is necessary to employ separation factors that are applicable to the population for which the abridged table is being constructed, not those given in the original proposal or in illustrative presentations of a method. Hence, the analyst must usually make an independent judgment regarding the factors to be chosen.[7]

Other Types of Basic Life Tables

Both the complete life table and the abridged life table I have discussed so far are period or calendar-year life tables. That is, they are based on the assumption that the schedule of age-specific death rates for a given calendar year (or short group of calendar years) applies to a single birth cohort throughout its lifetime. In fact, numerous birth cohorts are represented in such period tables. The two types of tables differ only in the age detail presented and in the ways by which the single ages were derived from the grouped data. As mentioned earlier, there are other varieties of life tables. One other type of life table, a generation, or cohort, life table, is based on a different combination of age-specific death rates than period tables. In addition, numerous extensions of the basic life table have been developed; these are discussed later in this chapter and in Chap. 8.

Selected Populations, Truncated Life Tables, and Life Table Extracts

Life tables may be constructed for populations having special characteristics. The best known of the life tables of this type are those describing the policyholders of large life insurance companies. For many years the Metropolitan Life Insurance Company published a life table for its industrial policyholders, those who paid their premiums weekly. These tables were in effect limited to those in a particular income class defined on the basis of the frequency of the payment of their insurance

[7]For further information about the construction of abridged life tables, refer to various editions of *The Methods and Materials of Demography*, particularly Vol. II of the unabridged version, published by the U.S. Census Bureau in 1977–1980, or the second edition of *The Methods and Materials of Demography* published by Elsevier/Academic Press in 2004.

premiums. A life table may exclude a certain class of persons. The table may cover the general population except for those with a particular health condition (e.g., diabetes, heart disease). Or it may refer exclusively to a group with a particular health condition (e.g., heart disease, diabetics). A table may cover everyone except those with a particular lifestyle or health risk such as smokers or obese persons. If deaths due to smoking or obesity are excluded, the table could tell us how many years of expected life are added by the elimination of such deaths.

From time to time only parts of a life table are published. If these parts are sections of the age cycle, the table is called a truncated life table. For example, the table may refer only to persons aged 50 years and over or under 10 years of age. More commonly, entries are shown only for selected functions, such as q_x, l_x, and e_x, and these functions may be presented for every age or, typically, for every 5th year of age (0, 1, 5, 10, 15, etc.). The United Nations follows this practice in presenting life table extracts for the countries of the world in the *Demographic Yearbook*. The reason for this practice is to preserve space while publishing the most widely used functions of the life table; the other functions are either easily derivable from those given ($_nd_x$) or are of secondary interest (L_x and T_x).[8]

Generation (or Cohort) Life Tables

Life expectancy at birth in the conventional or period life table is a hypothetical measure of life expectation reflecting mortality over the lifetime of a cohort in terms of mortality in a single calendar year. If the age-specific death rates relate to an actual cohort of births and the observed risks of death throughout the life span of this cohort (that is, if the death rates for successive ages relate to successive calendar years), the life table gives a more realistic measure of the life expectancy of a real cohort. A life table constructed with such a set of death rates is called a cohort, or generation, life table. The table begins with the infant mortality rate in the birth year of the cohort, then proceeds with the death rate in the next calendar year for children aged 1, then with the death rate in the following calendar year for children aged 2, and so on for about 100 years. For example, for the generation life table of 1905, the mortality rates would be selected as follows: q_0, 1905; q_1, 1906; q_5, 1910; . . .; q_{95}, 2000, $q_{100}+$, 2005+. Thus, to construct such a table requires a set of death rates for single ages for appropriate calendar years for over a century. The life-expectancy function of such a table gives the average future lifetime for a cohort of individuals born in a specified year for those surviving to each exact future age x. Generation life tables are conventionally identified in terms of the year of birth of the cohort.

Generation life tables can be used for a variety of purposes, among them the historical study of life expectancy, the projection of mortality, and the preparation

[8]Rounded estimates of T_x may be obtained as the product of e_x and l_x, and estimates of L_x may be obtained by taking the differences of the T_x's.

of estimates of orphanhood. A series of such tables could represent the transitions in mortality and life expectancy of real cohorts over time and thereby improve the basis for analyzing the changes in death rates over time.

In order to prepare a cohort or generation life table for a given country, a sufficient number of years must pass since mortality data have been systematically recorded (preferably somewhere over 80 years) in order to measure most of the observed lifetime mortality of the birth cohort of interest.[9] For the United States, enough years of officially recorded death statistics have elapsed to make possible the preparation of many historical generation life tables. The U.S. Actuary's Office (2005) has published a historical and projected series of generation life tables (for each sex) for the United States Social Security Area at decennial (1900–2100) intervals.[10] Inasmuch as mortality data for over a century are needed to complete a generation life table, death rates at the advanced ages for most of the historical tables had to be projected and all the death rates in the tables for 2010 and later had to be projected. Table 4.6 presents selected values from the generation life table for the cohort of females born in 1920 prepared by the U.S. Office of the Actuary. For this table the death rates for ages above 80 or so had to be projected (Fig. 4.2).

The expectation of life in a generation life table is higher than in the period life table for the starting year of the generation life table when secular trends in death rates are declining, as was the case in the twentieth century in the United States and the other more developed countries of the world. The reverse could be true if death rates increase. A comparison of life expectations at birth in two such tables reflects the relation between the initial mortality of a cohort and its later mortality experience in view of the general tendency for the mortality of an actual cohort to fall at each age over its lifetime. In the United States during the last century, period life tables typically understated the life expectancy at birth of an actual cohort of persons by several years. Table 4.7 presents a historical series of values for life expectancy at birth for the United States (Social Security Area) from 1900 to 2000 for the two types of life tables. A comparison of Figs. 4.3 and 4.4 shows the historical trend

[9] As stated, preparation of a generation life table requires compilation of data over more than a century on an annual basis and, usually, the projection of data for some years at the older ages for the cohort. The only practical way to prepare such a table and especially a series of such tables is by electronic computer. The basic input for an annual series of generation life tables, or even a decennial series, is the giant matrix of observed or projected central death rates for single ages and single calendar years for at least 110 years.

[10] U.S. Office of the Actuary, Social Security Administration, "Life tables for the United States Social Security Area: 1900–2100," by F.C. Bell and M.L. Miller, *Actuarial Study*, No. 120, 2005. The U.S. Social Security Area includes the outlying areas of the United States in addition to the 50 states and the District of Columbia. This report replaces an earlier report of the same title, *Actuarial Study* No. 112, 2002, of the same scope. Earlier, Jacobson (1964) prepared generation life tables for white females and white males born in the United States in 1840 and every tenth year thereafter through 1960, on the basis of the mortality rates experienced in each year in the life of these cohorts (through 1960) and projected mortality rates. L.I. Dublin, A.J. Lotka, and M. Spiegelman had described such tables in their classic work, *Length of Life*, rev. ed., New York, Ronald Press, 1949.

Table 4.6 Selected values from the SSA generation life table for the cohort of females born in the United States in 1920

Calendar year	Age (x)	Mortality rate (q_x)	Number of survivors (l_x)	Average future lifetime (e_x)
1920	0	0.06773	100,000	69.2
1921	1	0.01418	93,227	73.2
1925	5	0.00267	90,514	71.4
1930	10	0.00116	89,649	67.0
1935	15	0.00161	89,097	62.4
1940	20	0.00191	88,280	58.0
1945	25	0.00171	87,450	53.5
1950	30	0.00143	86,773	48.9
1955	35	0.00158	86,137	44.3
1960	40	0.00235	85,351	39.7
1965	45	0.00359	84,180	35.2
1970	50	0.00528	82,405	30.9
1975	55	0.00686	80,012	26.7
1980	60	0.00954	76,981	22.7
1985	65	0.01423	72,745	18.8
1990	70	0.02032	66,913	15.3
1995	75	0.03149	59,150	11.9
2000	80	0.07576	48,681	8.9
2005	85[a]	0.12186	35,003	6.4
2010	90[a]	0.19307	19,610	4.4
2015	95[a]	0.28845	7,124	3.1
2020	100[a]	0.37351	1,434	2.3
2025	105[a]	0.46279	152	1.7
2030	110[a]	1.00000	6	1.3

Source: Office of the Chief Actuary, U.S. Social Security Administration (2005), Table 4.7, pp. 95–97
[a]Projected

in the gap between life expectancy from the two types of tables. According to the generation life table for the cohort of females born in 1920, life expectation is 69.2 years, whereas the period life table for 1920 shows a life expectation at birth of 56.3 years. These figures indicate that the actual 1920 cohort is expected to live 13.0 years longer on the average than shown by the period life table for 1920.

The fact that the basic data for a single table pertain to many different calendar years means that the time reference for the generation life table is somewhat indefinite. Fully equating the two types of tables with respect to time reference is not possible. If a period life table is sought that would correspond more closely to a generation life table, we could select the period life table that has the same life expectation at birth as the generation life table. For example, the U.S. generation life table for 1950, with a life expectation at birth of 72 (M) and 78 (F), could be approximated by the period life table for 1990, which has similar life expectation figures (Table 4.7). Life expectancy for the female cohort born in 1900, 58.3 years, is reached in the period life tables for about 1925. In general, one can achieve greater

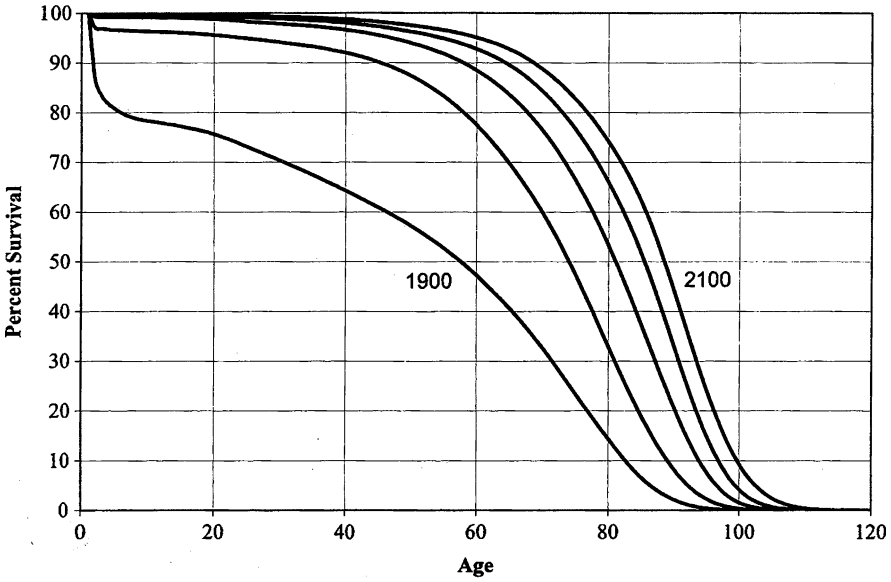


Fig. 4.2 Percent surviving by age: U.S. Social Security Area, 1900–2100 (For selected calendar years, 1900, 1950, 2000, 2050, and 2100; based on period life tables) (Source: U.S. Social Security Administration (2005), Table 4.6, page 26, and Fig. 4.5, page 16)

Table 4.7 Comparison of life expectation at birth, by sex, in period tables for specified years and generation life tables for cohorts born in these years: U.S. Social Security Area, 1900–2010

	Male			Female		
	Period table	Cohort table	Difference	Period table	Cohort table	Difference
Year	Years			Years		
1900	46.4	51.5	5.1	49.0	58.3	9.3
1910	50.1	56.2	6.1	53.6	63.7	10.1
1920	54.5	61.8	7.3	56.3	69.2	12.9
1930	58.0	66.1	8.1	61.3	72.8	11.5
1940	61.4	69.6	8.2	65.7	75.8	10.1
1950	65.6	72.2	6.6	71.1	78.4	7.3
1960	66.7	73.6	6.9	73.2	79.6	6.4
1970	67.2	75.4	8.2	74.9	80.9	6.0
1980	69.9	77.2	7.3	77.5	82.3	4.8
1990	71.8	78.9	7.1	78.9	83.3	4.4
2000	74.0	80.0	6.0	79.4	84.2	4.8
2005	74.9	80.5	5.6	79.6	84.6	5.0
2010	75.4	81.0	5.6	80.0	85.0	5.0

Source: Based on U.S. Office of the Chief Actuary, Social Security Administration (2005)

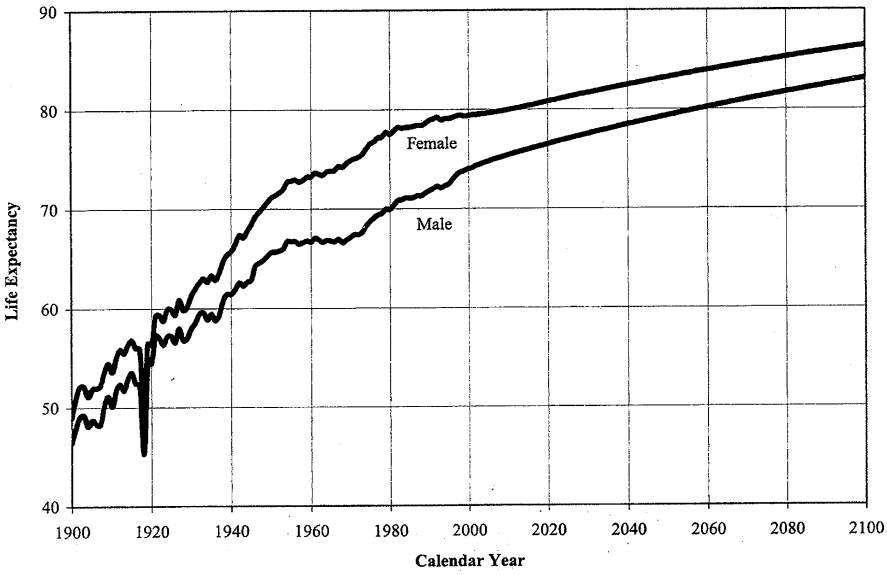


Fig. 4.3 Life expectancy at birth, by sex, based on period life tables: U.S. Social Security Area, 1900–2100 (Source: U.S. Social Security Administration (2005), Table 6, page 26, and Fig. 2a, page 10)

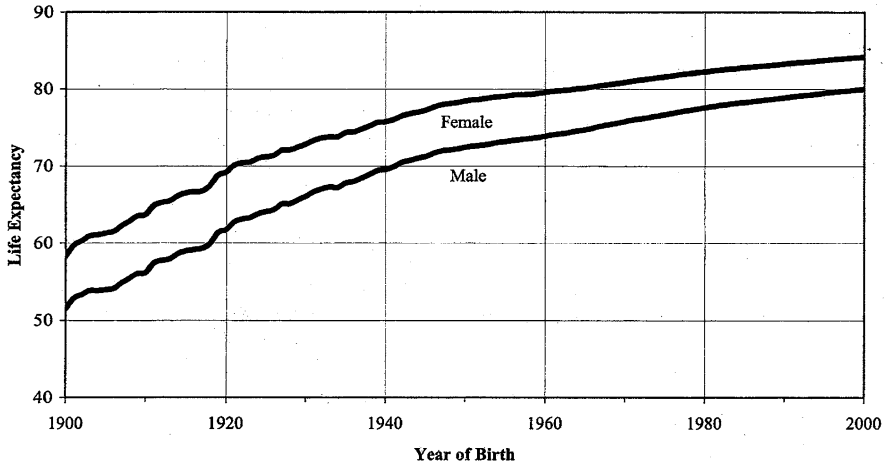


Fig. 4.4 Life expectancy at birth, by sex, based on cohort life tables: U.S. Social Security Area, 1900–2100 (Source: U.S. Social Security Administration (2005), Table 4.7, page 89, and Fig. 4.3a, page 13)

comparability between these two types of tables for past years by assigning the generation life table to the year 25–40 years after the cohort's year of birth. Whether this relation holds true in the future will depend on the future direction and rate of change in mortality.

A generation life table can involve the combination of mortality rates that are very different in nature, as a result of mortality improvement over a long period of time. Health conditions today are vastly different from those prevailing a 100 years ago, and the corresponding mortality patterns are also dissimilar. On the other hand, this type of table does reflect the actual combination of changing health conditions and intracohort influences to which particular cohorts are subject. A period life table reflects the influence of a more unitary set of health conditions and a single mortality pattern, but not ones to which a real cohort has been or can with assurance be exposed.

Analysis of the Standard Life Table

Level and Age Pattern of Deaths

In the life table, as in the observed population, deaths show a characteristic "bathtub" pattern, falling from a peak at infancy to a trough at about age 10, rising gradually to a second peak in the late adult ages, and finally falling away to zero as the original birth cohort is depleted. The distribution has two modal ages, but in referring to the modal age of life table deaths, the reference is normally to the second modal age. After this age, even with rising death rates there are too few survivors for increasing numbers of deaths to occur. The distribution of deaths by age may be described by various measures of central tendency, dispersion, and slope, some of which are discussed in the sections below.

Average Age at Death

The "average" age of deaths in the life table may be represented by three measures, the median age, the mean age, and the modal age. The values for these measures in the U.S. life table for 2000 are:

	Age at death		
Measure	Total	Male	Female
Median	80.3	77.7	82.8
Mean	76.9	74.1	79.5
Mode[a]	85.0	83.5	86.5

Source: Based on NCHS (2002b)
[a]Approximated. Modal age for adult deaths, disregarding the mode at infancy

These values differ substantially from one another. The mean age of deaths tends to be lower than the median age because the mean age takes all the deaths into account – unlike the median, which is a measure of position – and is pulled sharply in the direction of the greater number of years of life lost (i.e., to the left). The mode is the highest value and is least affected by the details of the level and shape of the age distribution. Note that the age pattern of deaths is quite unlike a normal curve (unless one disregards the ages under 20) and hence the relations between the averages seen in the normal curve are not found in the life table curve of deaths.[11]

The modal age of death is often difficult to determine exactly because the number of deaths may fluctuate from age to age or detailed data may not be available at the highest ages to identify a specific age. Nevertheless, in recent years some researchers have maintained that the modal age is an extremely useful measure for the analysis and comparison of the age distributions of deaths and survivors (Canudas-Romo 2006; Cheung and Robine 2007).

In 2000 the mean age at death in the life table was 77 years. The mean age of deaths is a weighted average of the ages at death, the weights being the deaths at each age. The mean age of deaths in the life table and the expectation of life at birth are equivalent. A simple proof is as follows:

$$e_0 = T_0 \div l_0 \tag{4.42a}$$

$$e_0 = \sum_0^\infty L_x \div \sum_0^\infty d_x \tag{4.42b}$$

$$e_0 = \sum_0^\infty a_x d_x \div \sum_0^\infty d_x \tag{4.42c}$$

where a_x refers to the age of a death and d_x refers to a death at age x. Since the sum of the ages of all the deaths of the members of the cohort equals the total years of life lived by the cohort, we can substitute $\sum a_x d_x$ for $\sum L_x$. Either formula (4.42a) or (4.42c) can be employed to determine average age at death according to the life table.

The corresponding values for the average ages at death in the observed population were given in Chap. 3. The life table values tend to be higher than the observed values because a life table stationary population with low mortality tends to be older than the corresponding observed population. A generation life table for 2000 (i.e., for the cohort born in 2000) would have an even higher average age at death than the period table for 2000 if the cohort is subject to lower age-specific death rates in the future than in 2000.

Average age of life table survivors and population. The modal age of survivors in the life table is clearly the age of the initial size of the birth cohort, i.e., birth (l_0). The modal age of the life table stationary population is also zero. These modal ages

[11] For further analysis it is useful to calculate the medians and means on the basis of life table data for ages 20 or 25 and over, so as to limit the analysis to the ages where the force of mortality is steadily increasing.

tell us nothing about the level of mortality. The median ages of survivors and of the stationary population are about 39.5 and 39.0, respectively – a median characteristic of an aged/low fertility/low mortality population. In contrast, the median age of life table deaths, that is, the age when the cohort numbers $^1/_2$ of the radix, or 50,000, is about 80 years in the U.S. 2000 life table.

The mean age of survivors and the mean age of the stationary population are approximately the same:

$$\sum \mathrm{al_x} \div \sum \mathrm{l_x} \sim \sum \mathrm{aL_x} \div \sum \mathrm{L_x} \tag{4.43}$$

The values of these two measures, calculated on the basis of the U.S. life table of 2000, are both about 39 years, with the mean age of the stationary population being slightly lower than the mean age of survivors.

Survival curve and median survival time with censoring. To this point it has been assumed that information on all deaths is available for every member of a cohort until the last member of the cohort expires. In many experimental studies the subjects enter the study at different dates and the study is terminated at a fixed date; as a result, some subjects survive through the experimental period without experiencing death or the other events covered in the study, or some subjects drop out and their fate is unknown. Biostatisticians refer to the data for those events for which no date or age is available as censored data. Under these conditions the method of deriving the survival curve described earlier does not apply and a modified method must be used that takes account of the absence of information for the censored data.

To derive the survival curve for experimental studies involving deaths that are censored, it is customary to use the Kaplan-Meier product-limit formula. This formula for calculating survivors involves alternately calculating a survival rate, determining the number of survivors, calculating the next survival rate, determining the number of survivors, and so on, and then taking the cumulative product of the survival rates. The formula is:

$$\mathrm{S(t_j)} = \prod [(\mathrm{n_i} - \mathrm{d_i}) \div \mathrm{n_i}] \tag{4.44}$$

where $\prod$, the operator for cumulative multiplication, indicates that the product is taken over all the times, t_i, at which deaths occurred up to and including time t_j. The symbol n_i is equivalent to l_x in the usual life-table notation. The calculation starts with $n_{i=0} = l_0$. A survival rate is calculated for each period during which deaths are known to have occurred:

$$(\mathrm{n_i} - \mathrm{d_i}) \div \mathrm{n_i} \tag{4.45}$$

A survival rate is not computed for the times of censored observations because no known deaths occurred at those times. Conventionally, biostatisticians draw the

survival curve as a series of step changes, with the steps occurring at the times of known deaths. Because some individuals are still alive at the end of the observation period for the study, the survival curve does not reach zero.[12]

To summarize this survival curve in a single number, it is customary to use the median survival time. This measure is analogous to the median age of survivors. It is derived from the data for the survival curve and is defined as the smallest observed survival time for which the estimated survival function is less than 0.5. Alternatively, the time corresponding to the point at which the survival function equals 0.5 may be derived by interpolating between the cumulative survival rates before and after 0.5 and reading off the corresponding time or age.

Shape of the Distribution of Deaths and Survivors

Many measures have been devised to describe the shape of the age distribution of life table deaths and survivors. Some focus on the shape of the survival curve and, specifically, on the "rectangularity of the survival curve." Others focus on the shape of the curve of life table deaths and, specifically, on the "compression" or "expansion of mortality." The various measures of these two related phenomena provide essentially the same information about them and are highly correlated. Several measures are explained below without linking them separately to the rectangularity of the survival curve and the compression of mortality.

Ratio of modal deaths. One simple measure of the shape of the age distribution of life table deaths is the ratio of the modal number of adult deaths to the number of infant deaths. This measure is illustrated with data from various U.S. life tables spanning a century, from the life table for the Original Death Registration States (ODRS) in 1900–1902 to the life table for the United States in 2000, and for the United Nations model life tables at very low and very high levels of life expectancy:

	Ratio of deaths, adult modal age to infancy		
Population	Deaths under 1 year	Deaths at –	Ratio
ODRS white females, 1900–02	11,061	Age 75 – 2039	0.18
U.S. white females,1950	2,355	Age 80 – 3446	1.46
U.S. white females, 2000	513	Age 86 – 3793	7.39
UN model table, females $e_0 = 35$	16,449	Age 67 – 1218	0.07
UN model table, females $e_0 = 75$	2,804	Age 82 – 3513	1.25

Source: U.S. National Center of Health Statistics or predecessor agencies. Coale et al. (1993)

[12]For an illustration with data and a graph, see S. A. Glantz, *Primer of Biostatistics*, Sixth Edition, New York: McGraw-Hill, 2005, Chapter 11.

Table 4.8 Measures of central tendency and dispersion of the age distribution of life table deaths: United States, 1900–2004

Year	Mode[a]	Mean[b]	Median	1st quartile	3rd quartile	IQR[c]	RIQR[c]
2004	85.5	77.8	81.3	71.7	89.1	17.4	21.4
2000	85.0	76.9	80.3	69.7	88.0	18.3	22.8
1989–1991	83.5	75.4	79.0	67.9	87.2	19.3	24.4
1979–1981	83.5	73.9	77.6	66.3	86.0	19.7	25.4
1969–1971	80.5	70.8	74.9	63.1	83.5	20.4	27.2
1959–1961	81.5	69.9	74.3	62.7	82.8	20.1	27.1
1949–1951	78.5	68.1	72.8	60.6	81.5	20.9	28.7
1939–1941	76.5	63.6	69.9	55.5	79.2	23.7	33.9
1900	72.8	47.7	56.7	20.2	73.3	53.1	93.7
Increase							
1970–2000	4.5	6.1	5.4	6.6	4.5	−2.1	−4.4
1940–1970	4.0	7.2	5.0	7.6	4.3	−3.3	−6.7
1900–1940	3.7	15.9	13.2	35.3	5.9	−29.4	−59.8

Source: Based on life tables published by U.S. National Center for Health Statistics or its predecessor agencies. Table for 1900: U.S. Office of the Chief Actuary, SSA (2005)

[a]Approximation

[b]Mean age of death, or life expectation at birth

[c]Interquartile range (IQR) and interquartile range as percent of median age at death (RIQR)

In very high mortality countries, which are also commonly the least developed countries, this ratio tends to be very low; here infant deaths dominate the age pattern of deaths. In very low mortality/more developed countries, the ratio tends to be high, but with great variations; here deaths at the very high ages tend to dominate the age pattern.

Interquartile range. The second measure, the interquartile range, is a common measure of the dispersion of distributions. Here it is applied to the deaths or the survivor population in the life table. The interquartile survival range is the difference between the ages to which 25% and 75% of the original cohort of births survive in the life table.

$$\mathrm{IQR} = \mathrm{x}_2 - \mathrm{x}_1 \tag{4.46}$$

where x_1 and x_2 represent the ages corresponding to $l_{x1} = 0.25$ and $l_{x2} = 0.75$. Table 4.8 presents a series on the IQR from U.S. official life tables from 1900 to 2000. The declines in the IQR in this period indicate that the survival curve was becoming more rectangular and that the deaths were becoming more compact with respect to age. The IQR fell by 35 years in this 100-year period. As the age distribution of life table survivors at the older ages becomes more rectangular, the IQR falls. Most of the decline in the IQR occurred in the first 40 years of the century; only a small share of it occurred between 1940 and 2000. Moreover, the decline in the IQR was greater from 1940 to 1970 (3.3 years) than from 1970 to 2000 (2.1 years). The process of rectangularization of the survival curve has been steadily slowing down as life expectancy has been rising.

An even slower shift in the "squaring" of the survival curve is indicated if the age range selected for calculating the IQR is limited to the adult ages, say 25 years and over, where the mortality curve is monotonic, as shown here:

Year	Md	1st Q	3rd Q	IQR	Tricennial change in IQR
1939–41	71.6	60.3	80.0	19.7	–
1969–71	75.6	65.0	83.9	18.9	−0.8
2000	80.6	70.5	88.1	17.6	−1.3

Source: Based on life table of the U.S. National Center for Health Statistics (1969–1971 and 2000) and U.S. Census Bureau (1939–1941)

These figures show that the decline in the IQR over the 60-year period, 1940–2000, was only 2.1 years for the adult ages, as compared with 5.4 years when the figures are computed for all ages. We should expect smaller changes in the IQRs when the range is restricted to the adult ages, because the omission of childhood mortality reduces the variability of deaths while the upper age limits of survival are essentially fixed. According to these modified figures, there is a slight quickening of the rectangularization process, but the movement is very muted.

Standard deviation. The next measure of dispersion of the age distribution of life table deaths, its standard deviation, is computed in the conventional way from life table deaths:

$$\sigma_d = \sqrt{\sum (X_d - X_m)^2 \div n} \tag{4.47a}$$

$$= \sqrt{\left(\sum X_d^2 \div n\right) - X_m^2} \tag{4.47b}$$

$$= \sqrt{\left(\sum X_d^2 \div n\right) - e_0^2} \tag{4.47c}$$

where x_d refers to the age of deaths, x_m and e_0 to the mean age of deaths, and n to the total number of deaths. The summation is over the n deaths. A lower standard deviation corresponds to a lesser dispersion of the deaths with respect to age.

Gini coefficient. The Gini coefficient, another measure of dispersion, is normally used to measure inequality in income distributions, but it may be adapted for measuring inequality in the age distribution of deaths (Hanada 1983; Wilmoth and Horiuchi 1999). The formula in its life table application is,

$$\text{Gini Coefficient} = 1 - \int [l(x)/l(0)]^2 \, dx \div e_0 \tag{4.48a}$$

$$\text{or} \qquad 1 - \sum (L_x/l_0)^2 \div e_0 \tag{4.48b}$$

Its computation involves deriving the sum of the squares of the stationary population (on a decimal basis) at each age, dividing the result by the expectation of life, and taking the complement of the quotient. A decline in the Gini coefficient means less variation in the age distribution of deaths.

Keyfitz' H. Keyfitz' H, the next measure described in this section, has a dual interpretation. It was originally designed to measure the relation between a proportional reduction in death rates and a proportional rise in life expectation. It can also serve as a measure of variation of the age of deaths (Wilmoth and Horiuchi 1999; Nusselder and Mackenbach 1996). I focus only on the second interpretation here; the other interpretation will be treated in a later section of the chapter.

As the variation in age at death declines, the value of H declines. The formula for H is, in continuous notation,

$$\mathrm{H} = -\int_0^\infty \mathrm{l(x)}\,\ln[\mathrm{l_x}/\mathrm{l_0}]\mathrm{dx} \div \int_0^\infty \mathrm{l(x)} \tag{4.49a}$$

This formula "translates" into the following equation in discrete notation,

$$\mathrm{H} = -\sum_0^\infty [\mathrm{L_x} * \ln(\mathrm{l_x}/\mathrm{l_0})\,] \div \sum_0^\infty (\mathrm{L_x}) \tag{4.49b}$$

$$\mathrm{H} = -\sum_0^\infty [\mathrm{L_x} * \ln(\mathrm{l_x}/\mathrm{l_0})] \div \mathrm{T_0} \tag{4.49c}$$

One can read the formula as the weighted average of the natural logarithms of the survivorship function, with the stationary populations serving as weights. H measures the degree to which the survivorship, or l_x, curve is concave upward; larger values of H are associated with a greater concavity of the l_x curve. An l_x curve that is convex has a rather small H. The H value of .1298 for the life table of the U.S. population in 2000 indicates a smaller variation in age at death than the H value of .1792 in 1949–1951.

Greatest downward slope. Finally, I list the greatest downward slope (GDS), which is the (negative) slope of the survival curve at its maximum over the adult ages. It indicates the steepness of the survival curve as it falls at these ages. The formula is,

$$\mathrm{GDS} = \max\{-l'(x)\} \text{ or } \max\{-S'(x)\} \text{ for } \mathrm{x} \geq 25 \tag{4.50}$$

In actual application, first differences are examined instead of derivatives and the maximum first difference is selected. The value of the GDS increases as the rectangularity of the survival curve increases.

Comparison of Life Table Functions

To this point I have discussed the functions of the life table and the measures for analyzing them without referring to their use in comparisons of mortality and survival between populations. Normally it is acceptable simply to take the absolute

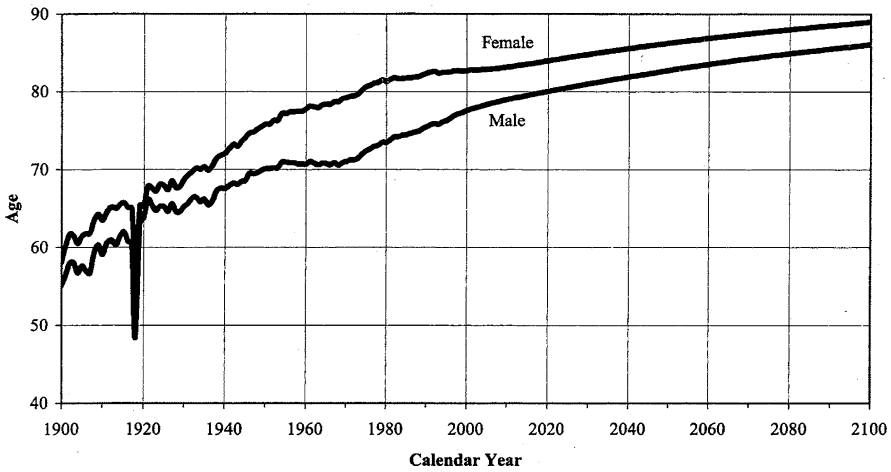


Fig. 4.5 Median age at death, by sex, based on period life tables: U.S. Social Security Area, 1900–2100

difference of the two values of any summary measure such as life expectancy, the life table death rate, the median age at death (or of survivors), and the modal age at death. In addition to these measures, other measures are available when whole distributions are being compared for differences in their patterns (Fig. 4.5).

Comparisons of survivors, deaths, and mortality rates. For the survivorship function, Chen *et al.* (2007) suggest, as one measure, the maximum value of the differences of corresponding values of the survival distributions:

$$\| \mathbf{l_1} - \mathbf{l_2} \|_{\max} = |l_{1x} - l_{2x}| \tag{4.51}$$

where l in bold font represents a vector of age-specific survivors from the life table. These measures can be applied also to comparisons of the patterns of life table deaths and mortality rates. Simply find the maximum of the differences of the distributions of life table deaths or the distributions of mortality rates being compared. Such measures resemble those for the observed deaths and death rates. Alternatively, take the sum of the differences without regard to sign:

$$\int \| \mathbf{d_1} - \mathbf{d_2} \| \, dx \tag{4.52a}$$

Or in discrete notation,

$$\sum\nolimits_x |d_{1x} - d_{2x}| \tag{4.52b}$$

The d_x's could represent absolute deaths or the percents of deaths in the total distribution of deaths. When percents are used, the measure is the same as the index

of dissimilarity (ID) (except that the index of dissimilarity calls for one-half of the sum of the differences). Two distributions of mortality rates can be compared by the following formula:

$$\sqrt{\int \| \mu_1 - \mu_2 \|^2 \; dx} \cong \sqrt{\sum (\mu_{1x} - \mu_{2x})^2} \tag{4.53}$$

In this case the differences of mortality rates are squared, the squares are summed, and the square root of the total is taken. This formula gives greater weight to the higher mortality at older ages than formula 4.52, expresses for death rates.

Scaling the mortality curve. Populations observed over long periods of time, populations of very different sizes, and populations at very different levels of socioeconomic development are likely to have very different levels and patterns of mortality and survival. Moreover, humans and other animal species (e.g., dogs, pigs, mice) have very different life spans. Hence, it is necessary to make some adjustment in the data before meaningful comparisons can be made between these different populations (Carnes et al. 2006). The mortality levels and patterns of these populations cannot be validly compared without taking account of these gross differences. We are usually interested in comparing the corresponding summary measures of two distributions rather than each element in the two distributions.

Relative interquartile range and coefficient of variation. Two basic approaches can be considered to effect comparisons between the mortality patterns of two or more widely different populations. One can adjust or standardize the age distributions of life table death rates for their general level and then compute a summary measure of dispersion. This scaling operation is accomplished by dividing the original rates (M_x) in a distribution by their mean, median, or total, and then calculating the measure of dispersion from the modified rates. When the adjusted rates are obtained by dividing the original rates, M_x, by their mean, $\sum M_x/n$, the new distribution has a mean equal to one and a standard deviation quite different from the original standard deviation.[13] The new standard deviation obtained in this way is already a relative measure, or the new coefficient of variation, since dividing it by the new mean (=1) equals itself.

Alternatively, one can compute the measure of dispersion first and then standardize it by dividing it by the mean or median. In this scaling operation, one divides

[13]After dividing the death rates in the different distributions being compared by their means, the standard deviations are equal to $\sqrt{\sum (M'_x - 1)^2 \div n}$. The new standard deviations differ greatly from the original ones. As mentioned in Chapter 3, division of the elements in a distribution by some factor, whether it is the median, mean, or total, modifies the shape of the distribution, including its standard deviation and variance (i.e., the square of the standard deviation). Division of a distribution by factor a changes the original variance by the factor $1/a^2$, the original standard deviation by a factor of $1/a$, and the original coefficient of variation by a factor of $1/a$ if the factor $(=1/a)$ is the mean.

the standard deviation of the original distribution by the mean of the distribution to derive the coefficient of variation. We have then σ/M_{mn}. Expressing the measure as a percent, we have,

$$C.V. = (\sigma \div M_{mn}) * 100 \tag{4.54}$$

It is simpler and equally logical to calculate the relative interquartile range (RIQR), standardizing the interquartile range (IQR) by dividing it by the median, or second quartile, of the distribution (M_{md}):

$$RIQR = (IQR \div M_{md}) * 100 \tag{4.55a}$$

For some analysts the measure of choice for comparing the dispersion of mortality distributions of populations that refer to very different dates is the relative interquartile range based on life table deaths.

$$RIQR_d = (IQR_d \div d_{md}) * 100 \tag{4.55b}$$

Thus, in 1940 for the United States, the relative interquartile range ($RIQR_d$) is computed as $(23.7 \div 69.9) * 100 = 33.9$, while in 2000 the $RIQR_d$ is computed as $(18.3 \div 80.3) * 100 = 22.8$ (Table 4.8). These results suggest a substantial degree of compression of the distribution of deaths over this 60-year period, amounting to 11.1 standardized years. If, however, we confine the range of ages for calculating the IQR and RIQR to the ages 25 and over so as to focus attention on that part of the survival curve most relevant to the compression/rectangularization issue, we derive substantially lower RIQRs. For example, in 1940, the RIQR was 27.6 years, and in 2000 the corresponding figure was 21.9. Over this 60-year period the decline in the RIQR was only 5.7 years as compared to 11.1 years when all ages are included in the computation.

Limitations of Standard Life Tables and Proposed Adjustments

Synthetic vs. Real Cohorts

The leading criticism that has been made of the conventional life table is that it represents a hypothetical construct combining age-specific death rates that do not apply to any real birth cohort. The conventional table is said, therefore, to describe a synthetic cohort rather than a real one. It is even possible that some schedules of annual age-specific death rates could not apply to any real cohort because of the changing historical levels of mortality affecting the different birth cohorts in any current year and the changing health composition of these birth cohorts ("heterogeneity effects"). Each birth cohort that the current schedule of rates comprises has a different experience with regard to migration, epidemics, war, economic privation, and so on, that affects its mortality history and health

composition. In short, the life expectancy from a period life table does not represent the actual mortality and life expectancy of the birth cohort for the year in which it is calculated or for any cohort composing it, it does not describe the future experience of the persons alive in that year, and it is possible that it could not realistically apply to any real cohort. Rather, the conventional life table is a hypothetical construct intended to summarize current mortality, not the past or future mortality history of the cohorts composing it.

To answer the principal criticism that the life table is an artificial construct that does not apply to any real group of people, the generation, or cohort, life table was developed. It applies to a single real cohort and, as the reader may recall, requires age-specific death rates for over a century to complete. On the other hand, the generation life table offers little information regarding mortality in any particular calendar year and represents the experience of only a single birth cohort for the year of reference. It is evident that one type of table could not substitute for the other and that the two types of tables are different analytic tools for measuring mortality.

Most demographic measures are hypothetical constructs (e.g., the age-adjusted death rate and the total fertility rate), and demographers often employ several different measures to describe different aspects of a particular demographic phenomenon to serve different analytic purposes. The "moral" is that it is useful to devise different types of life tables or life table measures to serve different analytic ends.

Other Limitations

The accuracy of a life table depends on the accuracy of the age-specific death rates on which it is based. In the more developed countries questions are raised most commonly with respect to the accuracy of the death rates at the advanced ages, such as ages 85 and over. The rates can be distorted by the sparseness of the data, the greater tendency for the ages of survivors and deaths to be misreported, and the obstacles in securing accurate data for respondents who may be too ill or frail to report accurately for themselves. Although these rates have little effect on life expectation at birth, they are quite important in analyses relating to mortality patterns and trends at the tail-end of life and issues concerning maximum human life span. Mortality rates at the upper end of the table are also affected by a ceiling effect – values "with little place to go." Given the fact that there is, at present, an apparent limit to the duration of human life, measures relating to this part of the age distribution may be distorted in relation to measures relating to the younger part. It is not possible to evaluate the rates at the advanced ages confidently by examining rates at both younger and older ages because of the very fact that rates at the higher ages are inaccurate. Fortunately at the present time, access to alternative sources of data, new evaluation methods, and historical improvements in the quality of data have made the management of this issue more feasible. For many of the less developed countries securing accurate sets of age-specific death rates for constructing life tables remains elusive. This situation is discussed in Chap. 12.

We have seen that life expectation at birth is equivalent to the average age of life table deaths and hence is a function of a set of age-specific mortality rates. It may also be viewed as a function of the set of life table death rates (${}_nM_x$). If the life table death rates are weighted by the stationary population in the life table (${}_nL_x$), we derive the life table death rate, the reciprocal of the life expectancy at birth:

$$\sum(M_x\ L_x) \div \sum L_x = \sum d_x \div T_0 = l_0 \div T_0 = l \div e_0 \qquad (4.56)$$

Since the stationary population is generated through a series of steps from a set of age-specific mortality rates, life expectancy at birth is independent of the observed population. Accordingly, it may be considered an age-standardized measure. Because, however, the populations serving as weights of the age-specific death rates are not the same from life table to life table (i.e., from area to area or date to date), life expectation figures for different populations are not fully comparable. While life expectancies are not based on a common weighting scheme, they maintain the "correct" rank order. The life table death rate is a standardized death rate *sui generis*.

Proposed Adjustments of the Standard Life Table

Demographers have been seeking a modification of the standard life table that describes mortality for a particular year while taking account of the mortality history of the individual birth cohorts living in that year. They have been distinguishing current mortality conditions and historic mortality conditions in formulating their criticisms and modifications of the conventional table. In recent decades the first analysts to describe an alternative measure of life expectation that serves this purpose was Brouard (1986) and Sardon (1994), and their ideas were elaborated by Bongaarts and Feeney (2002, 2003, 2005). The Bongaarts/Feeney proposal was widely criticized (Vaupel 2002; Guillot 2003; Wilmoth 2005). More recently, Schoen and Canudas-Romo (2005) proposed a different and more defensible measure. These alternative measures should be viewed as additions to the tools available for measuring longevity, not as substitutes for the conventional measure.

Bongaarts-Feeney Adjustment

Bongaarts and Feeney (2002, 2003, 2005) maintain that life expectancy at birth as conventionally calculated is a distorted representation of mean lifetime whenever the mean age of deaths is changing. If the mean age at death is rising, period life expectancy at birth overestimates "actual" life expectancy and if mean age at death is falling, period life expectancy underestimates "actual" life expectancy. They argue that period life expectancy at birth as now computed in the standard life table has a

"tempo bias" and should be adjusted downward for this bias, by the special formulas they provide. Bongaarts and Feeney derive a tempo-adjusted life expectancy for ages 30 and over, first by dividing the observed age-specific death rates by $1 - r_p(t)$, where $r_p(t)$ denotes the rate of change in the period mean age of deaths, and then by constructing a life table from these age-adjusted age-specific rates. Their formula for the adjusted life expectancy is,

$$e_0(t)^* = \int_0^\infty \exp\left\{-\int_0^a \mu(x,t) \div 1 - r_p(t)dx\right\} da \tag{4.57}$$

which, in effect, says in continuous notation what was said just above the formula. They also use another formula to calculate period mean age at death: $\int_0^\infty p(a,t)da$, with $p(a,t)$ denoting the proportion of the cohort born at time $t - a$ which survives to age a. This formula is identical to the measure called CAL, the cross-sectional average length of life, proposed by Brouard (1986) and Guillot (2003). The alternative way of measuring life expectation at birth they propose takes account of past fluctuations in age-specific death rates and so combines period and cohort mortality.

Where death rates are changing in the same direction over a series of years, CAL shows a trend that is similar to life expectancy in period life tables but at a somewhat different level. CAL is sensitive to fluctuations in period and cohort mortality and to tempo effects. It produces a consistently less favorable life expectation under conditions of declining mortality than the conventional life expectation. Bongaarts and Feeney estimated the upward distortions in female life expectancy at birth for 1970–1990 to be 1.5 years in Denmark, 1.9 years in Sweden, and 3.3 years in Japan (assuming no mortality under age 30).

Schoen/Canudas-Romo Average Cohort Life Expectancy

The Schoen/Canudas-Romo (2005) proposal, called the "average cohort life expectancy" (ACLE), is a weighted average of the life expectancies of all the cohorts present in a given year, the weights being the probability of survival of each cohort to that year. ACLE requires data that span over two centuries; calculations for contemporary populations require projections of mortality for nearly all cohorts alive in the subject year. ACLE is relatively insensitive to both period and cohort fluctuations in death rates and, unlike cohort life expectancy, which represents one cohort, takes all the cohorts affected by mortality in a given year into account. After comparing period life expectancy, cohort life expectancy, and CAL, with ACLE, Schoen and Canudas-Romo conclude that ACLE gives the best representation of survivorship of cohorts living in a given year.

Extensions of the Standard Life Table

As we have seen, the basic life table describes how a single hypothetical cohort of births is reduced by a single factor, deaths, as it ages. A cohort can be reduced by other factors and even augmented by still others during its lifetime. This section explains how the basic life table can be extended to measure other lifetime changes to a cohort in addition to death. Life tables that describe losses to a cohort from more than one factor are called multiple-decrement tables. Tables that describe increases as well as decreases to the cohort are called multiple increment-decrement tables. These tables are distinguished by the principal methods used in their construction, the complexity of the methods, the basic data required, and the types of information produced. Multiple-decrement tables may be constructed by the prevalence-ratio method, the occurrence-exposure method, or the multistate method. Multiple increment-decrement tables may be constructed by the occurrence-exposure method and the multistate method. The discussion of these special types of life tables is relatively brief in this chapter; the subject is treated further in Chap. 8.

Multiple-Decrement Tables

Uses and Methods

With multiple-decrement tables there are two or more ways to exit from the initial cohort, one of which is mortality and the other(s) some change in demographic, health, or socioeconomic status. Multiple-decrement tables incorporate conventional life table components and build on basic life table techniques. They have been employed in the analysis of cause of death, disability, the labor force, marital status, migration, school enrollment, and other demographic and socioeconomic characteristics of the population. The factor or factors reducing the cohort in addition to total deaths may be a specific cause or causes of death, becoming disabled, incurring a chronic disease, bearing a first child, marrying for the first time, joining the labor force for the first time, entering a nursing home for the first time, and similar factors.

Prevalence-ratio method. A common way of constructing such tables is the prevalence-ratio method (also called the Sullivan method). It involves the application of prevalence ratios observed in the general population to total life table deaths or the life table stationary population, to derive the life table numbers in the special status being studied. In this way, life table deaths or the life table stationary population are distributed into different statuses according to the prevalence of those statuses in the observed population. For a study of the disabled population, for example, the proportions disabled in each age group in the observed population, obtained from a census or sample survey, are assumed to apply to the life table stationary population, so as to obtain the life table stationary disabled and

nondisabled populations. Once this partition is made the analyst can generate the rest of the life table, working forward and backward to derive the other functions relating to disability, e.g., the number of nondisabled survivors to each age or average remaining years of nondisabled life. For the present discussion, I illustrate the prevalence-ratio method with so-called cause-of-death elimination life tables (see below) and leave to Chap. 8 the detailed discussion of methods of life table construction and analysis combining mortality and morbidity.

Occurrence-exposure method. Another method of constructing multiple-decrement tables involves the use of occurrence-exposure rates, in which occurrence rates for the phenomenon under study (e.g., becoming disabled for the first time or net number becoming disabled), along with mortality rates, are applied to survivors. In this method a synthetic cohort is diminished jointly by mortality rates and another risk factor (e.g., net disability rates). Such tables provide an indication of the average number of disabled years to be expected after a given age by all survivors or by disabled survivors. In addition, the tables provide information on the chances of ever becoming disabled and age-specific rates of net accession to disabled life. Tables of disabled life ("tables of healthy life") have been available for only a few decades and a few countries. Other tables could be constructed describing a cohort's history with respect to incurring a chronic disease (e.g., heart disease) or a specific health event (e.g., heart attack).

The treatment of the risk of the health event as a net number or a first-time event, without separate allowance for incurring a disease and recovering from it, makes the resulting tables approximate. The procedure used generally assumes that the death rates of the general population and the study population (e.g., disabled persons; persons with heart disease) are the same. This assumption is questionable and makes the results less realistic and less accurate.

Cause-of-Death Elimination Life Tables

Interpretation and Uses

A cause-of-death elimination life table is an extension of the conventional life table that measures survivorship, deaths, and life expectancy under the assumption that a particular cause of death or group of causes of death has been eliminated. The assumption that a cause of death has been eliminated means that one cannot die from this cause and it must be excluded from the total number of deaths. Table 4.9 presents an abridged life table for the population of the United States in 1989–1991 that eliminates cancer as a cause of death.

A series of such life tables, compared with the conventional life table allowing for all causes combined, provides some new measures of the relative importance of different causes of death, namely the gain in expectation of life that would occur if the particular cause of death were eliminated and the probability of eventually dying

Table 4.9 Abridged standard life table and life table eliminating cancer as a cause of death: United States, 1989–1991

Period of life between two exact ages stated in years x to x+1	Proportion of persons alive at beginning of age interval dying during age interval ${}_nq_x$	Of 100,000 born alive		Average number of years of life remaining at beginning of age interval e_x
		Number living at beginning of age interval l_x	Stationary population in age interval ${}_nL_x$	
Eliminating no cause				
0–1	0.00936	100,000	99,258	75.37
1–5	0.00189	99,064	395,814	75.08
5–10	0.00112	98,877	494,084	71.22
10–15	0.00133	98,766	493,578	66.29
15–20	0.00426	98,635	492,218	61.38
20–25	0.00554	98,215	489,742	56.63
25–30	0.00615	97,671	486,890	51.93
30–35	0.00771	97,070	483,548	47.23
35–40	0.00985	96,322	479,328	42.58
40–45	0.01278	95,373	473,966	37.98
45–50	0.01895	95,154	466,606	33.44
50–55	0.02936	92,370	455,539	29.03
55–60	0.04596	89,658	438,678	24.83
60–65	0.07036	85,537	413,441	20.90
65–70	0.10264	79,519	378,129	17.28
70–75	0.15287	71,357	330,664	13.96
75–80	0.22110	60,449	269,741	11.00
80–85	0.32525	47,084	197,518	8.40
85–90	0.46346	31,770	121,174	6.23
90–95	0.63147	17,046	56,032	4.50
95–100	0.77332	6,282	17,158	3.29
100 and over	1.00000	1,424	3,508	2.46
Eliminating malignant neoplasms (140–208)				
0–1	0.00934	100,000	99,260	78.72
1–5	0.00175	99,066	395,855	78,46
5–10	0.00096	98,893	494,206	74.60
10–15	0.00117	98,798	493,766	69.67
15–20	0.00405	98,682	492,499	64.75
20–25	0.00526	98,282	490,144	60.00
25–30	0.00572	97,765	487,464	55.30
30–35	0.00691	97,206	484,413	50.61
35–40	0.00834	96,534	480,736	45.94
40–45	0.00989	95,729	473,396	41.30
45–50	0.01327	94,783	470,980	36.69
50–55	0.01900	93,526	463,492	32.15
55–60	0.02890	91,748	452,556	27.72
60–65	0.04467	89,097	436,062	23.46
65–70	0.06790	85,117	411,798	19.44

(continued)

Table 4.9 (continued)

Period of life between two exact ages stated in years x to x+1	Proportion of persons alive at beginning of age interval dying during age interval ${}_nq_x$	Of 100,000 born alive		Average number of years of life remaining at beginning of age interval e_x
		Number living at beginning of age interval l_x	Stationary population in age interval ${}_nL_x$	
70–75	0.10827	79,337	376,115	15.66
75–80	0.17008	70,747	324,472	12.25
80–85	0.27173	58,714	254,085	9.23
85–90	0.41446	42,760	168,451	6.74
90–95	0.59446	25,037	84,814	4.78
95–100	0.74997	10,154	28,428	3.43
100 and over	1.00000	2,539	6,408	2.52

Source: U.S. National Center for Health Statistics (1999)
Numbers after cause of death are category numbers of the Ninth Revision of the International Classification of Diseases, 1975

from a particular cause. The gain in life expectancy from hypothetically eliminating a cause is derived by subtracting life expectation in the standard table from life expectation in the cause-elimination table. For example, a comparison of e_0's in the two life tables in Table 4.9 shows that 3.4 years of life ($= 78.72 - 75.37$) would be added to the life expectancy at birth of the U.S. population in 1989–1991 if cancer were eliminated as a cause of death (U.S. NCHS/Anderson 1999). The chance of ever dying from a particular cause is obtained by dividing the cumulated number of life table deaths from this cause over all ages from the cause-elimination table by the radix of the all-causes life table (*i.e.*, 100,000). For example, according to the U.S. 1989–1991 cause-elimination table for cancer, the probability at birth of eventually dying from cancer (not shown here) is $0.22 (= 22,023 \div 100,000)$.

Eliminating a single cause in this way assumes that the timing and levels of the rates for the remaining causes are not affected. We say then that these tables are constructed under the assumption of the independence of the various cause-specific death rates. Such cause-elimination tables were computed by the U.S. National Center for Health Statistics decennially from 1959–1961 to 1989–1991 for each of several dozen causes; similar tables for 1999–2001 are planned. Such tables were prepared for all the leading causes and an "all other causes" category, so that all causes are nominally covered by a set of such tables. Appendix Table 4.A1 shows the gains in life expectancy at birth from eliminating each of the leading causes and Appendix Table 4.A2 shows the probability of ever dying from them for 1989–1991.

In 1989–1991 the elimination of heart disease and cancer shows the largest gains in life expectancy at birth. Elimination of heart disease would increase life expectancy at birth by over $4\frac{1}{2}$ years as compared with the gain of nearly $3\frac{1}{2}$ years from eliminating cancer. The elimination of only two other cause-categories would result in gains as great as 1 year or so, diseases of the respiratory system (0.97) and accidents and other adverse effects (0.92). The residual category "Other than

the 15 leading causes of death" would add only 1.96 years. Table 4.A1 suggests that substantial future increases in life expectancy will have to come mainly from reductions in heart disease and cancer, which, according to the reported data, contribute half of all deaths. The number of deaths resulting from other individual causes is too small to effect a large future change in life expectancy, except in combination with several leading causes. Similarly, heart disease and cancer lead the list of causes with respect to the chance of ever dying. The proportions are 0.36 and 0.22, respectively. The chance of ever dying from any other cause is quite small. For example, the proportions are only 0.10 for respiratory diseases and 0.03 for accidents and adverse effects.

Method of Construction of Cause-Elimination Life Tables[14]

The first step in the preparation of an abridged multiple-decrement table eliminating a cause of death is to subdivide the total number of life table deaths at each age group in the conventional life table into the different causes or groups of causes of death. The subdivision of total life table deaths by cause is made on the basis of the distribution of deaths by cause in the actual population.

$$ {}_{n}\mathrm{d}_{\mathrm{x}}^{\mathrm{i}} = {}_{n}\mathrm{r}_{x}^{\mathrm{i}} * {}_{n}\mathrm{d}_{\mathrm{x}} \tag{4.58} $$

where i represents a particular cause of death, ${}_{n}r_{\mathrm{x}}^{\mathrm{i}}$ represents the proportion of deaths due to cause i according to the observed death statistics, and ${}_{n}d_{x}$ represents life table deaths. From the observed deaths,

$$ {}_{n}\mathrm{r}_{\mathrm{x}}^{\mathrm{i}} = {}_{n}\mathrm{D}_{\mathrm{x}}^{\mathrm{i}} * {}_{n}\mathrm{D}_{\mathrm{x}} \tag{4.59} $$

The next step is the calculation of the probabilities of survival with the ith cause eliminated, or ${}_{n}p_{\mathrm{x}}^{(-\mathrm{i})}$, for $x = 1, 5, 10, 15, \ldots, 105$.

$$ {}_{n}\mathrm{p}_{\mathrm{x}}^{(-\mathrm{i})} = {}_{n}\mathrm{p}_{\mathrm{x}}^{(1-\mathrm{nrxi})} \tag{4.60} $$

where ${}_{n}\mathrm{p}_{\mathrm{x}} = \mathrm{l}_{\mathrm{x}+\mathrm{n}} \div \mathrm{l}_{\mathrm{x}}$ was derived from the corresponding life table for all causes combined. Values of $l_{x}^{(-I)}$ were derived sequentially by the formulas,

$$ \begin{aligned} \mathrm{l}_{0}^{(-\mathrm{i})} &= 100{,}000 \\ \mathrm{l}_{\mathrm{x}+\mathrm{n}}{}^{(-\mathrm{i})} &= {}_{n}\mathrm{p}_{\mathrm{x}}^{(-\mathrm{i})} * \mathrm{l}_{\mathrm{x}}^{(-\mathrm{i})} \end{aligned} \tag{4.61} $$

[14] A more compete exposition of the method of constructing such tables is given in U.S. NCHS/Anderson (1999).

The age-specific probabilities of death eliminating the ith cause were then obtained as the complements of ${}_np_x^{(-i)}$:

$$ {}_{\mathrm{n}}\mathrm{q}_{\mathrm{x}}^{(-\mathrm{i})} = 1 - {}_{\mathrm{n}}\,\mathrm{p}_{\mathrm{x}}^{(-\mathrm{i})} \tag{4.62} $$

This formula represents the probability that a person surviving to age x will die within n years if the ith cause of death is eliminated.

The stationary population in the age interval x to $x+n$ ($L_x^{(-1)}$) is estimated by weighting $l_x^{(-l)}$ and ${l_{x+n}}^{(-l)}$ by factors representing the average number of years lived by those who died within the interval and its complement. These factors are taken from the life table for all causes combined. $T_x^{(-1)}$ and $e_x^{(-1)}$ are obtained in the usual way on the basis of values of $L_x^{(-l)}$ and $l_x^{(-l)}$. As stated earlier, the gain in expectation of life due to the elimination of a specified cause of death is derived as the difference between life expectancy in the life table eliminating this cause of death and life expectancy at the same age in the life table for all causes combined:

$$ \mathrm{g}_{\mathrm{x}}^{(-\mathrm{i})} = \mathrm{e}_{\mathrm{x}}^{(-\mathrm{i})} - \mathrm{e}_{\mathrm{x}} \tag{4.63} $$

The probability that an individual aged x will eventually die from the ith-cause was calculated by the formula,

$$ {}_{\infty}\mathrm{q}_{\mathrm{x}}{}^{\mathrm{i}} = \mathrm{l}_{\mathrm{x}}{}^{\mathrm{i}} \div \mathrm{l}_{\mathrm{x}} \tag{4.64} $$

where l_x is the number of survivors to age x in the life table for all causes combined and l_x^i is the total number of deaths due to the ith cause at all ages x and over.

Limitations: Competing risks and other issues. Cause-of-death elimination life tables are approximate guides as to the relative importance of the principal causes of death, expressed in terms of the gains in life expectancy from eliminating each cause and the chance of ever dying from each cause. They are not projections of mortality, but they are suggestive of the areas where mortality reductions are required for substantial improvements in life expectancy. Cause-elimination life tables as calculated by NCHS are based on the assumption that the risk of dying from a particular cause is independent of the risks of dying from other causes. In other words, the method of construction disregards the competing risks of the causes of death. Competing risk characterizes the situation where an individual is subject to more than one type of risk from a class of events. For this purpose we can consider the causes as divided into any number of classes – only two broad classes, such as endogenous and exogenous causes of death, or numerous specific categories of causes. These are the risks that compete for the individual's life, and the risk that is realized first determines how long the individual lives. Making allowances for competing risks in the construction of cause-elimination life tables would contribute to more realistic estimates of years gained by the elimination of the various causes of death. It would also contribute to a more informed analysis of trends in mortality and a sounder basis for projecting mortality. While the primary issue is evident – that

the elimination of one cause of death could have an effect on the timing and level of the other causes of death – the solution is complex and subject to considerable uncertainty, as explained further below.

The occurrence of death from one cause (e.g., stroke) removes numerous individuals from experiencing death from any other cause. As a consequence, the risk of dying from the other causes is reduced. This is a common situation affecting mortality trends. If, on the other hand, a cause of death is eliminated and individuals are thereby saved, they are at risk for death from the other causes. With the elimination or even substantial reduction of one or more causes, not only will more persons be at risk of dying from the other causes, but the risks for the other causes will tend to rise. This is because of the likelihood that the individuals saved are excessively vulnerable to other diseases, given the frequent occurrence of comorbidities (i.e., combinations of chronic illnesses) and the common association of many other causes with the underlying or reported cause of death. The outcome involving increases in other causes with the elimination or near-elimination of one or more causes is also affected by the possibility of an upper limit to human longevity and the near certainty of death from some cause or combination of causes before this possible limit is reached. The identification and timing of the other causes of death depend on the mean age of the cause eliminated and the mean age of the comorbidities and associated conditions.

We know intuitively that, if all causes are eliminated simultaneously, unlimited survival would result. On the other hand, according to the U.S. cause-elimination life tables in 1989–91, the sum of the years gained by eliminating every leading cause of death and the remaining causes as a group independently is merely 15 years. This is a paradox in longevity analysis. We can reasonably conclude that at low mortality levels the reduction of the death rate from any one cause has a damping effect on the decline in other causes.[15]

Because of the independence assumption, the gains are not simply additive. In other words, the sum of the gains from two causes or more is not equal to the gains from eliminating the combination of these causes. For example, the gain in life expectancy from the elimination of major cardiovascular diseases, 6.7 years, is somewhat higher than the sum of the gains from the principal categories of this group of diseases, 5.5 years. The most impressive example of this "rule" is the inconsistency between the sum of the figures for the gains from the 15 leading causes of death, 12.3 years, and the gain from eliminating "Other than the 15 leading

[15]The reduction of the death rate from some cause may induce either an increase in related causes in tandem or postpone the rise in the death rates from other related causes to later ages. The argument can be made that the elimination of some causes increases the likelihood that some similar or associated causes will be reduced. This was found to be the case with various infectious diseases in the early part of the last century when the reduction of mortality from one infectious disease was accompanied by a reduction in other infectious diseases. Whether this phenomenon generally applies to the chronic diseases of later life is doubtful, even though etiological connections between these diseases have been established.

causes of death," 2.0 years, on the one hand, and the gain from eliminating all causes of death at once, on the other. The sum is 14.3 years while the years gained from eliminating all causes of death at once is infinite.

The interpretation of the differences in the probabilities of ever dying from a particular cause also has its difficulties. The probabilities may differ for two groups of persons even though their death rates are identical. The probabilities depend greatly on the mortality level of all the causes of death. Accordingly, they are an acceptable measure of the importance of each cause of death for a single group of persons at a given time but they are only roughly informative about the differences between groups and trends for the same group.

Model calculations allowing for the dependence of the causes of death could be made but they would have wide confidence intervals. If it is assumed that a particular cause of death is eliminated and the risks of dying from the various causes are dependent, there is no generally acceptable way of estimating their mortality rates and the survival functions (Moeschberger and Klein 1995). Consideration could be given to the relation of the causes with respect to their ages of incidence and their comorbidities. The studies of the multiple causes of death reported on death certificates in association with the underlying cause would provide much valuable information for identifying the competing causes of death that might be most affected by the elimination of a particular cause. Model calculations should be made to evaluate the effect of the order of elimination of the causes as well. A more realistic goal than eliminating various causes of death might be to "eliminate" them partially, say by 25–50%, and at the same time allow for the dependence of the causes on one another.

Multiple Increment-Decrement Tables

A life table that incorporates factors that augment the number of survivors as well as decrease their number is designated a multiple increment-decrement table. This extension of the multiple-decrement table also measures exits from the new status, such as exits from the labor force because of disability or retirement, exits from a nursing home because of death or transfer, and mothers' losses of a child. There are two general methods of constructing multiple increment-decrement tables: The occurrence-exposure method and the multistate method.

Occurrence-Exposure Method

In this application of the occurrence-exposure method the incremental and decremental factors are expressed separately. Hence, at least three factors are modifying the numbers in the basic cohort, namely, deaths, an incremental factor, and a decremental factor. For example, in the study of disability, the factors might be deaths, the incidence of disability, and recovery from disability. The probabilities

are applied directly to the survivor function to yield separate columns of changes in this function – deaths, persons becoming disabled, and persons recovering from disability. The rest of the life table can be derived by standard methods outlined earlier. The table can provide data on the expected years with disability, the expected years free of disability, and the chance of ever becoming disabled, as well as the chance of becoming disabled at each age, the number of disabled survivors at each age, and the disabled stationary population.

The table could theoretically be more complex and show changes in additional factors at each age, but then a better vehicle for measuring and analyzing the characteristic or event is the multistate table.

Multistate Methods

When the factors are numerous and many states and transfers are involved, the systematic management of the multitudinous calculations calls for a different method – the so-called multistate method. Multistate methods are now widely used to analyze many population characteristics and events where individuals can move from one state to another, can return to one's original state, and are subject to different mortality risks, such as a health condition, marital status, or labor force status.

The multistate life table is based on assumptions corresponding to a Markov process and so the equations for constructing such tables are consistent with a Markov design. In the present context a Markov model is specified when the probability that an individual leaves his/her state depends only on the state and a person's age. Multistate models can accommodate more than one active state in addition to death and the "occupants" of any state can move from that state to any other state or to the state of death. The multistate model is assumed to involve $k+1$ states, the $k+1$th state being an "absorbing" state, or death. Occupants of at least two of the k states must be able to move between one another, so that there are entrants into at least one state. For example, a multistate life table relating to disability would begin with "transition probabilities" between pairs of states identified by death, disability, and nondisability, and these are used to "move" the occupants of the states of disability and nondisability to the same (original) states or to the states of nondisability, disability, or death over the time interval represented by the transition probabilities.

Because of the number and complexity of the computations required in multistate modeling, the formulas are expressed in matrix notation. It may be recalled that matrix notation is a shorthand algebraic notation for symbolizing and manipulating data in blocks. Multistate methods were not possible and manageable until the advent of computers. Further discussion of multistate tables, describing their application to the joint measurement of mortality and morbidity, is given in Chap. 8.

Applications of the Life Table

Demographic Applications

Average Annual Change in Life Expectancy

Various measures have been used to calculate the average annual change in life expectancy. Inasmuch as life expectancy is already a ratio and varies within a narrow numerical range from zero to 100, it is logical to measure change simply as the average annual amount of change, obtained by dividing the total gain in life expectancy during a period by the number of years. Some would extend this measure, and compute an arithmetic average percent of change in life expectancy by dividing the average annual amount of change by the expectancy value at the beginning or middle of the period. Because this is a time series, however, it is more common and logical to calculate the average on the assumption of continuous compounding:

$$(1) = e_{x+t} \div e_x = e^{rt} \tag{4.65a}$$

$$(2) = (1) \text{ in } \ln \text{ form} \qquad \ln(e_{x+1} \div e_x) = rt \tag{4.65b}$$

$$(3) = \text{Dividing(2) by t} \qquad \ln(e_{x+1} \div e_x) \div t = r \tag{4.65c}$$

where x represents exact age $x, x + t$ represents the exact age t years later, and r represents the average annual rate of change. Given e_0 in 1949–1951 and 2000, 68.07 and 76.87, respectively,

$$76.87 \div 68.07 = e^{rt(50)}$$

$$1.129 = e^{50r}$$

$$0.1213 = 50r$$

$$r = .0024 \text{ or } 0.24\%$$

Hence, between 1950 and 2000 life expectancy increased 0.24% a year. The average annual amount of change is 0.176 and the corresponding arithmetic average annual percent of change is 0.24%[=0.176 ÷ {(76.87 + 68.07) ÷ 2}]. Note that dividing by the mean value in computing the annual average percent gives a close approximation of the exponential rate of change.

Increase in Life Expectancy as Share of Total Possible Gain

To determine the share of the total possible gain in life expectancy that was achieved during a given period, we assume that the maximum possible gain is the difference

between 100, a theoretical target, and life expectancy attained by the initial date of the period. Accordingly, from the 2000 and 1949–1951 life tables for the United States:

$$\begin{aligned} SC_0^n &= (e_0^{t+n} - e_0^t) \div (100 - e_0^t) \\ &= (76.87 - 68.07) \div (100 - 68.07) = 8.80 \div 31.93 = 0.276 \end{aligned} \tag{4.66}$$

For comparisons of changes in life expectancy as a share of the total possible gain for periods of different lengths, the average annual change in the measure must be computed. For this, we divide the total "percent" change by the number of years in the period, $0.276 \div 50 = .0055$, or we can follow the more refined formula of Arriaga (1984),

$$\begin{aligned} ASC_0^n &= [1 - (1 - SC_0^n)^{1/n}] \\ ASC_0^n &= [1 - (1 - .276)^{.02}] = (1 - .9936) = .0064 \end{aligned} \tag{4.67}$$

This result says that in the half century from 1950 to 2000 life expectancy at birth gained 0.6% of its theoretical target each year on the average.

Age-Bounded Life Expectancies

Estimates of average years of life lived in a specific age interval may be derived by "differencing" the T_x values bordering the age interval and dividing by the l_x at the start of the interval. For example, average years of life lived in the interval from age 65 to 80 is calculated by the following formula:

$${}_ie_x = (T_x - T_{x+i}) \div l_x \tag{4.68a}$$

$${}_{15}e_{65} = (T_{65} - T_{80}) \div l_{65} \tag{4.68b}$$

Then years lived of the total possible gain, in percent, is derived by

$$({}_ie_x \div i) * 100 \tag{4.69a}$$

$$({}_{15}e_{65} \div 15) * 100 \tag{4.69b}$$

For the United States in 2000, the calculations are:

$$(1{,}467{,}498 - 437{,}240) \div 82{,}131 = 12.54$$

$$(12.54 \div 15) * 100 = 84\%$$

In this 15-year age interval, 12 ½ years of life are lived by the cohort on the average; that is, the cohort lives 84%, or about five-sixths, of the total possible gain in the

interval. Average years lived in the interval 80 years and over is the same as life expectancy at age 80. Formulas 4.68 and 4.69 can be used to gauge improvement in mortality in specific age intervals between different populations.

Arriaga's (1984) formula can be applied to derive the average annual percent change in age-bounded life expectation as a share of total possible gain in the age group:

$$ {}_{i}\mathrm{ASC}_{x}^{n} = [1 - (1 - {}_{i}\,\mathrm{SC}_{x}^{n})^{1/n}] * 100 \tag{4.70} $$

where ${}_{i}\mathrm{SC}_{x}^{n}$ represents the change in age-bounded life expectation over a period (e.g., 1950–2000) in an age group (e.g., ages 65 to 80) as a share of its total possible gain (e.g., $15 -{}_{15}e_{65}$,)

$$ {}_{i}\mathrm{SC}_{x}^{n} = ({}_{i}e_{x}^{t+n} -{}_{i}\, e_{x}^{t}) \div (i -{}_{i}\, e_{x}^{t}) \tag{4.71} $$

For example, for the period 1950–2000 at ages 65–80, gain in age-bounded life expectancy as a share of its total possible gain increased 0.9% annually:

$$ \begin{aligned} {}_{15}\mathrm{ASC}_{65}^{50} &= [1 - (1 - .3724)^{.02}] * 100 = [1 - (.6276)^{.02}] * 100 \\ &= (1 - .9907) * 100 = .0093 * 100 = 0.93\% \end{aligned} $$

If we simply want the average annual percent change in age-bounded life expectation in this period for this age group, the formula would be

$$ \begin{aligned} {}_{i}\mathrm{SC}_{xp}^{n} &= ({}_{i}e_{x}^{t+n} -{}_{i}\, e_{x}^{t}) \div ({}_{i}e_{x}^{t}) \\ {}_{15}\mathrm{ASC}_{65p}^{50} &= [1 - (1 - .1318)^{.02}] * 100 = [1 - (.8682)^{.02}] * 100 \\ &= (1 - .9972) * 100 = .0028 * 100 = 0.28\% \end{aligned} \tag{4.72} $$

Life expectation in the 65-to-80-year age group increased in the second half of the twentieth century by 0.3% per year. The arithmetic average percent was also 0.3% (= .1318 ÷ 50*100).

Decomposition of Life Expectancy at Birth by Age and Other Categories

As we have seen, death rates for the various age groups do not change at the same rate. Hence, life expectancy and age-bounded life expectancies do not change at the same rate at the various ages. The shift in the age-pattern of improvements in death rates over the last century has led to a considerable interest in measuring the shares of the gain in life expectancy at birth attributable to gains at various ages and various components of the age groups, such as sex, race, and cause of death. The method is complex because direct, indirect, and interaction effects of the declines in death rates in the life table have to be taken into account.

Contribution of age-specific changes in e_x to change in life expectancy at birth. In this section a method, devised by Arriaga (1984), of measuring the contribution of change in life expectancy at each age to change in life expectancy at birth between two dates is discussed. At first glance, the change in the age-bounded life expectancies for the various age groups would seem to represent the elements of this decomposition, but we observe, for example, that the sum of the changes between the age-bounded life expectancies over several age groups for the period 1949–1951 and 2000 differs greatly from the difference between life expectancy at birth in these 2 years. For example, the first figure is 5.6 years for five broad age groups, and the second is 8.8 years. Yet a logical requirement of the decomposition is that the sum of the contributions to change over all ages should equal the difference in the change in life expectation at birth.

A large part of this discrepancy results from the fact that there are other effects on the change in mortality in an age group than the "direct" effect. In addition to the direct effect, there are an indirect effect resulting from the increased number of survivors when mortality falls and an interaction effect that takes into account the mortality change in other ages resulting from the indirect effect.[16]

The steps in the procedure are described in Appendix 4.2 and an illustration of the procedure is shown in Appendix Table 4.A3. The illustration measures the contribution of change in five broad age groups to the total change in life expectancy for white females between 1949–1951 and 2000 in the United States. The age groups selected are 0–14, 15–44, 45–64, 65–84, and 85–99 (or 85 and over). We need data for the following functions, e_x, l_x, and T_x, at the initial ages of the age groups enumerated. From the values of e_x, l_x, and T_x, we can derive all the intermediate values required to solve the several equations given by Arriaga. The first step is to compute the age-bounded life expectancies for the age groups listed. Then use them and the values of e_x, l_x, and T_x as required in the Arriaga equations. Separate equations are given for measuring the direct effect, the indirect effect, and the interaction effect.

Our illustration of the Arriaga decomposition method informs us that in the second half of the twentieth century, for white females in the United States, one-quarter of the increase in life expectancy at birth occurred among children, one-third occurred among youth and "working adults," and the remaining two-fifths occurred at the older ages (Table 4.10).

[16]We may define these components more fully by paraphrasing Arriaga (1984). The direct effect is the change in life expectancy within a particular age group as a result of the mortality change in that age group. The indirect effect is the number of years added to a given life expectancy because the mortality change within a specific age group produces a change in the number of survivors at the end of the age interval. The difference in those surviving before and after the mortality change is added to (or subtracted from) "years lived" as they pass through successive ages, assuming that mortality has not changed and remains at the same level at these later ages. Both direct and indirect effects are generated because mortality has changed only within the age group under study. An interaction effect arises because the direct and indirect effects take into account only the mortality change at each age group,

Table 4.10 Decomposition of the change in life expectancy at birth into the contribution of changes in age groups, for U.S. white females between 1949–1951 and 2000

Age group	Total contribution	Percent of total	Direct effect	Indirect effect	Interaction
0–14	1.964	24.7	0.334	1.477	0.153
15–44	0.885	11.1	0.194	0.587	0.104
45–64	1.836	23.1	0.542	1.008	0.286
65–84	2.883	36.3	1.797	0.798	0.288
85+	0.371	4.7	0.371	–	–
All ages	7.939	100.0	3.238	3.870	0.831

Source: Appendix Table 4.A3

Other types of decomposition. –The same decomposition methodology can be extended to encompass the contributions at each age of the major causes of death to the total change in life expectancy over a specified period. Further, the same methodology can be applied to measure the contribution of the difference in mortality of the races or the sexes at each age group to the total race or sex difference in life expectancy at birth or some age-bounded segment of life expectancy. (See, for example, U.S. NCHS/Kochanek et al. 1994.)

Survival Rates

Except for the use of life expectancy values in comparative analyses of mortality levels of different populations, the most common application of life tables is to calculate survival rates (ratios) for use in estimating or projecting deaths or population. Survival rates are used to measure survival propensities both during past periods and future periods. For example, in the conduct of national longitudinal panel surveys, from wave to wave there are dropouts due to death. It is desirable to evaluate the number of these dropouts independently. This can done by comparing the survival rates between waves for survey participants with survival rates from appropriate life tables.

Calculation. A survival rate is the proportion of a cohort of individuals aged x (or x *to* $x+n$) surviving to age $x+t$ (or $x+n$ to $x+n+t$). It is usually based on the ${}_nL_x$ function of an abridged life table and is symbolized by ${}_nL_{x+t} \div {}_nL_x$. Survival rates are defined in terms of two ages (or age groups), and hence two time references, the initial age(s) and date(s) and the terminal age(s) and date(s). Both age references are equally significant, although survival rates are commonly identified or labeled in terms of the initial ages, the initial date, and the time period. Survival rates usually express survival from a younger age to an older age, but they can be applied as reverse survival rates to restore deaths to the older of two populations.

In order to allow for mortality in a population during a period, survival rates are multiplied against the initial population. When survival rates are used to restore deaths to a population during a prior period in a reverse calculation, they are divided

into the terminal population. The formulas for applying forward and reverse survival rates for 5-year age groups over a 5-year time period are:

$$\text{Forwardsurvival} : {}_5P_x^t * {}_5s_x^5 = {}_5E_{x+5}^{t+5} \tag{4.73a}$$

$$\text{Reverse survival} : {}_5P_x^t \div {}_5s_{x-5}^5 = {}_5E_{x-5}^{t-5} \tag{4.73b}$$

where the elements represent the initial or terminal populations (${}_5P_x^t$), the 5-year forward survival rates (${}_5s_x^5$), the 5-year reverse survival rate (${}_5s_{x-5}^5$), the expected terminal population (${}_5E_{x+5}^{t+5}$), and the expected initial population (${}_5E_{x-5}^{t-5}$.). Note that forward survival and reverse survival are equally natural operations demographically. (The expression "revival rate" is not viewed as a felicitous substitute for the expression "reverse survival rate," and is not used.)

The most common type of survival rate employed in population studies is the rate for a 5-year age group for a 5-year time period. The formula is

$${}_5s_x^5 = {}_5L_{x+5} \div {}_5L_x \tag{4.74a}$$

Using data from the 2000 U.S. abridged life table (Table 4.2), the proportion of the population 55–59 years of age that will survive 5 years is calculated as follows:

$$\begin{aligned} {}_5s_{55}^5 &= {}_5L_{60} \div {}_5L_{55} \\ &= 424{,}879 \div 447{,}124 \\ &= .95025 \end{aligned} \tag{4.74b}$$

Note that the later ages are represented in the numerator and that the earlier ages are represented in the denominator. Also note that the time periods and the age spans have to be consistent with one another and with real-life age-time relations.

The proportion of the population 75 years and over that will live another 10 years is

$$\begin{aligned} {}_\infty s_{75}^{10} &= {}_\infty L_{85} \div {}_\infty L_{75} \text{ or } T_{85} \div T_{75} \\ &= 221{,}293 \div 726{,}571 = .30457 \end{aligned} \tag{4.75}$$

The survival rates for terminal age groups can be computed from the T_x values of the life table. Note again that the L_x or T_x value for the initial age group is shown in the denominator of the rate and the L_x or T_x value for the terminal age group is shown in the numerator of the rate. A complete life table readily permits calculation of survival rates for single ages for 1 year or any other age-time period. A 1-year survival rate for a single age is represented by

$$s_x = L_{x+1} \div L_x \tag{4.76}$$

The proportion of the population 65 years of age on a given date that will survive 1 year, on the basis of the 2000 complete U.S. life table (Table 4.1), is

$$s_{65} = L_{66} \div L_{65} = 80{,}109 \div 81{,}472 = .98327$$

The proportion of 75 year-olds that are expected to live another 10 years is

$$s_{75}^{10} = L_{85} \div L_{75} = 33{,}268 \div 63{,}024 = .52786$$

Survival rates involving birthdays are computed using the l_x values. The proportion of newborn babies that will reach their fifth birthday (Table 4.1) is

$$l_5 \div l_0 = 99197 \div 100{,}000 = .99197$$

If survival from birth to a given age interval is wanted, then the form of the survival rate is ${}_nL_x \div nl_0$, or for a 5-year age group, ${}_5L_0 \div 5l_0$. Here n or 5 cohorts of 100,000 babies are subject to the risk of death until the end of the 5-year period of their birth. Survival from birth to the age interval 75–79 is measured by

$$289{,}331 \div 500{,}000 = .5787$$

That is, according to the mortality levels of 2000, more than half of the newborn babies are expected to survive to ages 75–79 years.

Use of life table survival rates. In the use of life table survival rates in population studies, decisions have to be made regarding the selection of the life table and the life table survival rates most appropriate in a particular case. Where a life table is not available for the particular year or period, but for prior or subsequent years or for the initial or terminal years of the period (rather than the midyear of the period), special survival rates may have to be computed on the basis of the available life tables. For example, life tables may be available for the census years, but we may need survival rates for the intercensal period. The desired rates may be derived by interpolating the survival rates from each of the two census-year tables linearly to the middle of the intercensal period. This assumes that mortality changes occurred evenly over these years. If mortality changes during the intercensal period are negligible, the interpolation may be unnecessary.

In other cases the application may require adjustment for the geographic area or for the racial or ethnic group, urban-rural residence, marital status, educational status, other socioeconomic characteristic, or health status. Examination of differences in observed age-specific death rates between the particular subpopulation of interest and the population for which life tables are available, when such comparisons of death rates can be made, should suggest whether some adjustment in the basic survival rates is needed for developing survival rates for the subpopulation. In some cases, as for the many of the states of the United States, national life tables for individual sex-race groups are adequate to represent mortality in the state.

Survival rates for the general population can be adjusted directly for geographic or socioeconomic variations by use of observed central death rates for the various categories, as follows:

1. Take the complement of the general survival rate;

2. Calculate an adjustment factor for the particular geographic, socioeconomic, or health category in terms of central death rates;
3. Multiply the adjustment factor by the complement of the general survival rate in (1); and
4. Take the complement of the result in (3).

For example, let us derive a 5-year survival rate for the married male population 55–59 years of age of the United States in 2000. Assuming ${}_5s_x^5$ is a 5-year survival rate with initial ages x to $x + 4$ and ${}_{10}M_x$ is a central death rate for ages x to $x + 9$:

Instructions	Calculations
(1) = $1 - {}_5s_{55}^5$ (= a 5-year mortality rate)	(1) = 1−.9502 = .0498
(2) = $({}_{10}M_{55}^m)$ *for the married population* $({}_{10}M_{55})$ for the general population	(2) = 0.773
(3) = (1)*(2)	(3) = .0385
(4) = 1 − (3) = ${}_5s_{55}^5$ for the married male population	(4) = .9615

Source note: NCHS (2002a, b)

We see that the general male survival rate has been increased to allow for the much lower mortality of the married male population.

In some cases official life tables are available for a country but they are based on seriously incomplete or erroneous statistics on deaths. In these cases and in others where life tables are not available, one has to decide whether to construct a new life table, "borrow" a life table from another country, or adopt a model life table from a set of available model tables. Commonly, the areas that lack life tables do not have adequate death statistics for constructing such tables. In some cases one may be able to construct a life table by use of the existing statistics on deaths and population and "stable population" estimation techniques. The selection of a model life table and the use of stable population techniques are discussed in Chap. 12.

The estimates of survivors will differ depending on the age detail employed in computing the survival rates. With few exceptions, 5-year survival rates are sufficiently detailed to take account of the significant age variations in the estimates of survivors. For some purposes, calculations for 10-year intervals will be adequate. The use of survival rates with a given age interval implies that the life table population within this interval has the same age distribution as the actual population. To avoid dependence on this assumption, survival rates for very broad age spans should not be employed, except for rough calculations. Accordingly, survival rates for the terminal open-ended age interval should relate to a relatively limited age span containing only a small percentage of the actual population. The terminal group 65 and over is to be avoided in survival calculations except for very "young" populations (e.g., Nicaragua, Pakistan, Syria). For countries with relatively "old" populations (e.g., France, Sweden, United States), even 75 and over is an unsatisfactory terminal group and should be subdivided into component age groups, e.g., 75–84, and 85 and over.

While life table survival rates represent the ratio of l_x or L_x values to one another, the actual level of survival rates is affected only by the mortality rates (q'_xs) in the age range to which the survival rates apply. Hence, life table survival rates for various populations are properly combined on the basis of the population distribution at these ages. For example, 5-year survival rates with initial ages 50–54, for the white and nonwhite populations of the United States, could be combined on the basis of the distribution of the population by color at ages 50–54, in order to obtain a survival rate for all races combined. In general, to combine survival rates from different life tables, they should be weighted in proportion to the observed population in the ages involved. When entire life tables for males and females of the same population are being combined for analyzing differences between the sexes, however, the weighting used in combining them should be on the basis of the sex distribution of births.

National census survival rates. Another means of allowing for mortality is the use of national census survival rates (ratios). They employ life table concepts but do not involve use of actual life tables. National census survival rates can be employed in measuring and evaluating the level of mortality in a country and can substitute for life table survival rates, especially for countries lacking adequate vital statistics. They may be an important basis for selecting a model life table or model stable population (see Chap. 12). (They are also useful in evaluating the quality of the census data on the sex and age distribution of the national population.)

National census survival rates represent the ratio of the population in a given age group in one census to the population in the same birth cohort at the previous census. Census survival rates for the children born in the decade link births during the intercensal period and children under age 10 in the second census. While national census survival rates essentially measure mortality, unlike life table survival rates they are directly affected by the relative accuracy of the two census counts employed in deriving them. Underlying their use as intercensal survival rates are certain basic assumptions. These assumptions are that the population is a closed one (i.e., no net migration affects it during this intercensal period), that there has been no abnormal influence on mortality (e.g., war or famine), and that the net census undercount patterns of the population under study are similar to those of the national population from which the census survival rates were derived.

Census survival rates may be rather irregular from age to age, often fluctuating up and down throughout the age scale and exceeding unity in some ages. The more inconsistent the data in the two censuses with the actual mortality changes, the more erratic the census survival rates tend to be. The requirement that the population be a closed one is rarely if ever completely met, and it is desirable, therefore, to adjust the census data for intercensal net migration before computing the survival rates if the necessary data are at hand. For a country with only a small net immigration, this can sometimes be achieved indirectly by basing the census survival rates on the native population of the country.

Table 4.11 Life endurancy at three probability levels, by sex, for the United States: 1900–2004

	50%		10%		1%	
Year	Male	Female	Male	Female	Male	Female
1900	55	58	81	82	91	92
1950	70	76	86	89	94	97
1980	74	81	89	94	98	102
2000	78	83	91	95	100	ca.103
2004	79	84	93	97	100	>103

Source: Various life table publications of the National Center for Health Statistics. 1900: U.S. Office of the Chief Actuary, SSA (2005)
Age in years corresponding to indicated survival rates

Life Endurancy

Life endurancy is another measure of survival probability. More specifically, it is the age to which specified proportions of a cohort of births in a life table survive. At a survival probability of 0.5, life endurancy corresponds to the age to which half of the original cohort survives, or the median age of life table deaths, and at a survival probability of.01, life endurancy corresponds to the age to which 99% of the original cohort survives. Life endurancy is a measure of special interest in studies of extreme human longevity. Consider the data in Table 4.11, which reflect a dramatic increase between 1900 and 2000 in the age to which 50%, 10%, and 1% of newborn children survive.

Relation of Reduction in Death Rates and Increase in Survival Rates/Life Expectation

Death rates and survival rates "move" at different speeds. This becomes apparent when we consider a simple example. If a mortality rate of .002 is reduced by 50%, the resulting rate is .001. The corresponding survival rate increases from .998 to .999, or by 0.1%. The lower the death rate the wider the gap. This relation is illustrated more extensively in Table 4.12 with historical data on changes in 5-year survival rates and their complements, 5-year mortality rates, for the periods 1900–1950 and 1950–2000.

Similarly, a considerable decline in age-specific death rates produces only a small increase in life expectancy at low levels of mortality. For this comparison, the age-specific death rates have been summarized in the form of age-adjusted death rates. Table 4.13 compares, for the period 1900–2000, the percentage reductions in age-adjusted death rates (standard population = 2000 for 1940–2000, and standard population = 1940 for 1900–1940) with the percentage increases in life expectancy at birth, for each decade. As expected, for every decade in this period, the percent increase in life expectancy at birth is smaller than the associated decrease in death rates. However, we see only a rough indication of an inverse correlation between

Table 4.12 Percent change in 5-year mortality rates and 5-year survival rates: United States, 1900–1950 and 1950–2000

	Percent reduction in 5-year mortality rates[a]		Percent increase in 5-year survival rates	
Terminal age group (years)	1900–1950[b]	1950–2000	1900–1950[b]	1950–2000
0–4	83.4	76.6	14.8	2.5
5–9	93.7	76.2	5.1	0.4
10–14	97.4	72.6	1.3	0.2
15–19	97.3	47.4	1.3	0.2
20–24	96.3	34.3	2.1	0.2
25–29	95.6	35.7	2.8	0.3
30–34	95.1	40.7	3.1	0.4
35–39	94.5	42.9	3.5	0.5
40–44	93.9	44.8	3.5	0.8
45–49	92.9	47.1	3.5	1.3
50–54	91.6	48.7	3.5	2.1
55–59	89.2	48.9	4.0	3.3
60–64	85.7	46.0	4.7	4.7
65–69	80.5	42.4	6.6	6.4
70–74	73.4	40.1	8.9	9.4
75–79	63.3	38.8	12.6	14.8
80–84	49.8	35.0	20.0	22.4
85–89	35.8	27.6	28.4	30.8
90–94	23.2	21.2	35.6	43.3
95–99	12.9	15.7	43.8	63.0
100–104	5.3	13.1	76.6	104.9
Selected differences				
(25–29)−(75–79)	+32.3	−3.1	−9.8	−14.5
(30–34)−(80–84)	+45.3	+5.7	−16.9	−22.0

Source: Based on life tables for 1900–1902, 1949–1951, and 2000, published by the U.S. National Center for Health Statistics or its predecessor agencies

[a]Five-year mortality rates are the complements of 5-year survival rates

[b]Life table for 1900–1902 relates to white males in the Original Death Registration States

decennial changes in age-adjusted death rates and life expectation over this 100-year period. This seeming irregularity can be explained in part by the fact that the actual changes in death rates have not been uniform over the age scale. When the figures are grouped in 30-year intervals, the indications of this pattern are stronger. At $e_0 = 71$ in 1970, a decline in the age-adjusted death rate of 29% between 1970 and 2000 corresponds to a rise of only 9% in e_0. In the previous 30-year period, with an initial life expectation of 63 years, the changes were 32% and 13%. The recorded relation between the relative change in a schedule of age-specific death rates and the relative change in the corresponding life expectancy at birth is of considerable interest because of its implications for the change required in age-specific death rates in the future to effect a specified change in life expectancy.

As we saw earlier, Keyfitz has expressed such a relation mathematically under the assumption that there is a proportional change in mortality at all ages, that is,

Table 4.13 Comparison of percent declines in age-adjusted death rates with percent increases in life expectation at birth: United States, 1900–2000, by decades

Year	Age-adjusted death rate[b]	Expectation of life at birth (e_0)	Percent change[a]	
			Death rate	e_0
U.S. 2000 population as standard				
2000	869.0	76.9	−7.4	+2.0
1990	938.7	75.4	−9.7	+2.3
1980	1039.1	73.7	−15.0	+4.1
1970	1222.6	70.8	−8.5	+1.6
1960	1339.2	69.7	−7.4	+2.2
1950	1446.0	68.2	−19.0	+8.4
1940	1785.0	62.9	NA	NA
1970–2000	X	X	−28.9	+8.6
1940–1970	X	X	−31.5	+12.6
U.S. 1940 population as standard				
1940	10.8	62.9	−13.6	+5.4
1930	12.5	59.7	−12.0	+10.4
1920[c]	14.2	54.1	−10.1	+8.2
1910[c]	15.8	50.0	−11.2	+5.7
1900[c]	17.8	47.3	X	X
1900–1930	X	X	−29.8	+26.2
1910–1940	X	X	−31.6	+25.8

Source: Based on U.S. National Center for Health Statistics 2002a, b; and *Vital Statistics of the United States, 1950*, Vol.1, Part 1, 1954

X not applicable

[a]For prior decade

[b]Per 100,000 population for 1940–2000; per 1000 population for 1900–1940

[c]Death registration states only

when the age-specific death rates are increased or reduced by the same proportion at every age. This assumption is made for mathematical convenience but very roughly it can be taken to describe the experience of many countries during many periods in the past. Keyfitz' equation relating change in life expectation at birth and the decline in age-specific death rates is

$$\Delta e_0 = -ke_oH \tag{4.77}$$

where k represents the proportional reduction in age-specific death rates and H is calculated by

$$H = -\sum_{0\infty}(L_x * \mathrm{In}(l_x/l_0)) \div T_0 \tag{4.78}$$

H is called the entropy parameter, where entropy refers to the "information content of a distribution" (Schoen 1987). H declines as e_0 increases.

Proportional declines in mortality produce smaller increases in longevity at higher values of e_0. The calculation of H is illustrated on the basis of the abridged life table for 2000 in Table 4.A4. The approximate result is.1298. Hence, for 2000,

when life expectancy was 76.87, a 50% decline in mortality at every age would result in an increase in life expectancy of 6.5% ($= -(-0.50)(76.87)*(0.1298) =$ 4.99 years). For 1949–1951, when $e_0 = 68.07$ and $H = .1792$, a 50% decline in mortality would result in an increase of 6.1 years, or 9.0%, in life expectancy.

In sum, life expectancy and death rates change at very different rates and the differences between them shift with different levels of life expectancy (Keyfitz 1977; Vaupel 1986). At the higher levels of life expectancy, the required declines in age-specific death rates are much greater than the associated increases in life expectancy and, in general, the higher the levels of life expectancy, the larger must the associated declines in age-specific death rates be to effect a given increase in life expectancy.[17]

The Longevity Dividend

To continue this discussion of demographic applications of the life table, I define an alternative measure of life expectancy, namely total or complete life expectancy. Total life expectancy is an indicator of the additional years of life gained by the experience of having survived to later ages that I call the longevity dividend.

Total life expectancy. Total life expectancy is the sum of the attained age of survivors and life expectation at that attained age. For example, total life expectation at age 85 in the United States in the year 2000 was 91.3 years (=85.0 + 6.3). It is the total number of years, on the average, that the cohort is expected to live from birth, given that the cohort will survive to age 85. This measure highlights the fact that survival to each later age eliminates the risk of dying at any earlier age and thereby grants the survivor additional years of total life. Life expectancy steadily declines over the age scale (with the exception of some early ages at very high infant mortality levels) but survivors to higher ages outlive the expectation of life indicated for them at birth. The additional years gained by survivors over their expectation at birth is part of the longevity dividend they win for having survived so far. Consider this survival scenario. According to the life table for 2000, life expectancy at birth is 77 years, but those persons who survive to age 77 years have an average remaining life of 10 years, or a total life expectancy of 87 years (Table 4.1). If they are successful in reaching age 87, they are expected to live to age 93, or 6 more years. Thus, 16 additional years of life will have been gained. Those who

[17]The Keyfitz' formula has its limitations. Applications to actual life tables do not produce the expected results, especially with very large reductions of death rates at low levels of mortality. Furthermore, H can exceed 1.00. At the low levels of mortality observed in many countries of the world today (e.g., United States 2000), empirical calculations show that a 50% uniform reduction in age-specific death rates would bring about a 11% increase in life expectancy at birthto 85.2 years (cf. 6.5% with $H = .1298$), and an 80% uniform reduction in age-specific death rates would bring about a 22% increase (to 94 years) in life expectancy at birth (cf. 10.4% with $H = .1298$). With the Keyfitz formula, it would take a 230% reduction in the age-specific death rates of 2000 to achieve a life expectation of 100 and this is patently implausible.

survive to age 93 have a life expectation of 4 years, and if they survive to age 97, life expectancy is 3.1 years. Newly minted centenarians have 2.6 years to go. There is always a bonus for having survived to a later age.

Contribution of secular declines in mortality. The bonus for survival is greater than so far indicated, however – usually much greater – since individuals who survive to later ages benefit also from the secular improvements in mortality that occur with the passage of the years. The total longevity dividend represents the combination of the gain achieved by survival to later ages of life and the gain through reductions in death rates over time. The total longevity dividend and its components is suggested by a comparison of period and generation life tables for the same year, as shown in Table 4.7, Figs. 4.3 and 4.4, and data not shown. For example, a girl born in 1920 who survives to age 56, her life expectancy at birth, as shown in the period life table for 1920, has won some 26 additional years over her original birth expectation. Of this considerable gain, 18 years are a result of her success in surviving to age 56 and 8 years are a result of the improvements in mortality over her lifetime [$82.17 - 56.27 = 18.17 + (25.90 - 18.17) = 25.90$]. The first component of the gain is life expectation at age 56 in the complete period life table for 1920 and the second component is the excess of life expectation at age 56 in the complete cohort life table for 1920 over life expectation at age 56 in the complete period life table for 1920 (U.S. Office of the Actuary 2005).

Here is another scenario: A woman who was born in 1920 and lives for 85 years until her death in 2005 may be thought of as achieving this age by gaining 18 years over her life expectancy at birth (56 years) and by gaining the remaining 11 years (for a total of 29 years) by a combination of secular improvements in mortality (about 8 years up to age 56) and her further survival to later ages (about 3 years).

Rectangularization of the Survival Curve

The various measures of the shape of the distribution of life table deaths can be invoked to describe the changing shape of the survival curve. As mortality rates have fallen over the last century, the life table survival curve has tended to fall more and more sharply after the modal age of deaths. The rectangularization of the survival curve is virtually equivalent demographically to the age-compression of mortality. Hence, it can be measured in the same general ways as discussed earlier for measuring the shape of the distribution of life table deaths. These include changes in the interquartile range of deaths, the standard deviation of deaths, Keyfitz' H factor, the Gini ratio, and the slope of the survival curve at the adult ages.

As the mean, median, and modal ages at death have reached the advanced ages, the further rectangularization of the survival curve has been questioned. It is pertinent to raise the question whether this process has been slowing or reversing itself because of its considerable implications for the compression of mortality and morbidity. If the analysis is confined to adults only, the process now appears to be showing little movement, although continuing. The possibility that it has dissipated

is suggested by the facts that the survival curve includes more and more survivors beyond age 100, the rate of increase in death rates at the highest ages clearly diminishes with increasing age, and the maximum age of deaths has been rising historically in several countries of Western Europe.

A measure of the rectangularization of the survival curve that allows specifically for the change in death rates at the most extreme ages may throw further light on the process. For this purpose, another measure, the trend in the ratio of life expectation at birth to total life expectation at age 100 (per 100), may be informative:

$$e_0 \div (100 + e_{100}) * 100 \tag{4.79}$$

Values for this measure in various years for the United States are as follows:

Year	Ratio (per 100)	Increase
1900	48.5	–
1950	66.8	18.3
1980	71.9	5.1
2000	74.9	3.0

Source: Based on NCHS decennial life tables for 1900,1950, and 1980 and the NCHS annual life table for 2000

Hence, even with a specific allowance for changes in death rates over age 100, the survival curve has been becoming more rectangular and reflects continued compression of mortality. Furthermore, using the projected life tables of the Actuary's Office, U.S. Social Security Administration, total life expectancy at age 100 would have to increase from its present level of 103 to about 107 for the value of this measure to begin to turn around.

We can add to the analytic utility of this measure by confining the numerator to the adult ages, say 25 years and over:

Year	Ratio (per 100)	Increase
1900	38.5	–
1950	45.7	7.2
1980	49.5	3.8
2000	51.8	2.3

Source: Based on NCHS decennial life tables for 1900, 1950, and 1980 and the annual life table for 2000

These figures also suggest that the rectangularization of the survival curve and the compression of mortality have continued through the last decades of the last century but that they are slowing down.

Biological Rate of Aging

The biological rate of aging is a concept for measuring the age pattern of the force of mortality. It is more widely used in interspecies comparisons than in the analysis of human mortality per se. A leading version of this measure – the one that I evaluate with U.S. historical data in Table 4.14 – defines it as the number of ages that are required for the mortality rate to double with increasing age. Inasmuch as this measure is intended to reflect endogenous/intrinsic mortality and mortality rates at the very advanced ages are known to rise at a diminishing rate, I have based my calculations on rates in the age range from 30 to 80 years. This range excludes much of the extrinsic/exogenous mortality experienced by humans and excludes the advanced ages where mortality rates decelerate. Biologists maintain that the rate of aging is distinctive for each species and that it does not vary over the age range of a species and for subgroups in a given species. According to Austad (1997), biologists believe that the mortality rate for humans doubles every eight ages.

As may be recalled (Chap. 3), Gompertz, the British actuary of the nineteenth century, posited an age model of human mortality that assumed a fixed rate of increase in mortality rates over much of the adult age range. We can express this model simply in the following equations:

$$\mu_x = ae^{bx} \tag{4.80a}$$

$$\text{In}\mu_x = \text{In } a + bx \tag{4.80b}$$

$$Y_x = A + bx \tag{4.80c}$$

where μ_x and $ln\,\mu_x (=Y_x)$ refer to the terminal mortality rate, a and $ln\,a (=A)$ refer to the initial mortality rate, b is a constant representing the increase in mortality, and x is the age indicator.

Analysis of mortality rates for the United States over the last century suggest that the biological rate of aging has been changing over this period, specifically that the rate of aging has been speeding up (Table 4.14). In 1900 it took about $12\frac{1}{2}$ ages for the mortality rate to double; now it takes between $8\frac{1}{2}$ and 9.0 ages. A major decline in aging time occurred during the first half of the last century, and the number of ages has stabilized at just above 8.5 since then. The rate of aging also varies over the age range, differs somewhat for males and females, and varies for period and cohort data. For all observation years from 1900 to 2003, the rate of aging was much slower at the younger ages than at the older ages, particularly in the early part of the last century. The male and female rates tend to differ in only small degree, with the male rates being lower in the early years of the century and the female rates being lower

Table 4.14 Ages required for the age-specific mortality rate to double over the age range 30–80 years, for the United States: 1900–2003

Estimates based on NCHS data									
	Average years			First doubling[a]			Last doubling[a]		
Year	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female
1900–1902[b]	NA	12.3	12.6	NA	20.9	22.5	NA	8.4	8.5
1929–1931[c]	NA	10.0	10.1	NA	13.0	16.0	NA	7.5	7.8
1949–1951	8.6	8.8	8.4	9.7	9.7	9.7	8.3	9.1	7.5
1979–1981	8.8	9.0	8.0	12.5	12.5	9.1	8.5	9.1	7.1
2003	8.6	8.9	8.0	10.0	10.9	8.3	7.6	7.7	6.9
Estimates based on SSA period data									
	Average years			First doubling			Last doubling		
Year	Male	Female		Male	Female		Male	Female	
1900	12.4	12.7		21.0	22.3		7.6	7.5	
1930	10.7	10.7		14.4	16.8		8.7	8.1	
1950	8.8	8.4		9.8	9.5		9.2	7.7	
1980	8.9	8.0		12.5	9.2		9.1	7.2	
2000	8.9	8.1		11.6	9.1		7.8	7.2	
Estimates based on SSA cohort data									
Birth cohort	Average years			First doubling			Last doubling		
	Male	Female		Male	Female		Male	Female	
1900	11.8	13.2		18.5	25.7		10.7	8.7	
1930	9.5	9.1		8.7	8.6		8.6	7.8	
1950	10.1	10.1		12.6	11.1		8.6	8.4	

Source: Based on various life tables of the U.S. National Center for Health Statistics and the U.S. Social Security Administration, Office of the Chief Actuary (2002)

Note: *NCHS* National Center for Health Statistics, *SSA* Actuary's Office, Social Security Administration

NA not available

[a]Earliest doubling years beginning at age 30 and latest doubling years ending at age 80

[b]Whites in Original Death Registration States

[c]Whites in continental United States

in the later years of the century. Cohort figures reflect changing mortality experience over time and so, for example, the birth cohorts of 1950 indicate a much slower rate of aging than the period figures for that year.

Some Socioeconomic and Public Health Applications

The application of the standard life table has been extended to many domains and used to deal with a variety of socioeconomic and public health issues. I consider a few of them here briefly. Items treated are the joint survival of husband and wife, the

relation between the age of widowhood and the age of inheritance, and measurement of years until death. The discussion of life table applications is continued in Chap. 8 with notes on the measurement of disability in relation to the period of working life and the period of total life, and the risk of entering a nursing home.

Survival of Husband and Wife

In family studies, a question is often asked about the probability that a newly married couple will live together as a couple for some specified number of years. Suppose we ask the question, what is the probability that a newly married couple, with the groom at age 25 and the bride at age 23, will remain married long enough to celebrate their 50th wedding anniversary.

Calculations to answer the question have to be based on a number of assumptions. These assumptions relate to the probability of the dissolution of the marriage by divorce, differences in the probabilities of dying according to marital status, and the possible changes in mortality rates during the lifetime of the couple. Let us simplify our task for this discussion by excluding the first possibility and assuming further that there is no difference in mortality rates between married persons and the general population and that mortality rates do not change in the lifetime of the couple. All of these assumptions are seriously contrary to fact, but they serve as a place to begin.

The basic calculation simply takes the product of the individual probabilities of surviving. Here are the figures from the U.S. life tables for males and females for 2000:

Male	$L_{75}/L_{25} = 55{,}990/97{,}645 = .57340$
Female	$L_{73}/L_{23} = 74{,}002/98{,}745 = .74943$
Joint probability	$.57340 * .74943 = .42972$

That is, the couple has a 43% chance of living together for 50 years under the assumptions made.

With a 50% or so probability of divorce from a first marriage in the United States today, the above answer vastly overstates the real probability of joint survival in the first marriage for U.S. newly married couples. The divorce factor greatly overshadows the survival advantage of married persons over each other marital group, which is on the order of 25–40%. Since a married couple statistically has a high chance of getting divorced, to account for their joint survival we should use a life table for married persons, a so-called nuptiality table that takes account of both divorce and death (specifically, a marriage-dissolution table). Appropriate data are currently available to construct nuptiality life tables for the United States and several other countries but few tables of this type have been prepared. Marriage dissolution tables can be constructed as multiple-decrement tables or as multistate life tables, and would take joint account of death and divorce for married persons. A simpler, less realistic, handling of the question is to adjust any answer based on a general life table for the fact that the persons are married.

The question could be extended to ask about survival of either spouse for 50 years. So, what is the chance of the man dying and the women surviving for the 50-year period? An answer, without complicating assumptions, is:

$$.74943 * (1 - .57340) = .74943 * .42660 = .3197$$

The chance of the women dying and the man surviving is:

$$.57340 * (1 - .74946) = .57340 * .25057 = .1437$$

These calculations disregard the chances that mortality rates are different for a widow(er) and a married person and that they are affected by bereavement. Continuing with our simplistic construction of the question, we can now determine the chance of either the husband or wife dying before the 50th wedding anniversary. This is measured as the sum of the probabilities of either one dying and the other one living and the opposite:

$$.3197 + .1437 = .4634$$

Period of Widowhood and Age at Inheritance

Here we use the life table to ascertain the period of widowhood of new widows or widowers and the age of family beneficiaries of deceased parents. As before, the calculations are made on the assumptions that the mortality levels of persons do not differ according to marital status and that there is no effect of bereavement on the mortality of the surviving spouse. We also assume that the surviving spouse does not remarry.

If a husband dies at age 75, the new widow is likely to be about age 73. She has a very low probability of remarrying and is likely to remain a widow for about 13 years. If, on the other hand, the wife precedes the husband in death and dies at age 73, the widower is likely to be about age 75 and, barring remarriage, can be expected to live about 10 years as a widower. In each case the surviving spouse would live to about age 85–86. These values are taken from unabridged life tables for U.S. males and females in 2000.

A half century ago these numbers would have been substantially lower. Marriage tended to occur at an earlier age, and survival of both spouses within marriage and survival of the widowed spouse was much lower in the 1950s than today. When parents in the United States and other low mortality countries bequeath their assets to their children currently on the occasion of their death, the children receive these assets at a much later age than they did formerly. The children are likely to be about 55–60 years of age, when they are well along in their adult life, have raised their own children, and have become adjusted to a particular standard of living. At this point in their family history any new assets would not be available to affect financial decisions relating to childbearing, childrearing, housing, and career choices for the children.

Years Until Death

Another way to measure progress (or lack of it) in increasing longevity is to track the age at which persons have a specified life expectation, such as 10 or 15 years, according to current life tables. Such figures can be used not only to measure general mortality progress but also to compare the differences between the sexes, the races, or other groups of interest. The ages corresponding to life expectancies of 10 and 15 years from 1950 to 2000 at decennial intervals in the United States are as follows:

	Age	
Year	$e_x = 10$ years	$e_x = 15$years
1950	71.7	63.1
1960	72.5	64.0
1970	73.7	65.0
1980	75.9	67.3
1990	76.9	68.4
2000	77.3	69.1
Increase, 1950–2000	5.6	6.0

Source: Based on NCHS U.S. decennial life tables and annual NCHS life table for 2000

In 1950 survivors had an average of 10 years of life remaining at age 71.7 years but by 2000 this age had risen to 77.3 years. The figures reflect a gain of 5.6 years in this half century. The ages corresponding to 15 years of life expectancy run lower, of course, but describe a similar trend. Data of this kind can be invoked to set guidelines for selecting the ages at which full benefits might be mandated under Social Security. In this way account can automatically be taken of changes in the retirement years over which benefits are received.

With such a series of ages, a next step would be to determine the share of the total population that fell above each age; e.g., the share of the population in 2000 that fell above age 69.1 years. The goal would be to see if it is practical to stabilize the burden of the Social Security program at a nearly fixed share of the total population.

Other Health Applications

Multiple-decrement and multistate life tables can be constructed that measure numerous aspects of health and mortality jointly. As mentioned, I illustrate some of these other applications in Chap. 8 with tables that measure years of disabled life among the general population and the working population and the chance of ever becoming disabled, years of nursing-home life and the chance of ever entering a nursing home, and similar health phenomena.

Finally, the life table is an important tool in animal research. It can be applied to animal lives as well as to human lives and hence is a quite valuable tool for making comparisons of animal and human mortality experience in longevity research.

Appendix 4.1 Some Life Table Relations in Continuous Notation

Demographic analysis with life tables is commonly expressed in continuous notation, that is, the language of calculus. This formulation uses an integral sign ($\int$) rather than a summation sign ($\sum$) to represent sums. Age is shown in parentheses, where it represents the value at which the function is to be evaluated. For example, in continuous notation the number of survivors at age x is $l(x)$ but in discrete notation it is given as l_x.

In theory the basic function of a life table is the hazard rate. The hazard rate is the instantaneous probability of dying, representing the chance of dying in an infinitesimally short period. It is also called the force of mortality. If $l(t)$ is the number of survivors to time t, then the hazard rate, $h(t)$, may be given as the limit of $[\mathrm{l(t)} - \mathrm{l(t} + \Delta\mathrm{t})] \div \mathrm{l(t)}\Delta\mathrm{t}$ as Δt tends to zero. More commonly $h(t)$ is formulated in terms of $S(t)$ rather than $l(t)$ since the concept does not apply only to life table analysis. Alternatively, the hazard rate may be represented by

$$\mathrm{h(x)} = -\frac{\mathrm{l'(x)}}{\mathrm{l(x)}} = -\frac{\mathrm{dIn\ l(x)}}{\mathrm{dx}} \tag{4.81}$$

The force of mortality at age x (μ_x), that is, the age-specific mortality rate in the life table (q_x), may be approximated by

$$\mu_\mathrm{x} = -\mathrm{In}(\mathrm{l_{x+1}} \div \mathrm{l_x}) \tag{4.82}$$

The central death rate for an age interval, m_x, converges to the hazard rate, $h(t)$, and the force of mortality, μ_x (or q_x), as the age interval becomes smaller and smaller.

The probability of surviving to age x (i.e., $l_x \div l_0$) is called the survival distribution function. It is equal to the quotient of the probability density function, or the "share" of deaths at age x, and the hazard rate, or the instantaneous risk of dying at age x:

$$\mathrm{S(x)} = \mathrm{f(x)} \div \mathrm{h(x)} \tag{4.83a}$$

where $S(x)$ is the survival distribution function, $f(x)$ is the probability density function (pdf), and $h(x)$ is the hazard function. The relations in this formula can be illustrated with data from the official life table for the United States in 2000. The value for $f(x)$ at age 75 is $d_{75} \div_\infty d_0$, or .02420 (=2420/100,000), and the value of $h(x)$ at age 75 approximated by q_{75}, is .037675. Then, the survival distribution value, or the cumulative proability of dying at age 75 or later, is obtained from Eq. 4.83a as .02420 ÷ .037675 = .64234. This result approximates the entry in the life table for $l_{75} \div l_0$(= .64244). These relations are also expressed in terms of the mortality distribution function, $F(x)$, the complement of the survival distribution function:

$$1 - \mathrm{F(x)} = \mathrm{S(x)} = \mathrm{f(x)} \div \mathrm{h(x)} \tag{4.83b}$$

Solving the expression for $h(x)$, the hazard rate, we have:

$$\mathrm{h(x) = f(x) \div S(x)} \tag{4.84}$$

Since $f(x) = -dS(x)$, the hazard rate may be described as the relative rate of decline of the survival function:

$$\begin{aligned}\mathrm{h(x)} &= \frac{-\mathrm{dS(x)}}{\mathrm{S(x)dx}} \\ &= (.64244 - .61823) \div .64244 = .02421 \div .64244 \\ &= 0.037684 \end{aligned} \tag{4.85}$$

The U.S. 2000 life table shows 0.037675 for the mortality rate at age 75.

The number of person-years lived in a single-year-of-age interval is given by

$$\mathrm{L(x)} = \int_{\mathrm{a=x}}^{\mathrm{x+1}} \mathrm{l(a)da} \tag{4.86a}$$

For an interval of n years the formula is

$$\mathrm{L(x, x + n)} = \int_{\mathrm{a=x}}^{\mathrm{x+n}} \mathrm{l(a)\ da} \tag{4.86b}$$

In continuous notation, life expectancy is, therefore,

$$\mathrm{e(x)} = \int_{\mathrm{a=x}}^{\infty} \mathrm{l(a)da} \div \mathrm{l(x)} \tag{4.87a}$$

$$\mathrm{e(0)} = \int_{\mathrm{a+0}}^{\infty} \mathrm{l(a)da} \div \mathrm{l(0)} \tag{4.87b}$$

The standard deviation of the ages of deaths for ungrouped data is computed as follows:

$$\sigma^2 = \int_0^{\infty} \mathrm{x^2\, f(x)dx} - \mathrm{x_m^2} \tag{4.88a}$$

$$= \int_0^{\infty} \mathrm{x^2\, f(x)dx} - \mathrm{e_0^2} \tag{4.88b}$$

$$\sigma = \sqrt{\sigma^2} \tag{4.89}$$

where $\mathrm{x_m}$ and $\mathrm{e_0}$ are the mean of the distribution of ages of life table deaths, or life expectancy at birth. For grouped data, life expectation at birth is:

$$e_0 = \int_0^{\infty} xf(x)dx \text{ or } \int_0^{\infty} S(x)dx \tag{4.90}$$

$$\text{Gini coefficient for ages of deaths} : 1 -^* \int_0^{\infty} [S(x)]^2 dx \div e_0 \tag{4.91}$$

$$\text{Keyfitz's H} : \left[\int_0^{\infty} (l(x)^* \ln l(x)dx)\right] \div \int_0^{\infty} l(x)dx \tag{4.92}$$

Appendix 4.2 Arriaga Method of Decomposing Change in Life Expectancy at Birth Among Age Groups

The steps in the Arriaga procedure of decomposing the change in life expectancy at birth among age groups over some time period are described and illustrated in this appendix. Specifically, the contribution of change in five age groups to the change in life expectancy at birth is measured for white females between 1949–1951 and 2000 in the United States. The ages selected are 0–14, 15–44, 45–64, 65–84, and 85–99 (or 85 and over). Data are needed for the following functions, e_x, l_x, and T_x, at the initial ages of the age groups listed. From these values of e_x, l_x, and T_x, one can derive all the intermediate values required to solve the several equations given by Arriaga. These equations measure the direct effect, the indirect effect, and the interaction effect of the mortality change in an age group on the change in life expectancy at birth (see Table 4.A3).

We first compute the direct effect, that is, the adjusted age-bounded life expectancies for the age groups. The Arriaga equation for the direct effect is as follows:

$$DE = (l_x^t \div l_a^t)\ ({}_ie_x^{t+n} - {}_ie_x^t) \tag{4.93}$$

where ${}_ie_x^t$ and ${}_ie_x^{t+n}$ represent age-bounded life expectancy at the earlier and later dates, l_a^t refers to the survivors at the age for which change in life expectation is being decomposed (for life expectancy at birth, $a = 0$), and l_x^t represents the survivors at the initial age at the earlier date for the contributing age group.

The equations for measuring the indirect effect are:

$${}_iCS_x = l_x^t * (l_{x+i}^{t+n} \div l_x^{t+n}) - l_{x+i}^t \tag{4.94a}$$

$${}_iIE_x = ({}_iCS_x \div l_a^t) * e_{x+2b}^t \tag{4.94b}$$

The equations for measuring the interaction are:

$${}_iOE_x = ({}_iCS_x \div l_a^t) * e_{x+i}^{t+n} \tag{4.95a}$$

$${}_iI_x = {}_iOE_x - {}_iIE_x \tag{4.95b}$$

For the open-ended interval, since this is the last age group, the indirect effect and the interaction do not apply; only the direct effect applies. The formula for the direct effect at these ages is different from the formula for the other ages; it calls simply for life expectancies at age x.

$$DE_{x+} = (l_x^t / l_a^t)(e_x^{t+n} - e_x^t) \tag{4.96}$$

Appendix Tables

Table 4.A1 Gain in life expectancy at birth from eliminating specified causes of death: United States, 1989–1991

Cause of death (ICD-9 codes)		Gain in years
Infectious and parasitic diseases	(001–139)	0.45
Tuberculosis	(010-018)	0.01
Septicemia	(038)	0.10
Human immunodeficiency virus infection	(*042 −* 044)	0.26
Malignant neoplasms, including neoplasms of lymphatic and hematopoeitic tissues	(140-208)	3.36
Malignant neoplasms of lip, oral cavity, pharynx	(140-149)	0.05
Malignant neoplasms of digestive organs and peritoneum	(150-159)	0.69
Malignant neoplasms of respiratory and intrathoracic organs	(160-165)	0.93
Malignant neoplasms of trachea, bronchus, and lung	(162)	0.89
Malignant neoplasms of breast	(174-175)	0.29
Malignant neoplasms of genital organs	(179-187)	0.31
Malignant neoplasms of urinary organs	(188-189)	0.11
Leukemia	(204-208)	0.12
Benign neoplasms, carcinoma in situ, and neoplasms of uncertain behavior and unspecified nature	(210-239)	0.04
Diabetes mellitus	(240)	0.27
Senile and presenile organic psychotic conditions	(290)	0.03
Alzheimer's disease	(331.0)	0.05
Parkinson's disease	(332)	0.03
Major cardiovascular diseases	(390-448)	6.73
Diseases of heart	(390-398, 402, 404-429)	4.64
Rheumatic fever and rheumatic heart disease	(390-398)	0.03
Coronary heart disease	(410-414, 402,429.2)	3.43
Hypertensive heart disease	(402)	0.11
Hypertensive heart and renal disease	(404)	0.01
Ischemic heart disease	(410-414)	2.77
Acute myocordial infarction	(410)	1.29
Hypertension with or without renal disease	(401, 413)	0.04
Cerebrovascular diseases	(430-438)	0.68
Atherosclerosis	(440)	0.07
Aortic aneurysm	(441)	0.08

(continued)

Table 4.A1 (continued)

Cause of death (ICD-9 codes)		Gain in years
Diseases of the respiratory system	(416-519)	0.97
Pneumonia and influenza	(480-487)	0.35
Chronic obstructive pulmonary diseases and allied conditions	(490-496)	0.45
Pneumonia due to inhalation of food or vomit	(507.0)	0.03
Diseases of the digestive system	(520-579)	0.46
Ulcers of the stomach and duodenum	(531-533)	0.03
Chronic liver disease and cirrhosis	(571)	0.20
Nephritis, nephotic syndrome and nephrosis	(580-589)	0.10
Urinary tract infection	(559.0)	0.04
Congenital anomalies	(740-759)	0.20
Certain conditions originating in the perinatal period	(760-779)	0.33
Sudden infant death syndrome	(798.0)	0.10
Accidents and adverse effects	(E800-E949)	0.92
Motor vehicle accidents	(E810-E825)	0.51
All other accidents and adverse effects	(E800-E807, E826-E949)	0.40
Residential fires	(E90-E899)	0.05
Suicide	(E950-E959)	0.30
Homicide and legal intervention	(E960-E978)	0.29
Other than the 15 leading causes of death		1.96

Source: U.S. National Center for Health Statistics 1999

The 15 leading causes of death in 1989–1991 are shown in italics. Beginning with data for 1987 NCHS introduced categories *042 −* 044 for classifying and coding human immunodeficiency virus infection; the asterisks before the category numbers indicate that they are not part of ICD-9

Table 4.A2 Probability at birth of eventually dying from specified causes of death: United States, 1989–1991

Cause of death (ICD-9 codes)		Total population	Male	Female
Infectious and parasitic diseases	(001-139)	0.02098	0.02550	0.01635
Tuberculosis	(010-018)	0.00076	0.00097	0.00056
Septicemia	(038)	0.00960	0.00823	0.01090
Human immunodeficiency virus infection	(*042 −* 044)	0.00697	0.01233	0.00155
Malignant neoplasms, including neoplasms of lymphatic and hematopoeitic tissues	(140-208)	0.22033	0.23899	0.20395
Malignant neoplasms of lip, oral cavity pharynx	(140-149)	0.00351	0.00464	0.00242
Malignant neoplasms of digestive organs and peritoneum	(150-159)	0.05445	0.05745	0.05172
Malignant neoplasms of respiratory and intrathoracic organs	(160-165)	0.06165	0.08206	0.04268
Malignant neoplasms of trachea, bronchus, and lung	(162)	0.05947	0.07870	0.04160
Malignant neoplasms of breast	(174-175)	0.01847	0.00025	0.03586

(continued)

Table 4.A2 (continued)

Cause of death (ICD-9 codes)		Total population	Male	Female
Malignant neoplasms of genital organs	(179-187)	0.02609	0.03285	0.02040
Malignant neoplasms of urinary organs	(188-189)	0.00940	0.01223	0.00681
Leukemia	(204-208)	0.00801	0.00877	0.00732
Benign neoplasms, carcinoma in situ, and neoplasms of uncertain behavior and unspecified nature	(210-239)	0.00308	0.00275	0.00339
Diabetes mellitus	(240)	0.02177	0.01832	0.02503
Senile and presenile organic psychotic conditions	(290)	0.00533	0.00358	0.00696
Alzheimer's disease	(331.0)	0.00747	0.00546	0.00936
Parkinson's disease	(332)	0.00368	0.00440	0.00305
Major cardiovascular diseases	(390-448)	0.46512	0.43269	0.09532
Diseases of heart	(390-398, 402, 404-429)	0.36308	0.34982	0.37533
Rheumatic fever and rheumatic heart disease	(390-398)	0.00270	0.00161	0.00372
Coronary heart disease	(410-414, 402,429.2)	0.29213	0.28634	0.29740
Hypertensive heart disease	(402)	0.01023	0.00808	0.01218
Hypertensive heart and renal disease	(404)	0.00121	0.00097	0.00144
Ischemic heart disease	(410-414)	0.24585	0.24477	0.24686
Acute myocordial infarction	(410)	0.11521	0.12184	0.10902
Hypertension with or without renal disease	(401, 413)	0.00458	0.00378	0.00534
Cerebrovascular diseases	(430-438)	0.07495	0.05752	0.09123
Atherosclerosis	(440)	0.01082	0.00786	0.01357
Aortic aneurysm	(441)	0.00764	0.00108	0.00540
Diseases of the respiratory system	(416-519)	0.09588	0.09986	0.09286
Pneumonia and influenza	(480-487)	0.04275	0.03896	0.04648
Chronic obstructive pulmonary diseases and allied conditions	(490-496)	0.04028	0.04734	0.03413
Pneumonia due to inhalation of food or vomit	(507.0)	0.00430	0.00428	0.00435
Diseases of the digestive system	(520-579)	0.03428	0.03426	0.03492
Ulcers of the stomach and duodenum	(531-533)	0.00311	0.00293	0.00328
Chronic liver disease and cirrhosis	(571)	0.00996	0.01279	0.00721
Nephritis, nephotic syndrome and nephrosis	(580-589)	0.01053	0.01030	0.01080
Urinary tract infection	(559.0)	0.00641	0.00421	0.00839
Congenital anomalies	(740-759)	0.00360	0.00376	0.00343
Certain conditions originating in the perinatal period	(760-779)	0.00434	0.00480	0.00386

(continued)

Table 4.A2 (continued)

Cause of death (ICD-9 codes)		Total population	Male	Female
Sudden infant death syndrome	(798.0)	0.00134	0.00160	0.00106
Accidents and adverse effects	(E800-E949)	0.03196	0.04043	0.02340
Motor vehicle accidents	(E810-E825)	0.01397	0.01880	0.00911
All other accidents and adverse effects	(E800-E807, E826-E949)	0.01799	0.02163	0.01430
Residential fires	(E90-E899)	0.00150	0.00179	0.00121
Suicide	(E950-E959)	0.00983	0.01584	0.00394
Homicide and legal intervention	(E960-E978)	0.00679	0.01039	0.00312
Other than the 15 leading causes of death		0.13610	0.12606	0.14548

Source: U.S. National Center for Health Statistics 1999

The 15 leading causes of death in 1989–1991 are shown in italics. Beginning with data for 1987 NCHS introduced categories *042 −* 044 for classifying and coding human immunodeficiency virus infection; the asterisks before the category numbers indicate that they are not part of ICD-9

Table 4.A3 Illustrative calculations for deriving the relative contribution of changes in age groups to the changes in life expectancy at birth between 1949–1950 and 2000, for the white female population of the United States

Age	l_x, 1949–1951	l_x, 2000	T_x, 1949–1951	T_x, 2000	e_0, 1949–1951	e_0, 2000
0–14	100,000	100,000	7,203,179	7,996,958	72.03	79.97
15–44	96,756	99,243	5,746,259	6,506,503	59.39	65.56
45–64	92,725	97,044	2,885,849	3,552,953	31.12	36.61
65–84	76,773	87,385	1,151,366	1,680,965	15.00	19.24
85–99	21,348	43,112	103,172	283,226	4.83	6.57
100+	294	2,244	566	5,351	1.92	2.38

	Age-bounded life expectancies				Indirect effect	
Age	1949–1951	2000	Difference	Direct effect	l^{t+n} quotient	${}_iCS_x$
0–14	14.571	14.905	0.334	0.334	.99243	2487
15–44	29.561	29.761	0.200	0.194	.97784	1887
45–64	18.706	19.290	0.584	0.542	.90047	6723
65–84	13.654	15.995	2.341	1.797	.49336	16,529
85–99	4.802	6.445	1.643	0.371	.05205	817
100+	1.925	2.385	0.460	–	–	–
All ages				3.238		

	Indirect effect				Interaction effect		
Age	e^t_{x+i}	${}_iCS_x \div l^t_a$	${}_iIE_x$	e^{t+n}_{x+i}	${}_{ii}OE_x$	${}_iI_x$	Sum of effects
0–14	59.39	.02487	1.477	65.56	1.630	.153	1.964
15–44	31.12	.01887	0.587	36.61	0.691	.104	.885
45–64	15.00	.06723	1.008	19.24	0.716	.286	1.836
65–84	4.83	.16529	0.798	6.57	1.08	.288	2.883
85–99	1.92	.00817	[0.016]	[2.38]	[0.019]	[.-0.003]	.371
All ages			3.870			.831	7.939

Source: Basic data from NCHS life tables for 1949–1951 and 2000; method based on Arriaga (1984)

Table 4.A4 Illustrative calculation of an approximation to Keyfitz' H in the life table of the total population in 2000

Age (x)	L_x (1)	l_x/100,000 (2)	Ln l_x/100,000 (3)	L_x^*Lnl_x/100,000 (4) = (1)*(3)
0–1	99392	1	0	0
1–5	396916	0.99307	−0.00695	−2758.6
5–10	495668	0.99177	−0.00826	−4094.2
10–15	495278	0.99095	−0.00909	−4502.1
15–20	494200	0.98992	−0.01013	−5006.2
20–25	492113	0.98654	−0.01355	−6668.1
25–30	489702	0.98181	−0.01836	−8990.9
30–35	487130	0.97696	−0.02331	−11355
35–40	483813	0.97132	−0.0291	−14079
40–45	479070	0.96349	−0.03719	−17816.6
45–50	472085	0.9521	−0.04909	−23174.7
50–55	461940	0.93522	−0.06697	−30936.1
55–60	447124	0.91113	−0.09397	−41613.8
60–65	424879	0.87498	−0.13355	−56742.6
65–70	392758	0.82131	−0.19685	−77314.4
70–75	348168	0.74561	−0.29355	−102204.7
75–80	289331	0.64244	−0.44248	−128023.2
80–85	215947	0.51037	−67262	−145250.3
85–90	133503	0.34959	−1.05099	−140310.3
90–95	62766	0.18839	−1.66924	−104771.5
95–100	20388	0.07252	−2.62389	−53495.9
100+	4636	0.01781	−4.028	−18673.8
Sum	T0 = 7,686,810			−997782
H		Sum of col. 4 ÷ T_0 = 997782 ÷ T_0 = 0.1298		

Source: Based on the U.S. life table for 2000 (U.S. National Center for Health Statistics 2002b) and N. Keyfitz, *Applied Mathematical Demography*, New York: Wiley, 1977

References and Suggested Readings

Sources

Coale, A. J., Demeny, P., & Vaughn, B. (1993). *Regional model life tables and stable populations* (2nd ed.). New York: Academic Press.

U.S. National Center for Health Statistics. (1975). U.S. life tables, 1969–71. By T. N. E. Greville. *U.S. Decennial Life Tables for 1969–71, 1*(1). Hyattsville, MD: U.S. National Center of Health Statistics.

U.S. National Center for Health Statistics. (1993). Advance report of final mortality statistics, 1991. By D. L. Hoyert & B. H. Hudson. *Monthly Vital Statistics Reports, 42*(2), Supplement. Hyattsville, MD: U.S. National Center of Health Statistics.

U.S. National Center for Health Statistics. (1998). United States abridged life tables, 1996. By R. N. Anderson. *National Vital Statistics Reports, 47*(13). Hyattsville, MD: U.S. National Center of Health Statistics.

U.S. National Center for Health Statistics. (2002a). Deaths: Final data for 2000. By A. M. Miniño, E. Arias, S. L. Murphy, & B. L. Smith. *National Vital Statistics Reports, 50*(15). Hyattsville, MD: U.S. National Center of Health Statistics.
U.S. National Center for Health Statistics. (2002b). United States life tables, 2000. By E. Arias. *National Vital Statistics Reports, 51*(3). Hyattsville, MD: U.S. National Center of Health Statistics.
U.S. National Office of Vital Statistics. (1955). *Vital statistics of the United States, 1950*, Vol. III, *Mortality Data*. Washington, DC: Public Health Service.
U.S. Office of the Chief Actuary, Social Security Administration. (2005). Life tables for the United States social security area: 1900–2100. By F. C. Bell & M. L. Miller. *Actuarial study* No. 120. Washington, DC: US Government Printing Office.

Anatomy and Construction of Standard Life Tables

Chiang, C. L. (1968). *An introduction to stochastic processes in biostatistics*. New York: Wiley.
Greville, T. N. E. (1943). Short methods of constructing abridged life tables. *Record of the American Institute of Actuaries, 32*(65), 29–42. Part 1.
Keyfitz, N., & Frauenthal, J. (1975). An improved life table method. *Biometrics 21*, 889–899.
Kintner, H. J. (2004). The life table. Chapter 13 In J. S. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed.), San Diego: Elsevier/Academic Press.
Preston, S., Keyfitz, N., & Schoen, R. (1972). *Causes of death: Life tables for national populations*. New York: Seminar Press.
Reed, L. J., & Merrell, M. (1939). A short method for constructing an abridged life table. *American Journal of Hygiene, 30*(2), 33–62. Tables reproduced. In J. S. Siegel & D. W. Swanson (Eds.), *The methods and materials of demography* (2nd ed.). San Diego, CA: Elsevier/Academic Press, 2004.
Schoen, R. (1978). Calculating life tables by estimating Chiang's *a* from observed rates. *Demography, 15*, 625–635.
Schoen, R. (1988). *Modeling multigroup populations*. New York: Plenum Press. Chapter 1.
U.S. National Center for Health Statistics. (1966). Comparison of two methods of constructing abridged life tables by reference to a standard table. By M. Sirken. *Vital and Health Statistics, 2*(4). Hyattsville, MD: U.S. National Center of Health Statistics.
U.S. National Center for Health Statistics. (1999). Method for constructing complete annual U.S. life tables. By R. N. Anderson. *Vital and Health Statistics, 2*(129). Hyattsville, MD: U.S. National Center of Health Statistics.

Limitations, Variants, and Extensions of Standard Life Tables

Birnbaum, Z. W. (1979). On the mathematics of competing risks. *Vital and Health Statistics: Data Evaluation and Methods Research, 2*(77), 1–58. Hyattsville, MD: U.S. National Center for Health Statistics.
Bongaarts, J., & Feeney, G. (2002). How long do we live? *Working Papers* No. 156, Policy Research Division, Population Council.
Bongaarts, J., & Feeney, G. (2003). Estimating mean lifetime. *Working Papers* No. 179, Policy Research Division, Population Council.
Bongaarts, J., & Feeney, G. (2005). The quantum and tempo of life-cycle events. *Working Papers* No 207, Policy Research Division, Population Council.

Brouard, N. (1986). Structure et dynamique des populations. La pyramide des années à vivre, aspects nationaux et exemples régionaux. *Espace-Populations-Sociétés, 214*(15), 157–168.
Crowder, M. J. (2002). *Classical competing risks*. London: Chapman and Hall/CRC.
Guillot, M. (2003). A cross-sectional average length of life (CAL): A cross-sectional mortality measure that reflects the experience of cohorts. *Population Studies, 57*(1), 41–54.
Jacobson, P. H. (1964).Cohort survival for generations since 1840. *Milbank Memorial Fund Quarterly, 42*(3), 36–51.
Moeschberger, M. L., & Klein, J. P. (1995). Statistical methods for dependent competing risks. *Lifetime Data Analysis, 1*(2), 195–204.
Namboodiri, K. (1991). *Demographic analysis: A stochastic approach*. San Diego, CA: Academic Press.
Sardon, J. P. (1994). A period measure of mortality. The example of France. *Population: English Selection, 6*, 131–150.
Schoen, R. (1987). *Modeling multigroup populations*. New York: Plenum Press.
Schoen, R., & Canudas-Romo, V. (2005). Changing mortality and average cohort life expectancy. *Demographic Research, 13*(5), 117–142.
U.S. Office of the Chief Actuary, Social Security Administration. (2005). Life tables for the United States Social Security Area: 1900–2100. By F. C. Bell & M. L. Miller. *Actuarial study* No.120, Washington, DC: US Government Printing Office.
U.S. National Center for Health Statistics. (1968). United States life tables by causes of death: 1951–61. By F. Bayo. *Life Tables*, 1959–61, 1(6). Washington, DC: U.S. National Center for Health Statistics.
U.S. National Center of Health Statistics. (1999). United States life tables eliminating certain causes of death. By R. N. Anderson. *U.S. Decennial Life Tables for 1989–91 1*(4). Hyattsville, MD: National Center of Health Statistics
Vaupel, J. W. (2002). Life expectancy at current rates vs. current conditions: A reflection stimulated by Bongaarts and Feeney's "How Long Do We Live?" *Demographic Research, 7*, 365–378.
Wilmoth, J. R. (2005, March 31–April 2). *A simple explanation of tempo effects in fertility – and why such issues are irrelevant for mortality analysis*. Paper presented at the annual meeting of the Population Association of America, Philadelphia, PA.

Analysis and Applications of Life Tables

Arriaga, E. E. (1984). Measuring and explaining the change in life expectancies. *Demography, 21*(1), 83–96.
Austad, S. (1997). *Why we age?* New York: Wiley
Canudas-Romo, V. (2006, March 30–April 1, 2006). *The modal age at death and the shifting mortality hypothesis*. Paper presented at the annual meeting of the Population Association of America, Los Angeles, CA.
Carnes, B. A., Holden, I. R., Olshansky, S. J., Witten, T. M., & Siegel, J. S. (2006). Mortality partitions and their relevance to research on senescence. *Biogerontology*, 7, 183–198.
Chen, J., Senturk, D., Wang, J. L., Miller, H. G., et al. (2007). A demographic analysis of the fitness cost of extended longevity in *Caenorhabditis Elegans. Journal of Gerontology: Biological Sciences, 62A*(2), 126–135.
Cheung, S. L. K., & Robine, J.-M. (2007). Increase in common longevity and compression of mortality: The case of Japan. *Population Studies, 61*(1), 85–97.
Hanada, K. (1983). A formula of Gini's concentration ratio and its application to life tables. *Journal of the Japan Statistical Society, 13*(2), 95–98.
Nusselder, W. J., & Mackenbach, J. P. (1996). Rectangularization of the survival curve in the Netherlands, 1950–1992. *Gerontologist, 36*, 773–782.

Robine, J. M., Cheung, S. L. K., Thatcher, A. R., & Horiuchi, S. (2006, March 30–April, 1). *What can be learned by studying the adult modal age at death*? Paper presented at the annual meeting of the Population Association of America, 2006.

Shryock, H. S., Siegel, J. S, & Associates. (1980). Life tables. Chapter 15 In *The methods and materials of demography* (Vol. II). Washington, DC: U.S. Bureau of the Census.

U.S. National Center for Health Statistics. (1994). Causes of death contributing to changes in life expectancy: United States, 1984–89. By K. D. Kochanek, J. D. Maurer, & H. M. Rosenberg. *Vital and Health Statistics 20*(23). Hyattsville, MD: U.S. National Center of Health Statistics.

U.S. Social Security Administration, Office of the Chief Actuary. (2002). Life tables for the United States Social Security Area 1900–2100. By F. C. Bell & M. L. Miller. *Actuarial study*, No. 116, Washington, DC: US Government Printing Office

Vaupel, J. W. (1986). How changes in age-specific mortality affects life expectancy. *Population Studies, 40*, 147–157.

Wilmoth, J. R., & Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography, 36*(4), 475–495.

Chapter 5
Measures of Health Status, Functioning, and Use of Health Services

Concepts of Health Conditions and Classification of Diseases

Concepts of Health Conditions

A great number and variety of terms are used to refer to health conditions. The principal terms are disease, illness, sickness, disorder, pathology, health condition, morbid condition, impairment, injury, disability, and handicap. Dependency, frailty, and the standard categories of self-perceived health (excellent, . . . , poor) also describe health conditions, but in more global terms. Health demographers and epidemiologists have given some of these terms formal definitions and we shall note the definitions that are available. Other terms are used more loosely and lack formal definition. The term morbidity has been used to refer both to health as a field of study and to specific health conditions. The terms health condition, pathology, morbid condition, and disorder are the other general terms among those enumerated but all can refer to a specific condition. They cannot be satisfactorily distinguished from one another.

Health conditions or health disorders may be viewed as including diseases, injuries, and impairments, or any condition, lethal and non-lethal, that can arise from the biological processes of aging or external assaults or trauma (Aging can lead to changes in the body that give rise to disease but the vast majority of scholars do not consider aging itself a disease). A disease may be defined as any health condition except for injuries and impairments; that is, a health condition resulting from infection or a genetic or other internal deficiency, and usually characterized by pain, discomfort, or limitation of functioning (Humber and Almeder 1997).

An injury is an acute physical disorder resulting from an accident or deliberate attack against a person. Injuries, unlike diseases, must be defined simultaneously by the causative event and the resulting pathology. It has also been given a theoretical definition, namely, "the physical and psychological damage to the person produced by energy exchanges that have sudden discernible effects" (U.S. NCHS 2003; Robertson 1998; Haddon 1980). The most common operational definition of an

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_5, © Springer Science+Business Media B.V. 2012

injury refers to those health conditions listed in the Injury and Poisoning Chapter (XVII) of the International Classification of Diseases (ICD-9) and all those events coded to ICD Supplementary External Causes of Injury and Poisoning (E codes in ICD-9); or those pathologies included in Chapters XIX and XX of ICD-10. (See Chap. 3.)

It is useful to consider impairment, disability, handicap, dependency, and frailty apart from the other terms because they refer to the consequences of disease. These terms refer to an inability to function in some capacity, in addition to having one or more health conditions. Impairment describes any mental, physiological, or anatomical disorder or condition resulting from disease, injury, the normal processes of aging, or congenital anomaly. It usually refers to some chronic condition and often describes some sensory or musculoskeletal defect, such as hearing loss or arthritis. Disability refers to a loss of a function, because of a health condition, in some area(s) of a person's life. It does not necessarily prevent the person from participating in most areas of his or her life. A handicap refers to a mental or physical impairment that substantially limits one in carrying out one or more major life activities, particularly in one's social life. Dependency refers to the lack of control exercised by an impaired, disabled, or handicapped individual over certain basic life activities and the corresponding need to depend on others to perform these activities.

Frailty is a descriptor of a general health state (like poor self-perceived health) that may involve multiple chronic diseases, disability, and a reduction in functioning. Biologically it represents dysregulation of the homeostatic mechanisms of the body (Ferrucci et al. 2005).[1] A frail person is likely to have a number of chronic conditions, such as sarcopenia (i.e., loss of lean body mass) and arthritis (i.e., joint stiffness and pain), that result in limited mobility and restrictions on usual activities (Bortz 2002; Fried et al. 2001). Reference is sometimes made to the frail elderly, older adults who require support from others because of an accumulation of debilities, such as described.

Classification of Diseases

As stated in Chap. 3, the World Health Organization's (WHO) *International Classification of Diseases* (ICD) serves for both the classification of diseases and the classification of deaths. National governments follow the classification of diseases recommended by WHO. The Tenth Revision of the *International Classification of Diseases* (ICD-10) was published by WHO in 1992–1994. The United States shifted to the new classification beginning in 1999 and by 2005 most countries were using the new classification.

[1] Homeostasis is the ability of a cell, organ, or physiological system to maintain its internal equilibrium in a healthy state by adjusting its physiological processes in accordance with the changes in related physical structures. One might say, it is the balance achieved by the body through communication of the body's parts to each other.

The ICD includes classifications for mental and physical diseases. The definitions of the types of mental illnesses are more complex than those for physical illnesses. Chapter V of ICD-10, "Mental and Behavioral Disorders," provides clinical descriptions in addition to diagnostic guidelines of the types of mental illnesses. In the United States a separate publication, *The Diagnostic and Statistical Manual of Mental Disorders, DSM-IV* (American Psychiatric Association 1994), also gives a classification of mental illnesses. The codes and terms provided in DSM-IV are fully compatible with ICD-10. In the classification of mental disorders, the clinician assigns a category on the basis of clinical syndromes, personality disorders, physical disorders and conditions, the severity of psychosocial stressors, and the highest level of adaptive functioning during some specified prior period, such as the previous year (Mausner and Kramer 1985). WHO has also promulgated a separate classification of impairments, disabilities, and handicaps that is described later.

Population Measures of Health

Prevalence Ratios and Incidence Rates

Population measures of health generally differ as to whether they refer to the health *status* of a population group at some particular date or at any time during a period, or to health *events* occurring to a population at risk over a year or similar period. The former measures are called measures of prevalence and the latter measures are called measures of incidence. Epidemiologists distinguish between two types of prevalence measures, one referring to an individual's health status at a single date during a period and the other referring to an individual's status at any time during a period. Both types of prevalence measures relate to having a specified health condition, regardless when the condition was incurred. Epidemiologists also distinguish two types of incidence measures. Incidence rates for specific conditions are computed either on the basis of the number of persons incurring a particular health condition during a year or the number of cases (episodes) of a particular health condition incurred during a year. The incidence rate for cases will be larger than the incidence rate for persons since some persons will experience the same event more than once during the year.

Measures of prevalence are simply proportions and, accordingly, they are termed ratios rather than rates. They express the proportion of the population that has acquired a disease, not the risk of the occurrence of the disease in the population. With appropriate data, one can compute several types of prevalence ratios for health conditions. For example, one may compute ratios for acute illnesses or injuries, and ratios for chronic illnesses or impairments. Alternatively, these conditions can be incorporated into a general prevalence ratio for acute conditions (acute illnesses and injuries) and a general prevalence ratio for chronic conditions (chronic illnesses and impairments). In addition to measures for the total population, separate measures

may be computed for the noninstitutional population because this population and the institutional population have notably different health levels.

The prevalence ratio for an acute disease (e.g., acute respiratory tract infection) for a given year is the proportion of persons with the acute disease in the population at the middle of the year, or the proportion of persons who have the acute disease at any time during the year, usually expressed per 100,000 (PR_c):

$$PR_c = (p_c \div P) * 100{,}000 \tag{5.1}$$

where p_c represents the number of persons with an acute disease (e.g., an acute respiratory tract infection) at the middle of the year (or at any time in the year) and P the total midyear population. The prevalence ratio for injuries is computed in the same way.

The types of incidence measures parallel the types of prevalence measures. The commonest types of incidence measures are the incidence rate for an acute illness and the incidence rate for a chronic illness. The first measure represents either the number of persons incurring an acute illness during a year, or the number of acute illnesses occurring during a year, per 100,000 midyear population. The second measure, the incidence rate for chronic illnesses, represents either the number of persons incurring a chronic illness during a year, or the number of cases of chronic illnesses occurring during the year, per 100,000 midyear population. Separate incidence rates can be computed for acute illness, injury, chronic illness, and impairments. For example, we may compute the number of persons incurring an injury during a year, or the number of injuries (episodes) occurring during the year, per 100,000 persons. The incidence rate for a chronic condition, IR_c, is calculated by the formula:

$$IR_c = (I_c \div P) * 100{,}000 \tag{5.2}$$

where I_c represents the number of chronic conditions occurring during the year and P represents the midyear population.

Age-Specific and Age-Cause-Specific Measures

All the measures noted above can be calculated for the entire population or a particular segment of it, such as an age group, age-sex group, or age-cause group. Thus, one can calculate such measures as age-specific prevalence ratios, age-specific incidence rates, age-cause-specific prevalence ratios, and age-cause-specific incidence rates. An age-cause specific prevalence ratio (PR_{ac}) is computed as follows:

$$PR_{ac} = (p_{ac} \div P_a) * 100{,}000 \tag{5.3}$$

where p_{ac} represents the prevalence of the disease for the age group and P_a represents the midyear population at that age. Note that the denominator population cannot be made specific for cause, only for age, sex, and other demographic

Table 5.1 Prevalence ratios for two infectious conditions of children under age 5 in most countries of Sub-Saharan Africa: selected years, 1995–2000

Country and year	Acute respiratory tract infection	Diarrhea
Benin, 1996	15.7	26.1
Botswana, 2000	38.5	6.5
Burkina Faso, 1999	13.5	20.0[a]
Camaroon, 2000	7.0	18.9[b]
Central African Republic, 1995	28.2	26.5[a]
Chad, 2000	12.5	31.2
Comoros, 2000	10.1	18.3
Côte d'Ivoire, 2000	3.7	20.1
Eritrea, 1995	23.0	23.6
Ethiopia, 2000	24.4	23.6
Gambia, 2000	7.7	21.5
Ghana, 1998	13.8	17.9
Guinea, 1999	15.9	21.2
Guinea Bissau, 2000	10.1	31.5
Kenya, 1998	20.1	17.1
Madagascar, 2000	10.5	12.8
Malawi, 1996	12.3	16.1
Mali, 1996	15.3	25.3
Mozambique, 1997	11.8	20.7
Niger, 2000	11.8	40.0
Nigeria, 1999	11.3	15.3
Senegal, 2000	6.6	21.3[a]
Sierra Leone, 2000	8.7	25.3
Tanzania, 1999	13.9	12.4
Togo, 1998	20.2	31.1
Uganda, 1995	27.1	23.5
Zambia, 1996	12.7	23.5
Zimbabwe, 1999	15.8	13.9

Source: UNICEF (2001)

Note: The prevalence of acute respiratory tract infection and diarrhea often varies by season. Country surveys were administered at different times and, hence, the prevalence ratios are not fully comparable across countries Percentage of children under age 5 reported as having an acute respiratory tract infection or diarrhea at the time of the survey

[a]For 1996

[b]For 1998

variables. Table 5.1 shows prevalence ratios for two acute childhood conditions for most countries of sub-Saharan Africa for various years from 1991 to 2000. For example, the prevalence ratio for acute respiratory tract infection for children under 5 years of age in Botswana in 2000 is 38.5. It is calculated as follows:

$$38.5 = (78{,}900 \div 205{,}000) * 100$$

Table 5.2 presents illustrative prevalence ratios for selected chronic health conditions for males and females for the United States in 2007.

Table 5.2 Prevalence ratios for selected chronic conditions for persons 18 years and over, for the United States: 2007 and 1998

Chronic condition and sex	2007	1998
Arthritis		
Total	20.3 (20.8)	20.0
Female	17.5	18.0
Male	22.8	21.9
Diabetes		
Total	7.6 (7.8)	5.3
Male	8.0	5.2
Female	7.3	5.4
Asthma		
Total	7.3 (7.3)	9.0
Male	5.4	7.6
Female	9.0	10.3
Cancer		
Total	7.3 (7.3)	6.2
Male	7.1	5.0
Female	7.6	7.2
Heart disease		
Total	11.2 (11.3)	11.5
Male	12.5	11.1
Female	10.2	11.8

Source: U. S. National Center for Health Statistics, *Vital and Health Statistics,* Series 10 (240), 2009; Series 10 (209), 2002
Percentage of population 18 years and over. Figures for 2007 are standardized on U.S. 2000 census-based population; figures for 1998 are crude ratios; figures for 2007 in parentheses are crude ratios

A sex-cause-specific incidence rate is computed as follows:

$$\mathrm{IR_{sc}} = \mathrm{I_{sc}} \div \mathrm{P_s} * 100{,}000 \tag{5.4a}$$

where IR_{sc} represents a sex-cause specific incidence rate, c a cause of illness, s a sex group, and P the midyear population. For example, the incidence rate for cancer of the stomach for females in Japan in 2000 is,

$$\mathrm{IR_{sc}} = (37{,}920 \div 64{,}650{,}000) * 100{,}000 = 58.7.$$

A sex-age-cause specific incidence rate is computed as follows:

$$\mathrm{IR_{sac}} = \mathrm{I_{sac}} \div \mathrm{P_{sa}} * 100{,}000 \tag{5.4b}$$

Table 5.3 Incidence rates for specified types of cancers, by sex, for the countries of East Asia: 2000

Country and sex	Melanoma	Colon/rectal	Stomach	Lung	Leukemia
Male					
China, People's Republic of	0.21	12.24	33.25	35.06	4.07
China, Hong Kong	1.16	43.01	23.95	91.62	6.80
Japan	0.63	77.74	124.63	76.78	7.75
Korea, Democratic People's Republic of	0.28	10.23	48.25	31.08	4.66
Korea, Republic of	0.33	13.01	61.09	40.32	4.92
Mongolia	0.45	2.47	22.54	16.63	3.16
Female					
China, People's Republic of	0.17	10.12	17.90	16.04	3.17
China, Hong Kong	1.06	39.48	14.16	46.07	5.16
Japan	0.49	54.06	58.66	27.17	5.20
Korea, Democratic People's Republic of	0.23	9.34	24.17	11.36	3.42
Korea, Republic of	0.26	11.48	29.33	13.91	3.56
Mongolia	0.38	2.11	15.30	7.16	3.27

Source: Ferlay et al. (2001), Reprinted with permission
Incidence rates represent the number of new cases of the specified type of cancer reported during the year per 100,000 persons

For example, the estimated incidence rate for prostate cancer for U.S. males 55–64 years of age in 2005 was 111.1 per 100,000 population:

$$\mathrm{IR}_{\mathrm{mc},}55{-}64 = (16{,}240 \div 14{,}618{,}000) * 100{,}000 = 111.1$$

Table 5.3 shows incidence rates for several types of cancers for males and females for six countries in East Asia for the year 2000.

The incidence rate may also be expressed in cohort or probability form, wherein the denominator is the population free of the condition at the beginning of the year and the rate indicates the chances that this population will incur the condition during the following year.

Incidence data in cohort form for age groups may be obtained directly from various health surveys or disease registers, or estimated from prevalence data for successive ages in the same year or in successive calendar years obtained from censuses or surveys. The formulas for such estimates are as follows:

$$\text{“Period” incidence rate, year } y, \text{ age } a = \frac{(\text{prevalence ratio}_{\mathrm{y,\,a+1}}) - (\text{prevalence ratio}_{\mathrm{y,\,a}})}{\text{prevalence ratio}_{\mathrm{y,\,a}}} \tag{5.5a}$$

$$\text{“Cohort” incidence rate, year } y, \text{ age } a = \frac{(\text{prevalence ratio}_{y+1,a+1}) - (\text{prevalence ratio}_{y,a})}{\text{prevalence ratio}_{y,a}} \quad (5.5b)$$

By this indirect and approximate method of deriving the measures, the denominators are derived from the same source as the numerators and the chance of relating figures that are not comparable is reduced.

Average Age of Persons with a Chronic Disease and Average Age at Onset

Measures of the average age of persons with a disease, the average age of onset of a disease, and the average duration of a disease are useful in evaluating progress in the control of the disease, measuring the relation between possible causative factors and the disease, and developing assumptions of mortality for population projections.

For ascertaining the average age of persons with any chronic disease or a specific chronic disease, one can simply calculate the median or mean age of this population from data derived from health surveys. For a measure of the average age of persons with a chronic disease independent of the population age distribution, one can calculate the median or mean age from the age-specific percents with the disease.

The average (e.g., median, mean) age of onset of a disease represents the average age at which the disease was first reported, diagnosed, or experienced by a group of persons, whether for a given calendar period or for a cohort born in a given period. One can sometimes secure data from a survey (e.g., NHIS) or disease register (e.g., cancer register) on the number of persons incurring a disease during a period. If such data are tabulated in age groups, one can determine the mean or median age of the new occurrences (i.e., average age at onset) directly.

Although the average age of the cases of the disease does not really measure the age of onset of the disease, comparison of the average ages of the cases of the disease for two dates, say 1995 and 2005, may give a rough indication of the change in the age of onset of the disease between these years. The limitation of this measure is that the average age of persons with a disease could have remained unchanged while the age at onset may have risen or fallen because of a change in the dispersion of the durations of the illness (possibly as result of a change in the distribution of the population at risk). A more exact measure of the change in the age of onset of the disease would be based on an analysis of the age distribution of the occurrences of the disease. In this analysis we can select the age corresponding to the lowest 10th percentile of the distribution, for example, or the age corresponding to the mean minus two standard deviations from the mean.

Such cross-sectional estimates of the average age of onset and of changes in this measure would differ from those calculated for real birth cohorts. The latter are not only more descriptive of the experience of actual groups of people but also are more analytically useful and informative. The data for deriving such figures would come from longitudinal panel surveys or, with more approximate results, periodic cross-sectional surveys that permit relating data for the same birth cohort over time.

Average Duration of Illness and the Compression of Morbidity

The average duration of an illness refers to the period of time that the illness lasts among a group of subjects with the illness. Accordingly, the average duration of a chronic disease of later life is approximated by the average time between the age of onset of the disease and the age of death. It may be represented as the quotient of the combined years of duration (D_i) of all the cases of the chronic disease prevalent in a calendar year and the number of cases (C_i) of the disease prevalent in the year (i.e., $\sum D_i \div C_i$). Alternatively, it may be taken as the quotient of (1) the difference between the total for the ages at onset for the cases of a chronic disease in a year and the total for the ages at death of these cases and (2) the total number of cases of persons with the chronic disease [i.e., $(\sum A_{oi} - \sum A_{di}) \div C_i$]. The practical difficulties of matching the disease with the date of death are evident, but it may be feasible to obtain the necessary data in long-term panel surveys.

Linking the incidence rate and the prevalence ratio. The average duration of an illness is the main logical link between the incidence rate for the illness in a given year and the prevalence ratio for the condition in the year. This relation is represented approximately as:

$$\text{Prevalence ratio} \approx \text{incidence rate} * \text{average duration of illness} \tag{5.6a}$$

or

$$\text{Average duration of illness} \approx \text{prevalence ratio} \div \text{incidence rate} \tag{5.6b}$$

The average duration of chronic illness incorporates the effect of the mortality of the cases of illness and the recovery rate. Here is an example of how the formula works. Under constant conditions from year to year, if the incidence rate of an illness is 200 per 100,000 population in a year and the prevalence ratio is 1,000 per 100,000 population, then the average duration of the illness is 5.0 years:

$$\text{Average duration} \approx 0.010 \div .002 = 5.0$$

The relations in Eqs. 5.6a and 5.6b approach an identity when the elements in the equations tend to be constant from year to year.[2] The actual relation between the three measures may vary from this simple formulation because the demographic situation is dynamic. The incidence rate changes, persons migrate into and out of

[2]Note that under steady-state conditions the prevalence ratio is directly proportional to the incidence rate and the average duration of the illness. A more precise formula for calculating the relation between the prevalence ratio, the incidence rate, and the average duration of illness than shown is:

$$\text{Prevalence ratio} = \frac{\text{incidence rate} * \text{average duration of illness}}{[1 + \text{incidence rate} * \text{average duration of illness}]}$$

the area (and some of them have the illness in question), the case-survival rate and recovery rate may change over time, and the extent of reporting-error of the health data changes.

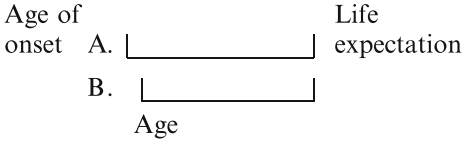
The case-survival rate is the proportion of the cases with an illness at a given date that continues to experience the health condition at a later date, that is, the proportion excluding deaths, recoveries, immigrants, and emigrants. If we disregard recoveries, immigrants, and emigrants, the case-survival rate is the complement of the case-fatality rate (= the proportion of cases succumbing to an illness):

$$\text{Case-survival rate} = 1 - \text{case-fatality rate} \tag{5.7}$$

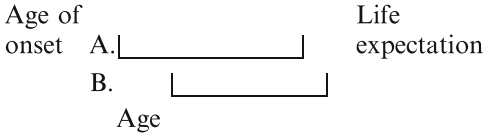
(The case-fatality rate is discussed further below.) Note that prevention of death may increase the number of persons with a chronic health condition, raise the case-survival rate, and thus raise the prevalence ratio. In other words, success at treating an illness without curing it adds to the burden of disease although death has been averted.

Compression of morbidity. The average age of onset of chronic morbidity bears an important relation to the change in life expectancy because this relation determines the duration of chronic morbidity. When the average age of onset rises faster than the increase in life expectancy, the period of chronic disease shrinks. The reduction in the average duration of chronic illness has been called the compression of morbidity (Fries 2003). The compression of morbidity can be diagrammed as follows:

1. When life expectation remains unchanged, but the age of onset rises:



2. When life expectation increases and the age of onset rises even more:



If the incidence rate is low or if the disease has a high fatality rate, this formula will approximate formula (5.6a) (See Kleinbaum et al. 1982).

The relation between the prevalence ratio and the incidence rate, or prevalence cases and incidence cases, may be structured in other ways. By analogy, we may apply the conventional population estimating equation with appropriate modifications, as follows:

$$PR_0 + I - D + Im - Em - R = PR_1$$

where the elements refer to prevalence of cases for a given disease at the start of a period (PR_o), new cases during the period (I), deaths of persons who have the disease (D) during the period, immigrants with the condition (Im), emigrants with the condition (Em), recovering cases reduced by relapses (R), and prevalence of cases at the end of the period (PR_1).

In both cases, age at onset is closer to life expectation in year B than in year A and hence morbidity has been compressed in this period. Compression of morbidity can also occur with a decline in life expectation and stability of age at onset of chronic disease, but this is a less likely possibility than the others. The compression of morbidity may also be defined as a faster decline in age-specific morbidity rates than in age-specific death rates at older age. In practice, it is difficult to prove that morbidity compression has occurred with specific chronic diseases because of the need to limit the deaths to those that occurred from the chronic diseases being studied. The argument has been made that compression of morbidity has been occurring, however, using data on disability in relation to all deaths (see below).

Prevalence of a Chronic Health Condition Among Segmented Population Groups

Age-specific prevalence data and ratios for particular chronic health conditions are directly obtainable from national health surveys, specifically the National Health Interview Survey or the public-use data based on it. The prevalence ratios state a very broad risk for the population of having the specific chronic disease. The accompanying data on demographic, economic, and health characteristics of the population and life-style and behavioral characteristics permit a division of the broader risk into more or less homogeneous segments of the population with respect to the risk. By subdividing the population into mutually exclusive risk groups that are homogeneous within them and heterogeneous among them, one can obtain separate risk statements for important risk groups. Age, sex, marital status, income class, education, smoking practice, degree of exercise use of alcohol, and such characteristics can be analyzed for their relative importance in identifying the appropriate risk groups. Regression analysis is useful in identifying the risk groups into which the population should be divided. A widely used methodology for this purpose is to determine the odds ratios for each characteristic by logistic regression.

The researcher may begin with estimates of the probability of being afflicted by a particular chronic disease (e.g., diabetes) in each of eight mutually exclusive segments of a population, i.e., these identified by sex, smoking/nonsmoking, under age 50/over age 50, or in each of sixteen segments, i.e., adding two marital classes. The results of such research can directly answer the question, what is the probability that an unmarried male over 50 who smokes will be afflicted with diabetes? If the question is, rather what is the probability that a male over age 50 will be afflicted with diabetes, simply compute the weighted average of the probability estimates of the relevant segments (married and unmarried, and smokers and nonsmokers among males over age 50), using as weights the sample sizes of the segments.

Probability of Incurring a Health Condition

Age-specific central rates of incurring an illness, injury, or impairment during a year may be converted to age-specific probabilities in the same way that age-specific death rates are converted to age-specific mortality rates. For single ages,

$$\mathrm{IR_{ac}}^{\mathrm{p}} = \mathrm{I_{ac}} \div (\mathrm{P_a} + 1/2\mathrm{D_a}) * 100 \tag{5.8a}$$

or more exactly,

$$\mathrm{IR_{ac}}^{\mathrm{pc}} = \mathrm{I_{ac}} \div ({}_{\mathrm{nc}}\mathrm{P_a} + 1/2\mathrm{D_a} + 1/2\mathrm{I_{ac}}) * 100 \tag{5.8b}$$

where I_{ac} refers to the number of persons incurring the health condition at some time during the year at age a, ${}_{nc}\mathrm{P_a}$ refers to persons aged a free of the condition at mid-year, and D_a refers to deaths at age a during the year. (There is a slight overlap in the terms of the formula inasmuch as some persons who became disabled during the year also died during the year.) These probabilities may be used, in combination with age-specific mortality rates, to construct a double-decrement table or a multistate table designed to measure at each age the chance of ever incurring the health condition and the expected years of life after incurring the condition, as explained further in Chap. 8.

Probability of ever incurring a disease or disability. As just suggested, estimating the lifetime probability that a population group will ever incur a disease or disability calls for the calculation of a double-decrement life table. I have previously described how the lifetime chances of dying from a particular cause can be derived from a cause-elimination life table (Chap. 4). For the present purpose, in addition to a schedule of mortality rates, we need a set of age-specific incidence rates for the disability or disease in question. The planned double-decrement life table would have two exit factors – death from any cause and occurrence of the disease. In the United States, national data on death rates are available from the vital statistics registration system, and national data on the occurrence or prevalence of a particular disease may be obtained from one or another of the national health surveys. Such a calculation usually makes the unrealistic assumption that the schedule of mortality rates selected applies both to the population with, and the population without, the disease or disability. This assumption can be avoided by developing a separate schedule of death rates for persons with the disease from the reported death rates for this cause shown in the vital statistics tabulations and the data on persons with the disease given in the National Health Interview Survey.

An illustrative truncated double decrement table showing the survival history of a synthetic cohort between the ages of 55 and 65 with respect to death and heart disease is shown as Table 5.4. The steps in the construction of a multiple decrement table showing the history of a real cohort with respect to death and disability are set forth in Chap. 8.

Individual's risk of ever incurring a chronic disease. The methodology may be modified to express the risk of incurring a particular chronic disease for individuals. Such a formula provides health counselors with a basis for advising individuals

Table 5.4 Illustrative truncated double-decrement table showing the survival history of a cohort between the ages of 55 and 65 reduced by death and heart disease

	Rate[a]		Survivors free of heart disease	Number dying or incurring heart disease in interval	
Age x to x+1	Mortality[b] (1)	Heart disease[c] (2)	$(3_{x-1})-(4_{x-1})-(5_{x-1}) = (3)$	Deaths (1) * (3) = (4)	Heart disease[d] (2) * (3) = (5)
55–56	.00476	.05160	100000	476	5160
56–57	.00508	.05144	94364	481	4875
57–58	.00545	.05128	89408	487	4585
58–59	.00590	.05112	84336	498	4311
59–60	.00645	.05096	79527	513	4048
60–61	.00710	.05080	74966	485	3808
61–62	.00784	.05064	70673	554	3579
62–63	.00859	.05048	66540	572	3360
63–64	.00931	.05032	62608	583	3150
64–65	.00999	.05016	58875	588	2953
65–66	:	:	55334	:	:
66–67	:	:	:	:	:
67–68	:	:	:	:	:

[a]Probabilities are only rough approximations; actual data or estimates are not available
[b]Mortality rates of persons who do not suffer from heart disease at the start of each age group and die of other diseases than heart disease in the interval. Rates were estimated from the U.S. life table for 2005, adjusted downward by 25% to reflect the lower mortality rate of persons without heart disease. The adjustment factor was estimated from an analysis of the death rates in the life tables eliminating deaths from heart disease
[c]Roughly estimated from the sparse prevalence ratios for heart disease reported in the NHIS for 2006. Intended to include the deaths of persons who incur heart disease in the age interval and subsequently die of heart disease or other disease
[d]The numbers for heart disease include those who incur heart disease in the age interval, whether of not they subsequently die of heart disease or other cause

about possible courses of action in reducing their risk of incurring a disease. Such a tool has been developed, for example, at the National Cancer Institute to measure a woman's individualized risk for invasive breast cancer over a 5-year period and over her lifetime (Gail et al. 1989). This tool applies statistical methods to data from the Breast Cancer Detection and Demonstration Project, a mammography screening project conducted in the 1970s.

Gail et al. developed the model for estimating the risk of developing breast cancer among a group of women who are enrolled in a program of annual mammographic screening, have no previous breast cancer, and show no evidence of breast cancer at the time of their initial screening mammogram. The model estimates the absolute risk or probability that a woman in a program of annual screening will develop invasive or *in situ* breast cancer over a defined age interval. The risk factors in this model include, in addition to current age, age at menarche, age at first parenthood, number of previous breast biopsies, presence of atypical hyperplasia[3] on biopsy,

[3]Hyperplasia is the abnormal increase in the number of cells in a tissue or organ, causing its enlargement.

and number of affected first-degree relatives. The individualized breast cancer probabilities are calculated from information on relative risks and a baseline hazard, that is, the rates for a patient without identified risk factors.

Similar measures of individualized risks of incurring other specific diseases could be devised for use in counseling individuals as to their options. Relevant risk factors need to be selected in the case of each disease. For example, one could develop a model for anticipating a heart attack in the subsequent 5 years for persons asymptomatic at the baseline date (i.e., free of symptoms perceived by the subject). The list of risk factors could include age, sex, race, number of first-degree relatives who have or have had coronary artery disease, blood pressure, cholesterol level, degree of coronary artery stenosis, and number of coronary arteries involved. If the results are to be used in decision-making regarding angioplasty or by-pass surgery, the formula could be adjusted to allow for such negative consequences of the surgery as the risks of the procedure, the immediate and lifetime trauma to the body caused by the procedure, and the cumulative effects of all associated diagnostic tests (e.g., multiple chest x-rays).[4]

Survival Period for a Disease

In the analysis of cancer morbidity, in particular, it has been of special interest to track the number of persons in the population with various types of cancer and to measure their periods of survival since the onset of the disease. Accordingly, the National Cancer Institute reports the number of cancer survivors, defined as the total number of persons living with cancer 5 years after diagnosis, and the probability of surviving 5 years since the onset of their disease ($_{y+5}S_c$) among those incurring the disease in a specified period, say, a 1-year period ($_yI_c$):

$$\text{Proportion of survivors after 5 years} = {}_{y+5}S_c \div {}_yI_c \tag{5.9}$$

An increase in the number of cancer survivors tells nothing about success in treating cancer, because the increase may be a reflection of an increase in the population, the proportion elderly in the population, or higher cancer incidence rates. For example, the number of cancer survivors tripled between 1970 and 2000, while the (unadjusted) cancer death rate increased 23%. For this period the National Cancer Institute reported that the 5-year cancer survival rate increased from 50% to 64%. The divergent paths of these rates can be explained in terms suggested above.

[4] Stenosis is the blockage of a blood vessel, typically an artery, with possible negative consequences in the form of angina (pain in chest), claudication (pain in calf or foot), heart attack, or stroke. Angioplasty is the procedure of catheterizing (coronary, peripheral, abdominal) arteries in order to reduce the degree of stenosis and improve the flow of blood though the arteries.

The survival period of persons with cancer may also be based on life tables designed to take account of the joint risks of mortality and incurring cancer. Such tables provide measures of the chance of ever dying from cancer, average years lived with cancer at each age, and related measures. They resemble the multiple-decrement life tables on heart disease mentioned earlier in this chapter and can be constructed by methods described in Chap. 8 as well.

Case-Fatality Rate

The case-fatality rate is a type of cause-specific mortality measure, with the population exposed to risk being those who already have the specific health condition. It answers the question, on a period basis, what percent of persons having a disease, injury, or impairment die from it during a given year? It is conventionally computed as the ratio of deaths from the disease, injury, or impairment during the year to persons who have the disease, injury, or impairment at the middle of the year or, alternatively, at any time during the year, per 100.

$$\mathrm{CFR} = (\mathrm{D_c} \div \mathrm{C}) * 100 \tag{5.10}$$

where *CFR* is the case-fatality rate, D_c represents the number of deaths from the cause during a year, and C the number of persons having the health condition at midyear. For example, the case-fatality rate for HIV disease in the United States in 2003 is calculated as,

CFR = (Deaths from HIV/AIDS disease ÷ number with HIV/AIDS at midyear) × 100

$$\mathrm{CFR} = (13{,}658 \div 41{,}831) * 100 = 32.7\%$$

A study of the "Global Burden of Tuberculosis" estimated the global case-fatality rate for tuberculosis as 23% (Dye et al. 1999). The case-fatality rate for tuberculosis exceeds 50% in African countries with high rates of HIV infection.

The case-fatality rate can be structured in cohort form as well as in period form. In cohort form it would be defined as the proportion of persons having a disease, injury, or impairment at the beginning of a year who die from that disease, injury, or impairment during the subsequent year. The case-fatality rate in cohort form tends to be lower than the case-fatality rate in period form because the base population tends to be higher at the beginning of the year.

A reduction in the case-fatality rate of a disease tends to increase the prevalence of the disease, particularly for chronic conditions. Specifically, if a larger proportion of persons are saved from death from a particular chronic disease, a larger proportion of persons will have the disease over a particular period and will need care. Such relations have important implications for the analysis of the relative costs of death and illness, and hence the costs of disease-treatment and disease-prevention programs.

Comorbidities

The frequency of chronic health conditions rises sharply in later life so that such conditions are very common among elderly persons. An elderly person often has more than one chronic health condition. This is characteristic of persons over age 80. Combinations of chronic health conditions are known as comorbidities. One simple way to compare the health status of two population groups is to compare the distributions of the populations by the number of morbid conditions affecting the members of each population. Populations may be distinguished by the proportion of the population that has none, one, two, or three or more chronic health conditions.

Several indexes and scores have been developed by epidemiologists on the basis of comorbidities to measure the probability of future hospitalization and death of individuals from chronic conditions. The higher the cumulative score the greater the probability of hospitalization or death. The most common among these indexes is the Charlson Comorbidity Index; others are the Davies Score and the Chronic Disease Score. The Charlson Comorbidity Index uses 19 categories of morbidities (asked as binary variables) and age, each associated with weights representing the adjusted risk of 1-year mortality. The Davies Score uses comorbidities like the Charlson Index but without age. The Chronic Disease Score is an aggregate measure of comorbidities based on current use of medications.

Data on morbidity and comorbidities may be obtained from self-reports in health surveys and respondent reports of current use of medications, as well as from an analysis of the multiple causes of death reported on death certificates. As the reader may recall, a death certificate may include, in addition to the underlying cause, contributory causes and associated causes. The latter sources were used by Ruzicka et al. (2004) in their study of the comorbidities associated with suicide in Australia during 1997–2001 and by Stallard (2002) in his study of the accuracy of the reporting of cause of death in the United States.

Age-Adjusted/Age-Standardized Measures

Generally the measures of morbidity for the total (all-ages) populations I have discussed so far are crude measures in that they are affected by the underlying age distribution but take no explicit account of it. In order to increase the comparability of these measures in analyzing health differences between populations from one date or one area to another, they may be recalculated as age-adjusted measures of morbidity. As we saw in Chap. 3, age-adjustment or age-standardization can be used to derive such age-adjusted measures for mortality and to disaggregate the difference between two crude rates into the contribution of differences in age composition and the contribution of differences in age-specific rates. This type of analysis can be applied to any general measures that differ either in space or time, including morbidity measures, using disaggregated data.

Like age-adjusted measures of mortality, age-adjusted measures of morbidity are summary measures of morbidity that represent weighted combinations of age-specific ratios or rates of disease, injury, or impairment and that employ as weights the same age distributions for all the populations being compared. One can calculate both general prevalence ratios and general incidence rates adjusted for age. The methods of computation and the limitations of such age-adjusted measures are the same as those described for age-adjusted death rates in Chap. 3. See that chapter for a detailed discussion of this topic.

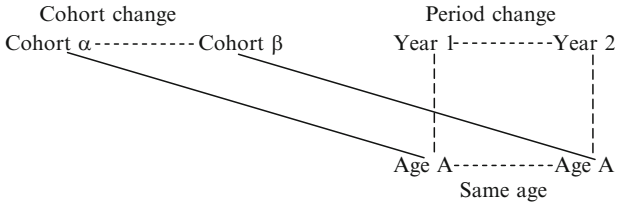
Age, Period, and Cohort Effects on Health Measures

A different type of analysis, known as age-period-cohort (APC) analysis, may be applied to a matrix of ratios or rates for age groups for a series of years, in order to determine the separate contributions of so-called age effects, period effects, and cohort effects to the levels and patterns of the rates and their changes. A common interest in APC analysis is to determine the role of these three factors in the difference between the rates for a health condition for different age groups at the same date or the same age group at different dates. For example, one might wish to determine the contribution of the three factors to the rise in the death rate from heart disease between ages 50–54 and ages 80–84 in the year 2000 or to the decline in the death rate from heart disease at ages 65–74 between 1975 and 2005.

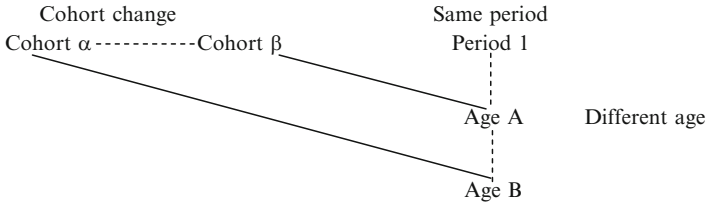
Period effects refer to the influence of events occurring in specific calendar years (e.g., a war, an epidemic, major political event) on the phenomenon studied. Age effects refer to the influence of different ages on the variable of interest. Cohort effects refer to the influence of the characteristics and experience of different birth cohorts, especially early life events, on the phenomenon studied. Interpretation of age-specific changes over the age cycle, or at a particular age over a series of years, in terms of one of these factors (e.g., the age effect or the period effect) or even two of these factors (e.g., the age and period effects) without regard to the third may lead to an erroneous interpretation of the changes; the effects of all three factors must be considered jointly. Figure 5.1 is a stylized sketch of the contribution of age, period, and cohort effects to the differences or changes in age-specific morbidity ratios.

Consider a standard APC table consisting of cross-sectional data for 5-year age groups at a series of dates 5 years apart, or data for 10-year age groups for a series of dates 10 years apart, specifically a matrix consisting of a series of age-specific prevalence ratios for persons with heart disease in 10-year age groups (20–29 to 80–89) for decennial years from 1950 to 2000. Note that the width of the age group must correspond to the time interval. This array represents 6 different (1-year) periods of observation, 7 different (10-year) age groups, and 12 different (10-year) birth cohorts (Exhibit 5.1). The data might come from periodic sample surveys providing cross-sectional or panel data, disease registries, or clinical trials. The analyst's attention may be focused on one age group, one period, or one birth cohort, but

(1) Change in age-specific morbidity ratios at age A between year 1 and year 2:



(2) Difference between age-specific morbidity ratios at age A and age B in year 1:



(3) Difference between age-specific morbidity ratios between age A in year 1 and age B in year 2 for the same cohort:

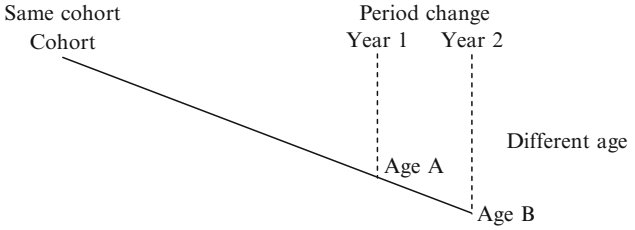


Fig. 5.1 Stylized sketch of the contribution of age, period, and cohort effects to the change in age-specific morbidity ratios (1) at age A between year 1 and year 2, (2) between ages A and B in year 1, and (3) between age A in year 1 and age B in year 2 (Note: Two of the factors must be different to provide meaningful comparisons. Not all combinations are possible; e.g., same cohort, different age, same period; same cohort, same age, different period; and different cohort, same period, same age)

all three effects are embedded in each of them. Even if the focus is on the data for one period, the concern would be the embedding of age and cohort effects in the period data as well as the period effects.

Before illustrating the role of these factors and considering how to measure them, it is useful to call attention to a few further complexities in the interpretation of the data with respect to APC effects. Real cohorts, whether defined in terms of year of birth or other sociodemographic characteristic, tend not to be fully "closed," as is often assumed, but to change as a result of the entry or departure of various members. That is, their characteristics and their composition change, most notably from losses through death, but also from migration, socioeconomic status, and other factors. These demographic changes in the cohort may be correlated with the

	Period (P)					
	1950	1960	1970	1980	1990	2000
Age (A)	P_1	P_2	P_3	P_4	P_5	P_6
20–29 A_7	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
30–39 A_6	C_6	C_7	C_8	C_9	C_{10}	C_{11}
40–49 A_5	C_5	C_6	C_7	C_8	C_9	C_{10}
50–59 A_4	C_4	C_5	C_6	C_7	C_8	C_9
60–69 A_3	C_3	C_4	C_5	C_6	C_7	C_8
70–79 A_2	C_2	C_3	C_4	C_5	C_6	C_7
80–89 A_1	C_1	C_2	C_3	C_4	C_5	C_6

Exhibit 5.1 Age-period-cohort table covering the periods 1950–2000 and ages 20–29 to 80–89
Notes: *A* age group 1–7, *P* period 1–6, *C* cohort 1–12. None of the cohorts are shown in entirety in this exhibit

independent variable under study and cumulatively may be so great as to give the cohorts an essentially different composition at their later years as compared with their earlier years. Selective mortality and changes in socioeconomic characteristics may have collaborated to produce a different "corps" of individuals with somewhat different characteristics in the later years of adult life from that at its beginning (e.g., different membership with changed health and socioeconomic composition).

Another confounding influence on the "dependent variable" that applies in particular to panel data is the possible change in the variable that results from repeated participation of the same persons in the survey. Such panel conditioning affects the data increasingly through the various waves of the study as additional collections of the data are made. Panel conditioning of effects may result from the participant's greater understanding of the questions, greater knowledge of the subject matter, or greater readiness to respond truthfully or untruthfully. Repeated cross-sectional surveys do not present this problem because the respondents are different in each survey.

Age Effect

With advancing age among adults in the same calendar year, health surveys and vital statistics show increasing levels of morbidity, disability, and death. A large share of persons from ages 60–64 years on report that they are afflicted with at least one chronic disease. Severe disability is reported by many at the advanced years of life, that is, at ages 85 years and over. Multiple chronic diseases and continuing severe disability are characteristic of the advanced years of life, even though some individuals experience only a brief period of end-of-life disability. Yet, the health changes over the age scale in any given year may exaggerate the extent of the rise in negative health indicators with age for a real cohort because the persons at the higher ages represent survivors of persons born in an earlier year when morbidity and mortality may have been much higher than at present. Hence, while

the health differences from age to age may appear to reflect the age effect only, these differences cannot be explained by the age factor alone; cohort effects are also embedded in the age figures. Period effects are also involved. These additional effects may be inconsequential or they may be quite important.

An event of massive impact (e.g., war, epidemic, depression) may have occurred and affected the level and pattern of the rates in a given year differently. Moreover, each of the many birth cohorts represented in an annual age distribution of rates or ratios has different demographic characteristics, early life experiences, and life histories. In short, when the levels of some measure of morbidity or mortality for a schedule of ages at the same date are compared, these age differences require explanation not only in terms of age effects but cohort and period effects as well.

Period Effect

Major historic events as well as some extraordinary current events have had their distinctive effects on the annual record of morbidity and mortality. Such period effects explain some of the differences between age-specific ratios and rates at the same age at different dates. They may be the gradual social changes that occur with time, a major economic, political, or social trauma, such as the influenza epidemic of 1918, the Great Depression of the 1930s, World War II, the Vietnam war, the explosion of the nuclear power plant at Chernobyl in 1986, or a great political or economic transformation, such as the dissolution of the Soviet Union, the reunification of Germany, or the introduction of Social Security and Medicare in the United States.

Kleinbaum et al. (1982) illustrate the period effect by citing the decline in alcohol consumption from 1920 to 1933 and the decline in mortality from liver cirrhosis during this period following the passage of the Prohibition Amendment to the Constitution. The cirrhosis death rate rose again in the 1930s after the Amendment was repealed and alcohol consumption resumed.

The events may affect the rates for all age groups uniformly or in a variable way. The effects may differ, not only because of the individuals' ages but also their prior health status, socioeconomic status, race, ethnic group, and other factors. The ages in any year represent many different birth cohorts with different experiences at birth and in childhood that affect their adult responses to period influences.

The major events tend to have some immediate effects and some long-range effects on the health of the individuals alive in the years they occurred. The events may directly affect individuals' health or affect their access to health services and social support systems. War, particularly, tends to affect the food supply and social support system of the entire population, but war mortality is usually concentrated among the youth. We recognize readily the variable period effects of an epidemic or accidents on different age groups. Even though historic events in a given year tend to affect the different age groups differently, there may be little evidence to show this variation, so that often the period effect is assumed to be age-neutral.

Cohort Effect

Each birth cohort is stamped with early life influences that it carries with it through the life course. First, birth cohorts differ in size. Next, cohorts differ in their characteristics at the time of birth, their experiences in the early years of life, and their experiences in later years as they age. Note that the experiences of one cohort are essentially independent of the experiences of the many other birth cohorts (i.e., age groups) in the calendar years through which it passes. Changes in membership within birth cohorts that tend to distinguish them from one another also occur. Cohorts gain members through immigration and lose members through emigration and death in different degrees. The exact extent of the rise or fall of illness with age from one calendar year to another for any cohort partly reflects the loss of persons with poor health prospects early in the cohort's lifetime and the retention of persons with lower health risks to later ages in a later year. Each cohort is heterogeneous *de sui generis* with respect to the "frailty," or health conditions, of its members. In general, persons with greater health risks are removed from a cohort as it ages, creating a so-called "healthy survivor effect."

The cohort effect is illustrated by the early life imprint of an unhealthy condition that persists in some form through the life of the cohort. Finch and Crimmins (2004) demonstrated the considerable influence of early-life sickness on later-life morbidity and mortality, particularly the effect of the exposure to infectious diseases and other sources of inflammation during childhood. They describe the work of Jones (1956), who concluded, on the basis of an analysis of mortality data for Sweden in the eighteenth and nineteenth centuries, that "mortality curves across the life span were displaced to lower values in more recent cohorts" and who hypothesized therefore that the "physiological age of each new generation is remaining more youthful at the same chronological age." Cohort effects were described as early as 1934 by Kermack et al. They observed that, as mortality at younger ages improved in successive cohorts in England and Sweden, the adult survivors in those cohorts also had lower mortality. They inferred from this observation that maternal health and the child's environment were major determinants of health in later life and that the health experience of the cohort was carried by it throughout its life. An epidemic of an infectious disease that afflicts many pregnant women (e.g., rubella) and that can be passed on to their children with harmful side-effects (e.g., congenital deafness) creates the circumstances for a generational cohort effect (Kermack et al. 1934a, b).

In an important finding, Frost (1939) showed that the rise in the modal age of mortality from tuberculosis in the period data for a series of years should not be interpreted as a real rise in the age of tuberculosis mortality. Rather, he observed that the age pattern of cohort data from cohort to cohort hardly varied and that the level of the cohort data showed a steady decline in tuberculosis with each later cohort. Another example of the confounding of age and cohort effects is provided by Dorn and Cutler (1959) relating to lung cancer mortality among U.S. white males in 1949–1950. The mortality rate is seen to increase up to age 65 and to decrease after age 70 (Fig. 5.2). If this is accepted as the pattern of age-specific change in lung cancer mortality, we would conclude that at the older ages there is a

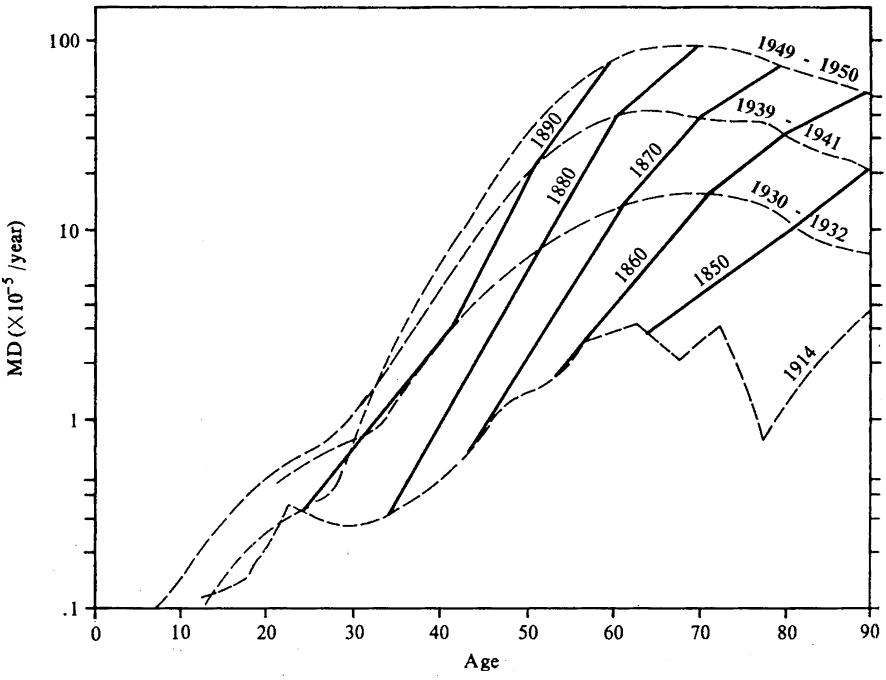


Fig. 5.2 Death rates from lung cancer (MD) for U.S. white males, by current age, 1914–2000, and by birth cohort, 1850–1910 (Source: Dorn and Cutler (1959). See also Kleinbaum et al. (1982), Figure 7.3, p. 134)

reduced tendency to acquire the disease because of increased resistance or changes in the environment. The period data show the same pattern at some earlier dates. If, however, we examine similar data for a series of birth cohorts covering the years from 1850 to 1890, we find a steady increase with age in the mortality rate. The cohort data do not show a drop in lung cancer mortality at the higher ages. Moreover, the more recent cohorts show higher rates than the earlier cohorts. The lethal role of the increase in cigarette smoking up through the first half of the last century is the likely cause of this increase. An extension of the Dorn/Cutler data to 2000 reflects the turnaround in tobacco use since 1950.

These illustrations make it quite clear that focusing only on period data on a health characteristic for age groups could result in a misleading interpretation of the basis of the age variation of the health characteristic.

Estimating Age, Period, and Cohort Effects

The estimation of age, period, and cohort effects on age variations in health characteristics is an important and informative analytic task, whether for a wide

range of ages or a single age. The analysis is complicated by the fact that the three types of effects – age, period, and cohort – must be involved in the analysis simultaneously, and it is difficult, if not impossible, to disentangle them without employing strong qualifying assumptions. As is made evident below, there is an "identification" problem, which precludes a definitive solution and allows only "scenarios" as ways of handling the analysis. In this section I consider some of the issues in the measurement of the separate effects, but the description of the attempts at formal measurement are beyond the scope of this book. References to such measurement methods are cited in the text and the list of references.

Analysts disagree as to whether there are two or three distinct effects and whether the three effects can be disaggregated (Mason and Fienberg 1985; Halli and Rao 1992; Glenn 2005). Most analysts deal with the issue as if there are three effects even though information on two of the factors provides information on the third. Thus, if you know the period and the age of a rate, you know the cohort with which it is associated, i.e., cohort = period – age. Since any one of the factors can be viewed as a form of interaction between the other two factors, all three factors cannot vary independently. Thus, any two of the three effects constrain the third; e.g., the age and period effects taken together constrain the cohort effect (Fig. 5.1). We are presented, therefore, with a redundancy or over identification problem.

Further issues in disaggregating a matrix of rates for APC effects. The general age pattern for incidence rates or prevalence ratios can usually be detected by observation of one or more of the period sets of data in the matrix, but the precise age pattern cannot be determined in this way. We can readily note the approximate effect of World War I, World War II, and the influenza epidemic of 1918 on the death rates for some age groups in the years in which these events occurred. The major historical (period) events, especially the catastrophic ones, tend to be relatively unique and can be viewed as anomalies for which appropriate allowances in general mortality and morbidity rates can be made. Making precise allowances for the effects of unusual events on highly disaggregated data, such as individual sex-age groups, for a specific health condition (e.g., diabetes prevalence ratio) is much more difficult, if not impossible. Period effects on age data may be roughly similar from age to age but they cannot be measured closely. The underlying level and pattern of age data are blurred by a combination of period and cohort effects.

In undertaking a disaggregation of a matrix of rates for APC effects, first organize the data, then determine the algorithm to be used, and finally solve for the parameters. The first task is to organize the set of rates in such form that they show sufficient age detail at frequent enough intervals for the matrix to be reasonably dense with observations and for the cohorts to be aligned with the age groups in the appropriate years. A set of rates for 5-year age groups at 5-year time intervals over a period of 30–40 years is one acceptable way of organizing the data. In practice, the data collected on health variables tend to present a problem in this regard and have to be smoothed or interpolated (Brady and Elms 1999). The data may be tabulated in one way in one year and in another way in another year, and single-year-of-age

tabulations may be irregular because of the sample size. The reader is advised to refer to other publications for guidance in handling the problems of smoothing and interpolating raw data.

The second task is to decide on a design for disaggregating the total effect into APC effects. Refined procedures for doing this currently involve determining the parametric design, but simpler procedures may be considered. One simple approach to the problem is to derive two generalized age patterns of the age-specific ratios or rates, one from the period data and one from the cohort data, using age data from the matrix over a number of years (Siegel 1993). The generalized age pattern for periods can be derived by averaging the age-specific rates or ratios for the same ages for a series of years. The generalized age pattern for cohorts can be derived by averaging the age-specific rates or ratios for the same ages for a series of birth cohorts. The generalized period and cohort patterns will be essentially the same and represent the age effect. Comparison of these generalized period and cohort patterns with the recorded sets of rates for the individual period data or birth cohorts provides a rough indication of period and cohort effects for each original array by age and by cohort. This type of analysis does not, however, provide precise information on these effects.

More refined approaches to APC analysis have been applied by many analysts. For example, some have employed various regression models, including linear, logistic, and Poisson regression models, in separating age-period-cohort effects in a wide variety of data.[5] Some analysts have sought to apply simple linear regression to solve the problem of separating age, period, and cohort effects on the dependent variable of interest, but it cannot be done by simple linear regression because of the perfect multicollinearity between any two of the effects and the third. This complication does not exclude the possibility of assuming a nonlinear relationship between the variables (Mason et al. 1973; Glenn 2005; Brady and Elms 1999). Even so, valid estimates can be obtained only with simplifying assumptions. Because of the identification problem, the disaggregation of the three effects is not solved by making the three factors the independent variables in a single regression equation or by using three regressions, each with two of the variables. One could set the value of the regression coefficient of one of the effects, for example the period effect, equal to zero, and attribute all changes to age and cohort effects in a single regression equation, but the validity of the results of this scenario would again depend on the validity of this assumption. One could design three regressions, fixing one effect in each regression equal to zero, and then evaluate the results of the three regressions on the basis of external information, seeking to determine which of the various solutions was most plausible or valid (Firebaugh 1997).

Hobcraft et al. (1982) and Halli and Rao (1992) describe a linear, additive regression model in which "dummy" variables for each of the effects are fitted to a logit transformation of the dependent variable; that is, they applied logistic regression. The logit transformation is $ln\ [p \div (1 - p)]$, where ln is the natural

[5] See Chap. 7 for further explanation of logistic and Poisson regression.

logarithm and p is the rate expressed as a proportion. In Halli and Rao's example of the use of logistic regression for the APC disaggregation, the results obtained include the proportion of explained variation in the overall rate for each of following models: age; period; cohort; age and period; age and cohort; period and cohort; and age, period, and cohort combined. In addition, age effects are obtained for each of eight age groups, period effects for each of seven time periods, and cohort effects for each of 14 cohorts. Carstensen (2005) applied Poisson regression and Yang et al. (2006) developed an "intrinsic estimator" in a complex nonlinear solution to the age-period-cohort problem.

In sum, analysts have devised a variety of complex methods to derive approximate solutions to the APC problem, even though it is generally recognized to be an insoluble problem unless drastic and generally unrealistic assumptions are made. Glenn (1976, 2005) maintains that APC models cannot be relied on to provide accurate estimates of age, period, and cohort effects. He maintains that analysts who are specialists in their subject field and understand their data can usually make valid inferences regarding the relative role of age, period, and cohort effects without employing complex methods of doubtful accuracy. Glenn suggests the use of applicable theory, external information, and some simpler statistical techniques to solve the APC puzzle. This could involve alternative regression models in which some *characteristics* of cohorts, ages, or periods are substituted for the cohorts, ages, or periods themselves in the regression equations.

Concepts and Measures of Functioning

Concepts and Classification of Functioning

I turn next to measures of health that focus on what one cannot do, or has difficulty doing, as a result of one or more health conditions. As noted previously, a health condition that limits functioning is termed a disability or, less commonly, impairment or handicap. Earlier measures of disability were based upon questions that inquired about limitations in performing one's "major activity" due to health reasons, such as limitations on holding a job or going to school, or inquired about "days of restricted activity" and "days of bed disability." Such questions on activity limitations were used to determine long-term disability. They were asked in decennial censuses and health surveys, such as the U.S. National Health Interview Survey and the Canadian Health Survey. In more recent years, a much greater variety of questions are used, including questions on personal-care disability ("activities of daily living" or ADL's), physical disability/domestic disability ("instrumental activities of daily living" or IADL's), work/employment disability, transportation (go-outside-the-home) disability, communications/sensory disability, and mental disability.

International Classifications

The World Health Organization's (WHO) *International Statistical Classification of Diseases and Related Problems* (ICD)[6] was considered to be unsatisfactory as a classification of the functional states of persons resulting from a health condition. This classification deals with the etiology of health problems, not with their practical effect on the individual, and does not provide sufficient detail on their severity to make correct inferences regarding the effect on functioning. Accordingly, the WHO proposed the compilation of a manual on the consequences of diseases and the result was the publication of the *International Classification of Impairments, Disabilities, and Handicaps* (ICIDH) in 1980. This manual standardized concepts in the disablement process. Because disability is viewed as a consequence of disease and the identification of a disease does not adequately portray the full consequences of the condition, the ICIDH was to be an extension of the ICD scheme.

The classification distinguished three types or levels of health-related restrictions on activity: Impairment, or injury level; disability, or functional level; and handicap, or social level. In the ICIDH classification, an impairment is the loss or abnormality of a psychological function, or loss or abnormality of a physiological or anatomical structure or function; a disability is an inability to perform an activity in the range considered normal for a person; and a handicap is an inability to perform one's social role(s) due to an impairment and/or disability (WHO 1980).

The original ICIDH scheme has been difficult to use for a number of reasons: (1) overlaps between the various concepts, especially between disability and handicap, (2) lack of clarity about the specific meanings associated with the categories, (3) difficulty in applying the classification scheme to various theories and models of disablement, (4) difficulty in adapting currently collected health data to conform with the ICIDH framework, and (5) failure of the ICIDH scheme to encompass the possibility for reversals, either reductions in severity or recovery. The ICIDH scheme has also been criticized because it does not allow for variations with age in the standards for functioning. However, many gerontologists would consider a uniform standard for all adults appropriate for functioning, as for health.

Since 1995 there have been efforts to revise the ICIDH framework and definitions. The ICIDH work was viewed as experimental, and in 2001 the WHO adopted a new classification. The new version of ICIDH was renamed the *International Classification of Functioning, Disability, and Health* (ICF) and designated ICIDH-2. ICIDH-2 represents a restructuring of ICIDH to cover both the positive and negative aspects of functioning, and focuses on: (1) body functions and body structures, (2) activities performed by an individual, and (3) participation in, and classification of, the areas of life in which an individual is involved. In addition, environmental factors are also part of the new classification scheme. "Environmental factors" is broadly defined to encompass "assistance" products and technologies, the natural environment and human changes in the environment, support and relationships

[6]The latest version is: World Health Organization (1992, 1993, 1994).

provided by people and animals, and formal services, governmental systems, and policies that affect the disabled. In sum, ICIDH-2 makes it possible to take into account the social and physical environment affecting a person's functioning.

The specific definition of disability can vary greatly depending on the social context, and the choices made with regard to its definition can result in very different estimates of the disabled population (Ravaud et al. 2002). One can employ an objective or subjective basis for making this determination. A subjective measure examines disability from the point of view of the individual and tries to determine whether the individual considers him/herself as having a disability or not. A more objective measure is the need for assistance to perform certain tasks. Disability can be measured, then, as did Ravaud et al., in their comparative analysis of various concepts of disability, by the (1) presence of functional limitations, (2) need for assistance, (3) self-reporting, (4) administrative recognition of a disability (e.g., participation in a program open only to disabled persons), and in other ways. Definitions of disability also differ in whether or not they allow for use of technological assistance devices and adaptations of the built environment.

Securing accurate and comparable data on disability for the countries of the world is a daunting task. Differences in cultures, concepts of functionality, technological development, and methods of obtaining the data on disability preclude the ready compilation of accurate and comparable data on disability prevalence among countries. The definition of disability can be controversial. One culture may interpret a condition as a disability that another may interpret as falling within the standard range of functioning. The situation is even more complicated in the mental health area. In the effort to secure better data on disability, the United Nations Statistics Division published a set of guidelines regarding statistical concepts and methods of data collection on disability in 2001, under the title *Guidelines and Principles for the Development of Disability Statistics*. The UN Statistics Division has also designed a questionnaire on disability statistics that it has been testing, along with tabulation and publication plans for use with the questionnaire.

United States and European Concepts

Several different definitions of disability are employed by the U.S federal government in censuses, sample surveys, and public entitlement programs. Among these are the varied definitions used by the National Center for Health Statistics, the Census Bureau, and the Social Security Administration. All definitions emphasize functional limitations. The National Health Interview Survey defines disability as "any temporary or long-term reduction of a person's activity as a result of an acute or chronic condition." The recent censuses of the United States (1980, 1990, and 2000) and the American Community Survey had questions on sensory impairments, physical disability, and activity limitations with respect to work, use of transportation, or personal care resulting from a physical, mental, or emotional condition lasting 6 months or more. The American Community Survey in 2010,

which will be adopted as part of the 2010 census, has essentially the same questions on disability as the 2000 census, but with some small modifications.

Specifically, the year 2010 U.S. census/the American Community Survey has the following questions on disability:

17. a.	Is this person deaf or does he/she have serious difficulty hearing?	Yes	No
b.	Is this person blind or does he/she have serious difficulty seeing even when wearing glasses?	Yes	No
18. a.	Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering, or making decisions?	Yes	No
b.	Does this person have serious difficulty walking or climbing stairs?	Yes	No
c.	Does this person have difficulty dressing or bathing?	Yes	No
19.	Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting a doctor's office or shopping?	Yes	No

In other words, these questions try to ascertain the types of disability that have been called communication/sensory disability (17a, b), mental disability (18a), physical disability/domestic disability/"instrumental activities of daily living" (18b), personal- or self – care disability/"activities of daily living" (18c), and transportation/ "go-outside-the-home" disability (19).

To secure information on disability in the Survey of Income and Program Participation (SIPP), the U.S. Census Bureau asked persons 15 years and over if they met any of the following criteria: Had difficulty in performing such activities as seeing, hearing, speaking, lifting or carrying, using stairs, and so on.; had an ADL limitation (activities of daily living/personal-care activities); had an IADL limitation (instrumental activities of daily living/simple activities inside and outside the home); had a mental or emotional condition such as a learning disability, or mental retardation; were limited in ability to do housework; were limited in ability to work at a job or business; or were in receipt of federal benefits for work disability.

The disability data in various national surveys and administrative records can be exploited to define degrees of severity of disability, such as severe, moderate, or light. The data on ADLs and IADLs in the NLTCS sample, which is drawn from the Medicare files, were used by Manton et al. (2006) to define moderate disability. This definition was having any health-related difficulty in performing at least one "activity of daily living" (ADL) or "instrumental activity of daily living" (IADL) for 90 or more consecutive days, or currently residing in an institution providing medical services. All other situations were considered nondisabled.

The definitions of disability employed for administrative uses by government agencies tend to differ from those asked in censuses and surveys. Consider the definition of disability employed by the U.S. Social Security Administration (SSA) for purposes of entitlement to monthly benefits. The SSA defines disability as "inability to engage in any substantial gainful activity (SGA) by reason of any

medically determined physical or mental impairment that can be expected to result in death or that has lasted or can be expected to last for a continuous period of not less than 12 months."[7]

The European Community Household Panel (ECHP), a longitudinal survey covering many aspects of daily life including health for the EU Member States, compiled comparable data on disability for 12 of the Member States in 1994. The question on disability in English was, "Are you hampered in your daily activities by any chronic physical or mental health problem, illness, or disability?" with possible responses, "Yes, severely," "Yes, to some extent," and "No" (Robine et al. 2001) Comparable data could not be secured for the institutional population, and it was assumed that the years lived in an institution were partitioned into years lived with and years lived without disability in the same proportion as for those living in ordinary households.

Assessing Disability

Activities of daily living or ADLs, also known as the Katz measure of disability, is the most widely used measure to assess disability. It evaluates the ability to perform six basic personal-care tasks, namely, dressing, grooming, eating, bathing, toileting, and transferring into and out of bed, without assistance (Katz et al. 1983). They represent a more limited level of functioning than instrumental activities of daily living, or IADLs, which assess the ability to use the telephone, prepare meals, do the laundry, shop for groceries or clothing, manage one's finances, take one's medication, and do other light housework. Sometimes a distinction is made between basic activities of daily living (BADL) and advanced activities of daily living (AADL). Branch et al. (1984) employed six items in their functioning index, four from the Katz ADL list (bathing, dressing, eating, and transferring) and two of their own (personal grooming, such as shaving, brushing hair, and trimming toe nails; and walking across a small room).

A measure developed by Rosow et al. (1966) assesses the ability to perform three tasks requiring mobility and strength; it includes walking up and down stairs, walking half a mile, and doing heavy work around the house. Another measure, the Nagi (1976) measure, assesses the ability to perform four basic tasks without difficulty: pushing and pulling large objects; stooping, crouching, or kneeling; reaching or extending arms above shoulder level; and writing or handling small objects. In their analysis of trends in rates of onset of disability in the 1980s and 1990s, Wolf et al. (2007) express the view that the Katz measure, the Rosow

[7]The SSA definition requires the impairment to be of a degree of severity that renders the individual unable to engage in any kind of substantial gainful work that exists anywhere in the national economy. If the determination of disability cannot be made on the basis of medical evidence only, consideration is given to the person's age, education, and work experience.

measure, and the Nagi measure satisfactorily cover the entire disability profile in older age for use in trend analysis of disability. Many other scales and measures of disability are found in public health and gerontological publications, however.

As stated, there has been a general shift away from measuring disability based upon general assessments of activity limitations affecting one's major activities. Clearly questions on activity limitations are unsuitable for portions of the population that do not engage in a major activity such as an occupation (e.g., the unemployed, persons who have retired). More recent research has utilized specific activities of daily living (ADLs), based upon the work of Katz et al., and certain more complex routines associated with independent living, the so-called instrumental activities of daily living (IADLs). Sometimes, the level of disability is defined in terms of the number of ADLs and IADLs that an individual is unable to perform. It is commonly assumed that the residents of nursing homes are all severely disabled (that is, that they suffer from several ADLs and IADLs), inasmuch as it is not usually practical to question them individually and they are known to be sharply limited in their functioning.

Population Measures of Disability

Disability Prevalence Ratios and Disability Incidence Rates

Distinguishing prevalence measures from incidence measures, we can compute general and age-specific disability prevalence ratios, and general and age-specific disability incidence rates. Disability prevalence ratios are calculated as the percent of the population 5 years and over or 16 years and over, or the percent of the population in an age group, with a disability. One specific form of such a measure is an age-cause-specific disability prevalence ratio, for example, the percent of the population in an age group with a chronic disease that limits functioning, as shown in Table 5.5. General disability prevalence rates and general cause-specific disability prevalence ratios may also be converted to age-adjusted ratios.

A (central) disability incidence rate for a year is the ratio of the number of persons that incurred a disability in a year to the midyear total population (per 100) and is usually based on a first-time diagnosis of a specific condition that limits activity. The central rate for a specific age group can be converted into a cohort rate, or probability, on the basis of an assumption of rectangularity in the distribution of deaths and disablements during the year and within the age according to the formula:

$$\mathrm{D_iR^p} = \mathrm{D_i} \div (\mathrm{P} + 1/2\mathrm{D_e}) \tag{5.11a}$$

$$\text{or, more exactly,} \quad \mathrm{D_iR^{pz}} = \mathrm{D_i} \div (\mathrm{P_{nd}}1/2\mathrm{D_e} + 1/2\mathrm{D_i}) \tag{5.11b}$$

where all elements refer to age a and $D_i\,R^{pz}$ represents the probability of incurring a disability during the year, D_i disablements during the year, P_{nd} the midyear

Table 5.5 Age-adjusted prevalence ratios for limitations in usual activities, ADL/IADLs, and work activity, by sex and by age, for the United States: 2007

	18 years and over			Age			
Limitations in activities	Both sexes	Male	Female	18–44	45–64	64–74	75+
Limitations in usual activities							
Limited	14.3	NA	NA	6.3	16.1	27.2	44.8
Limited due to 1 or more chronic conditions	13.8	NA	NA	6.0	15.7	26.5	43.3
Limitations in ADL/IADLs							
ADLs	2.0	1.6	2.3	0.5	1.7	3.3	11.0
IADLs	3.9	3.0	4.6	1.2	3.6	6.3	20.0
Limitations in work activity							
Unable to work	5.8	5.7	6.0	3.1	9.6	12.5[a]	
Limited in work	3.0	2.9	3.2	1.9	4.2	7.8[a]	

Source: U.S. NCHS, *Vital and health statistics* series 10(238), 2008
Percentage of population 18 years and over. Figures are standardized on U.S. 2000 census-based population
NA not available
[a]Ages 65–69

nondisabled population, and D_e deaths of persons during the year. Formula 5.11b expresses the probability that a nondisabled person at exact age a at the beginning of a year will become disabled between exact age a and exact age $a + 1$ during the year. (There is a slight overlap in the terms of the formula inasmuch as some persons who became disabled during the year also died during the year.)

The probability of becoming disabled at any age may be approximated from data on the percent disabled by single ages at a single date by differencing the percents at two successive ages and dividing the difference by the percent at the earlier age. The formula is, then,

$$\Pr(\text{becoming disabled between age 1 and age 2}) = (d_{r2y} - d_{r1y}) \div d_{r1y} \quad (5.12)$$

where the percents disabled (d_r) relate to two successive ages (1 and 2) in the same year (y).

Average Age of Disabled Persons, Age at Onset of Disability, and Duration of Disability

The reader will recall the earlier discussion of the average age of onset of a disease, the average duration of a disease, and the compression of morbidity. We consider here the average age of onset of disability, the average duration of disability, and the compression of disability. Disability is treated here as a distinctive type of morbidity, for example, a chronic disease that limits functioning. Disability and disease may move at different rates and even in different directions.

Since disability is often mentioned in legislation and the federal government may be paying for disability support, the age of first occurrence of a disability has important financial implications for the government. It affects the costs of Social Security, Medicare, and all health services. A small upward shift in the age of onset of Alzheimer's disease or other chronic disability, for example, could translate into savings of billions of dollars of public funds. On the other hand, a reduction in the case-fatality rate of a chronic disability like Alzheimer's disease without a corresponding rise in the age of onset would have the opposite effect.

For ascertaining the average age of disabled persons, one can simply calculate the median or mean age of the disabled population taken from census or survey data. For a measure of the average age of disabled persons independent of the population age distribution, calculate the median or mean age from the age-specific percents disabled.

One can sometimes secure data from a survey or from administrative records on the number of persons becoming disabled during a period. If such data are tabulated for ages, one can determine the mean or median age of the disablements (i.e., average age at onset) directly. If, on the other hand, only the numbers of disabled persons for age groups are available for various dates, an alternative method, one normally used to estimate net migration and net retirements during a period, can be employed to measure net disablements for the intervening periods and, from these figures, the average age of the onset of disability. More specifically, the method consists in estimating the number of disablements for birth cohorts in groups as a residual on the basis of the number of disabled persons classified by age, converting these cohort figures into estimates of disablements for the conventional age groups, and then calculating the mean or median age for the time interval. The method is spelled out in Appendix 5.1.

The age of onset of disability can also be measured as the age at the first decile or the mean age minus two standard deviations from the mean of the distribution of disabled persons. If, further, we can assume, as is generally true, that chronic disability lasts from age at onset to death, then we can compare the change in the duration of disability with the change in life expectancy and determine whether compression or expansion of disability is occurring. For compression of disability to have occurred during a period, age of onset of disability would have to have risen more than life expectancy during the period. Note, however, that it is possible for disability compression to occur while morbidity expansion is also occurring. This may have happened in the United States during the decades of the 1980s and 1990s, when chronic disability declined (Fries 2003; Freedman and Martin 2002; Manton et al. 2006). By comparing the relative change in age-specific disability ratios and in age-specific death rates, Fries (2003) sought to demonstrate that disability compression has been occurring in the United States since the early 1980s. These are important positive outcomes from a personal and social point of view: Patients suffer from their illness a shorter period of time before death and, potentially, society has a smaller health bill.

Other Measures of Functioning

In addition to the measures of disability/functioning based on the definitions given earlier, there are measures that are based on the concepts of limited functioning and restricted activity, and still others that may be described as functioning-related measures, such as those pertaining to frailty and long-term care.

Restricted Activity

Measures of limited functioning describe the frequency of restricted activity. A selected list of such measures is as follows:

Number of days of restricted activity associated with acute conditions per 100 persons, or number of days of restricted activity per 100 persons with acute conditions, per year.
Average number of days of restricted activity associated with acute conditions per acute condition in a year
Number of days of restricted activity associated with chronic conditions per 100 persons, or number of days of restricted activity per 100 persons with chronic conditions, per year
Average number of days of restricted activity associated with chronic conditions per chronic condition in a year.

Other, more specific, measures of restricted activity are defined in terms of days of work-loss or school-loss, and bed-disability days. A work-loss day or a school-loss day is a day on which one would have worked or attended school but did not do so for a whole day because of an illness or injury. A bed-disability day is a day on which a person stays in bed for all or most of the day because of an illness or injury. Here are some measures based on these concepts:

Days of bed disability associated with acute illness in a year per 100 persons per year.
Days of bed disability associated with an injury per 100 persons per year.
Days of bed disability associated with acute conditions (illnesses plus injuries) per 100 persons per year.
Days of bed disability associated with chronic conditions per 100 persons per year.
Days of work loss associated with acute conditions per 100 currently employed persons per year (restricted to persons 18 years and over).
Days of work loss associated with chronic conditions per 100 currently employed persons per year (restricted to persons 18 years and over).

Other Functioning-Related Concepts and Measures

The concepts of frailty, dependency, and long-term care all suggest physical dependency on others for care, continuously or occasionally.

Frailty. Understanding the concept of frailty requires a different paradigm for measuring functioning. Frailty can be defined simply as physical weakness and can be measured on the basis of various performance tests. Such a definition is inexpensive to apply and easy to interpret. On the other hand, it is too narrow and fails to capture the multisystem nature of the frailty syndrome. Hence, some have proposed a more comprehensive definition that takes the broad range of the frailty syndrome into account.

According to Ferrucci et al. (2005) and Bergman et al. (2007), at the core of frailty is the dysregulation of some fundamental biological mechanism(s) responsible for the homeostatic equilibrium of the individual, reflected in multiple physiological systems. The basic clinical features of the frailty syndrome include the following domains: (1) abnormalities in lower extremity performance and gait; (2) muscle weakness; (3) poor tolerance for exercise; (4) unstable balance; and (5) problems with body composition, such as undernutrition, sarcopenia (i.e., loss of lean body mass), and weight loss. In older persons, individual components of this syndrome are associated with certain classical geriatric conditions (e.g., falls, urinary incontinence, functional impairment, and depression). These are independent risk factors for disability, hospitalization, and death. The elements of the frailty syndrome are related to one another so that, if some are present, others tend to develop. For example, sarcopenia and poor muscle strength lead to weight loss and further aggravate sarcopenia. Frailty usually develops in older persons and this development is exacerbated by normal aging.[8]

Ferrucci et al. (2005) suggest that the theoretical definition of frailty can be operationalized in surveys and epidemiological studies as: (1) unexplained weight loss; (2) poor grip strength; (3) self-reported exhaustion; (4) slow walking speed; and (5) low physical activity. Persons with three or more of these characteristics are at significant risk of disability, hospitalization, and death. Poor lower extremity performance may be used as a single proxy biomarker for frailty in the identification of older persons who are likely to benefit from intervention for prevention of disability.

Other researchers have also sought to give the concept of frailty more precision by developing a frailty index. Rockwood and Minitski (2007) calculate a frailty score as the proportion of a list of potential deficits that an individual has accumulated. Their frailty index uses a range of deficits that are readily available in survey and clinical data. They adopted a list of 40 items from a standard Comprehensive Geriatric Assessment and evaluated each subject as 0 (=good), 0.5 (=fair), or 1.0 (=poor) on each item. The scores are given equal weight and summed, and their proportion of the maximum score of 40 is taken as the frailty index. For example,

[8] Ferrucci et al. (2005) believe that the intrinsic cause of frailty should be sought in common pathways for multiple impairments, such as hormonal changes, inflammation, disequilibrium between the production and scavenging of free radicals, and failures of the dynamic equilibrium between the complementary parts of the autonomic nervous system (i.e., the sympathetic and parasympathetic systems).

a total of 8 deficits from a total of 40 items gives a score of 0.20. The higher the index, regardless of the components of the index, the higher the likelihood that the subject will accumulate additional deficits, and will be frail and have a greater risk of adverse outcomes such as disability, institutionalization, and death. According to the frailty index, deficits tend to accumulate with rising age, women accumulate more deficits than men do, and at all ages institutional residents have higher scores than the community-dwelling population.[9]

The measurement of the frequency of frailty in the population can be effected by paper tests, performance tests, biometric tests, and self-assessments in surveys. With these data, a variety of ratios and rates, such as frailty prevalence ratios and frailty incidence rates, can be calculated. The years of frail life for individual respondents and then the average years of frail life for the pool of survey respondents can be ascertained from longitudinal surveys. "Frail span" can be derived by constructing a multiple decrement or multistate "table of healthy life" designed to apportion years of remaining life into frail years and non-frail years. These measures have a structure parallel to the morbidity/longevity measures described later in Chap. 8.

Long-term care. Long-term care is generally defined as the provision of health care, personal care, or social services over a substantial period of time to individuals who have functional limitations. It encompasses home care, community services, and institutional care. Most persons in long-term care settings in the United States reside in nursing homes. Both prevalence and incidence measures can be calculated for these dependent states. Note the very different values for the "stock" (prevalence) and "flow" (incidence) measures of residence in nursing homes at ages 65 and over in the United States: At any given time about 5% of the population 65 and over resides in nursing homes, in the course of a year some 8% of the elderly population enter nursing homes, and over the course of their lives 33–50% of the elderly will enter a nursing home. An initial analysis of the flow of population in nursing homes can be made by use of the standard population estimating equation, that is, on the basis of the resident totals at different dates and data on admissions, discharges, and deaths in the intervening period:

$$R_{y+t} = R_y + \text{Admissions - Discharges} - \text{Deaths of residents} \tag{5.13}$$

The median and mean lengths of stay can be calculated from an array of data on residents, including data on admissions, discharges, and deaths, or direct reports on length of residence. Further discussion of the risk of ever entering a nursing home is given in Chap. 8.

[9] Whitson et al. (2007), Bergman et al. (2007), and Rockwood et al. (2007) suggest other ways of composing a frailty index or defining freshly in their contributions to a special section of the *Journal of Gerontology: Medical Sciences, Vol. 62A*, No. 7, July, No. 7.

Measuring Mental Health

Two distinct areas of mental health may be measured in surveys, the affective and emotional dimension (i.e., moods) and the cognitive or intellectual dimension (i.e., thinking and memory). Both dimensions of mental health are important health attributes in their own right but they are important also because they are both associated with mortality risks. Lower scores on mental health tests are associated directly with poorer physical health and lower survival rates (e.g., Bassuk 2000; Lavery 2005).

Various batteries of questions have been employed in surveys to measure the affective aspect of mental health. A variety of terms (e.g., disorder of affect, negative mood, depression, unipolar disorder) have been employed to characterize the phenomenon being measured. Among these test scales, affective disorders have been measured in surveys by the Bradburn Affect Balance Scale. Respondents indicate whether they had experienced each of five negative moods (namely, being depressed, lonely, restless, bored, and upset) during the 2 weeks preceding the interview (U.S. NCHS 1997). Questions relating to each of these five negative moods are asked, each with a 5-level scale from 0 to 4. The negative-mood score is the sum of the scores of the five negative moods. The maximum score is 20 and the minimum is zero.

Another such test is the Center for Epidemiologic Studies Depression Scale (CES-D). This scale secures responses to 20 items representing depressive symptoms. Respondents are asked to give the frequency with which the listed depressive symptoms were experienced in the previous week. The possible responses are (1) rarely or none of the time, (2) some of the time, (3) occasionally, and (4) most or all of the time. These scores are summarized into an index and persons who scored over a specified level are coded clinically depressed and those who scored at or below this level are coded as not clinically depressed.

A different type of test is used to measure the second aspect of mental health, cognitive functioning. One such test is the Mini-Mental State Examination (MMSE). This test involves a standard battery of questions to measure memory and reasoning. More specifically, it measures orientation, attention, immediate and short-term recall, language, and the ability to follow simple verbal and written directions. Scores are assigned to distinguish severe cognitive impairment, mild cognitive impairment, and (high and low) normal cognitive functioning. The total possible score is 30 and scores under 24 determine assignment to one of the two impaired groups.

Other tests of cognitive functioning include the Clock Drawing Test (Royall et al. 1999), the Enhanced Mental Skills Test (Shankie et al. 2005), and the Delayed Word Recall Test (Knopman 1989). These tests differ in the specific domains of cognitive functioning tested, their qualities as test instruments, and their ability to predict health and survival in later life.

Measures of Overall Health Status

Subjective Health Status and Subjective Survival

As indicated in Chap. 2, it is not uncommon for health surveys to inquire about respondents' own views of their (physical) health status, called self-rated health, self-reported health, or subjective health. For this purpose a hierarchical scale such as poor, fair, good, and excellent is commonly used. A more detailed scale, including the classes very poor and very good, can be considered but it is not likely that respondents can deal with these further divisions. Studies have evaluated the efficacy of these subjective reports as proxies for objective measures of health and mortality. They can be compared with information from medical reports, health examinations, and hospital discharge records. In general, subjective health status has proven to be a rather accurate measure of "actual" health status. However, the question may provide problematic results in describing the age variation of health status because, with increasing age, respondents tend to lower their expectations on the basis of prevailing health norms or observed averages.

Most of the test studies of the use of subjective health status or self-rated health have been carried out in the more developed countries. The method has considerable potential utility in the less developed countries (LDC) also because of the great simplicity of the question, minimal cost, and the frequent lack of actual medical records in these areas. Accurate validation data are not usually available for these countries, however. There is evidence now that the question may be feasible in the LDC and that it may be a very valuable tool where objective data are not available. In a study in Bangladesh (Rahman and Barsky 2003), the use of the question proved to be both feasible and successful.

In addition to the question on self-rated health, respondents have been asked to offer their predictions of the number of years they expect to live. This measure may be called subjective survival. Here too, respondents have proved that they have a reasonable sense of their prospects for survival (Idler and Benyamini 1997; Idler and Kasl 1991). It may be useful to consider a composite measure of population health based on subjective reports of respondents relating to their general health condition and subjective reports on expected survival.

Many older persons have probably given this matter some previous thought because their evaluation of their own health status and their judgment as to its implications for their longevity may be a factor in their decision as to when to make their first claim for Social Security benefits – whether to retire on disability, to retire with reduced benefits at an early age (62 years), or to wait until they can receive full benefits (e.g., 66 years or later). Workers presumably consider their prospective financial viability and their expectations regarding their health as primary factors in the timing of their retirement, but they probably include consideration of their longevity prospects as well.

Evaluating Health Progress

It is evident from the above discussion that there are many different concepts of health and functioning, and that health conditions manifest themselves in different degrees and dimensions. The prevalence of disability among the elderly may be falling at the same time as the prevalence of chronic disease among them may be rising. Severe disability may be falling while light-to-moderate disability may be unchanged. To answer the question, is the health of the elderly population getting better or worse?, a general question on the health of the population is inadequate. Several questions reflecting the many facets of health, including the nature and severity of the conditions, are required. We may expect a variety of responses, which collectively may provide different indications of the trends in the health of the population.

The comparative evaluation of trends can be confounded by differences in designs from one health survey to another. Subjective measures of health may provide results inconsistent with objective measures. The differences in question wording (e.g., difficulty in performing an activity vs. needing help in performing the activity) may render comparisons difficult. The form of the interview (i.e., use of proxies, direct interview, self-administered questions, telephone interviews, computer-assisted interviews, etc.) may affect the nature of the responses. Different surveys may have different populations as targets (e.g., including or excluding the institutional population) or time references. Finally, the quality of the responses from the same person may differ from wave to wave of a longitudinal survey, both because of the diminished health status of the respondents and because of their increased familiarity with the questions. All of these types of problems were observed in the general review of survey findings on recent health trends in the Western countries, particularly Sweden and the United States, conducted by Parker and Thorsland (2007).

Even if these problems did not exist and questions are asked where the indicators are essentially the same for the same type of health phenomenon (e.g., using stairs vs. walking outside), we may expect not only real but also artificial differences in the findings for the sexes, age groups, socioeconomic classes, health-history groups (e.g., those with a previous history of a disease vs. those without such histories), and countries. This is so both because of variations in the ways that these groups tend to respond to health questions and also because of real differences in the health statuses of the categories. It is apparent that even with the best of data and with careful efforts at intercountry harmonization of data,[10] health data and trends are variegated and complex.

Since the different health and functioning measures represent different dimensions of health status and health outcomes, it is wise to apply multiple measures of

[10]Harmonization in this context refers to the adjustment of the data, definitions, and measures so as to achieve comparability and consistency between countries.

health status and functioning in research on these topics. The variety of health trends contributing to the overall trend, each with different costs and requirements for the resources needed, complicates the management of the various health conditions and calls for a variety of administrative responses. It also complicates the evaluation of future costs of health care.

Physiological Measures

The next chapter will review the risk factors – biological, individual, social, environmental, and stochastic – associated with the principal health conditions. Researchers have drawn on this information to develop an array of clinical measures, for use in epidemiological and social research, to measure the biological rate of human aging, the physical and functional decline of respondents in longitudinal surveys, and the effectiveness of various interventions. Two parallel streams of thinking regarding the determination of these clinical measures and their outcomes have developed. One stream has been engaged in identifying so-called biomarkers, anthropometric and biological measures of aging that are expected to be more predictive of life span than chronological age. The second stream has been engaged in identifying a group of measures, termed allostatic load, that combines a variety of risk factors affecting multiple physiological systems and that can be expected to predict future health risks better than any single risk factor by itself. Biogerontologists have followed the first path and epidemiologists the second although the two efforts have the same general goal and complement one another. If I merge these research approaches, I arrive at the definition of allostatic load as the cumulative dysregulation of bodily functions resulting from the combined negative impact of several biomarkers on health conditions.

Biomarkers

Biomarkers may be viewed as physical and mental traits that anticipate the onset of a disease, measure the progress toward a disease, are diseases that cause more serious diseases, or show the effect of some intervention to slow the progress of a disease. Such biological parameters are sought with the goal of measuring quantitatively the rate of aging more accurately than chronological age (Ingram 2004). The levels of the biomarkers tend to increase with advancing age, the most common biomarkers rising very roughly linearly with age. Various biological risk factors, symptoms, and precursors of disease have been used as biomarkers of aging. The statistical goal of determining the level of biomarkers is to calculate biological age, defined for this purpose as the residual between actual and predicted age. The predicted age is derived from a statistically modeled combination of many biomarkers. (See Karasik et al. 2005.)

Baker and Sprott (1988) were the first to offer a formal definition of a biomarker: "A biological parameter of an organism that, either alone or in some multivariate

composite, will, in the absence of disease, better predict functional capacity at some later age than will chronological age." Among the criteria for a biomarker of aging the following have been mentioned (Ingram et al. 2001; Ingram 2004; Baker and Sprott 1988):

1. Critical to effective maintenance of health and prevention of disease,
2. A measurable parameter that can be predicted at a later age or that can predict functional capacity at a later age, life span, or the age of onset of specific age-related diseases,
3. Significant cross-sectional correlation with age,
4. Significant longitudinal change in the same direction as the cross-sectional correlation,
5. Predictive of life span,
6. Consistent with a rate of age-related change and proportional to differences in life span among related species, and
7. Significant stability of individual differences over time.

To be useful for epidemiological studies, a biomarker should have the following characteristics: (1) it should be able to predict, better than chronological age, declines in one or more organ systems, the likelihood of the onset of chronic disease, or the age-related loss of function; (2) testing for it in surveys or field examinations should be minimally invasive and easily effected; (3) it should be able to measure a trait reliably, preferably on a quantitative scale; and (4) it can be readily collected from large numbers of persons with minimal expenditure of resources.

The criteria for defining and validating biomarkers of aging have been subject to considerable debate in the gerontological community. Most of the biomarkers selected so far relate to the cardiovascular system, but the physiological systems work synergistically so that some biomarkers measure changes in more than one system in the body. One particular group of biomarkers has been identified in a cluster as the metabolic syndrome. The metabolic syndrome is characterized by high levels of blood glucose and triglycerides, high insulin resistance, low levels of high density lipoproteins, high blood pressure, an increased tendency to form blood clots, and being overweight. These risk factors collaboratively increase the risk of developing cardiovascular disease and diabetes. The National Cholesterol Education Program (ATP Clinical Panel III) has set up the following criteria for identifying metabolic syndrome:

Abdominal obesity (waist circumference)	
Men	>102 cm (>40 in.)
Women	>88 cm (>35 in.)
Triglycerides	>150 mg/dL
HDL cholesterol	
Men	<Under 40 mg/dL
Women	<50 mg/dL
Blood pressure	>135/>85 mg/dL
Fasting glucose	>100 mg/dL

Source: Grundy et al. (2004)

A diagnosis of metabolic syndrome is made when three or more of the risk factors in the above list are present.

The Framingham Risk Score in combination with the coronary-artery-calcium score is another way of identifying patients at risk. The first takes into account such variables as age, sex, smoking history, cholesterol level, blood pressure, and diabetes, and the second is determined by a CAT (i.e., computerized axial tomography) scan of the coronary arteries that can show calcium deposits and the formation of plaques.

A comprehensive list of biomarkers combines standard clinical measures with non-clinical measures. Exhibit 5.1 enumerates some leading biomarkers classified according to physiological system. The standard clinical measures include body mass index, systolic blood pressure, diastolic blood pressure, total cholesterol, ratio of total cholesterol to high density cholesterol (HDL), and glycosylated hemoglobin. Non-clinical biomarkers cover neuroendocrine and immune system responses, and include cortisol, DHEA-S, norepinephrine, epinephrine, IL-6, IGF-1, and dopamine. The neuroendocrine and immune system biomarkers are important because they appear to provide warning signs of deteriorating health and function beyond what can be learned from the biomarkers relating to cardiovascular disease and metabolic function (Goldman et al. 2006).

Many other global and system-specific biomarkers have been proposed and tested for their ability to predict health outcomes and measure biological aging. Among them are presbyopia (e.g., inability to focus on nearby objects), skin elasticity, changes in telomere length (i.e., the length of the caps on the ends of chromosomes, which are the ordered collections of genes in cells), hearing loss, number of healthy teeth, sarcopenia (i.e., loss of lean muscle mass), clotting factors, grip strength, forced expiratory volume, bone mineral density, and cognitive and neuropsychological findings. From 29 physiological variables Nakamura and Miyao (2007) selected five variables (i.e., forced expiratory volume in 1 second, systolic blood pressure, hematocrit, albumin, and blood urea nitrogen) as their candidate biomarkers of aging.

Allostatic Load

None of the above biomarkers taken individually shows a strong ability to predict major health outcomes such as mortality and declines in physical and cognitive functioning. A summary index of several biomarkers, designated allostatic load, does have the ability to predict these types of health outcomes. Such a summary index appears to have the ability to measure potential dysfunctionality in a range of organ systems prospectively, even over an adult lifetime.

Allostasis has been defined by Sterling and Eyer (1988) as the body's ability to adapt its internal physiologic milieu to match external demands. As the body is challenged by external pressures (e.g., attack and accompanying fear), the body responds dynamically to regulate internal physiological parameters (e.g., heightened blood pressure). At younger ages allostasis is a normally adaptive response to stress,

Cardiovascular:
Resting systolic blood pressure, resting diastolic blood pressure, sum or mean of systolic and diastolic blood pressures, electrocardiogram (ECG)
Total cholesterol, high density lipoprotein (HDL), ratio of total cholesterol to HDL, triglyceride level, lipoprotein (a) Neck murmurs, tachycardia/bradycardia, heart rate, heart rate variability/arrythmia, arterial stiffness, ankle/brachial pulse ratio
Hematocrit, white blood cell count, red blood cell count, hemoglobin
Inflammatory:
C-reactive protein, interleukin 6, fibrinogen, albumin, homocysteine
Musculoskeletal:
Bone density, muscle mass, lean body mass, waist/hip ratio, motor functioning, grip strength, height/waist ratio
Respiratory:
Maximum vital breathing capacity, forced expiratory volume in 1 s (FEV_1)
Endocrinological:
UR cortisol, UR epinephrine, UR norepinephrine
Urological:
Creatinine clearance, proteinuria, protein-specific antigen (PSA), blood urea nitrogen
Metabolic:
Glycosylated hemoglobin, fasting insulin, fasting glucose, insulin sensitivity
Neurological:
Ability to touch left earlobe with right hand, balance, patellar reflex
Cognitive function:
Memory, psychological analysis

Exhibit 5.2 Selected biomarkers used in measuring biological aging and predicting adverse health outcomes, according to physiological system
Biomarkers listed are not mutually exclusive and are not necessarily limited to one physiological system

from which the body tends to return quickly to homeostasis. At the older ages, however, the levels of the biomarkers reflect the dysregulation in the body resulting from the repeated challenges to it. The cumulative dysregulation across multiple physiological systems resulting from the strain on the body produced by repeated challenges to the body's physiological systems as a person gets older constitutes its cumulative allostatic load (McEwen and Stellar 1993). Alternatively, cumulative allostatic load may be defined simply as the body's declining capacity to respond to stress, or the "wear and tear" on the body of older adults.

Exhibit 5.2 enumerates the biomarkers or components of allostatic load that are now commonly used in epidemiological research. Among the biomarkers are blood pressure, glucose and lipid metabolism, certain anthropometric indicators, and hormonal excretions. Specific biomarkers that have been individually linked to increased risks for disease are urinary cortisol, norepinephrine, and epinephrine excretions, serum DHEA-S, systolic and diastolic blood pressures, ratio of waist-to-hip circumference, high density lipid (HDL) cholesterol and the ratio of total to HDL cholesterol, and blood glycosylated hemoglobin (Karlamangla et al. 2002). These biomarkers are well-established risk factors for

obesity, hypertension, atherosclerosis, lipid imbalance, loss of bone mineral density, sarcopenia, immune dysfunction, and ultimately diabetes, cardiovascular diseases, and dysregulation in several physiological systems.

The allostatic load for an individual can be determined by assigning scores to each risk and combining the scores for the various risks (Seeman et al. 2001; Karlamangla et al. 2002). The scores can be weighted equally or unequally when combining risks. In this way individuals can be assigned to allostatic-load categories from lowest to highest, and their scores can be analyzed separately for each individual or for the population sample. In addition, the contribution to total mortality risk can be determined separately for each of the biomarkers and for component classes of biomarkers, e.g., cardiovascular, metabolic, inflammatory, neuroendocrine, and so on. It is evident that the physiological systems "talk to one another" because of the frequent appearance of risk scores indicating joint dysregulation for various physiological processes. Certain combinations are frequent and others are not. As will be explained further in Chap. 6, genetic background and life experiences affect how the systems act and interact.

Age variation in allostatic load. The percent of the population with high allostatic loads tends to increase with age. This is true also for the component biomarkers, such as systolic blood pressure, HDL cholesterol, glycosylated hemoglobin, and so on. The practical significance of the level of the risk factors varies with age and sex also. Each age-sex category – from childhood to very advanced ages for each sex – is physiologically different and the salutary levels of the risk factors may differ for the various age-sex groups. There is ambivalence in the interpretations of some risks, however. For example, the significance of high levels of various risk factors on the longevity of the population of advanced age has been questioned. Such questions have been raised with respect to lipid cholesterol, body mass index, and hypertension. Some researchers find that, at the advanced ages, reducing cholesterol levels, lowering weight, and controlling hypertension are associated with increased mortality. They maintain, for example, on the basis of various epidemiological survey data, that persons over age 80 who are free of cardiovascular disease do not have to concern themselves with reducing their cholesterol levels by taking medication or observing dietary guidelines. Other research suggests that, above the age of 80, overweight is not a risk factor for premature death. There are those analysts also who maintain that administering anti-hypertensive medication for high blood pressure may increase the risk of mortality after age 85 (Goodwin 2003). On the other hand, there is perhaps an equally large body of work that demonstrates increased survival of persons over age 80 when blood pressure has been lowered with medication, or shows an insignificant rise. (See further discussion in Chap. 6.)

Relative Index of Overall Health

Many proposals have been made for constructing a relative index of overall health for each person as a way of determining an individual's biological age. In order to

derive such an index, one could calculate the (weighted or unweighted) average of a selected combination of physiological biomarkers and measures of performance at each age, and then derive an index of health for the person by comparing the summary measure for the individual with the corresponding summary measure for persons in some band of young ages, such as 25–34, taken as standard. It would be interesting also to compare the summary measure for the individual with the summary measure at the same age for the larger population, representing "normal" (i.e., typical) performance at the age. Such a comparison would tell us little about the health condition of the individual, however, since the "average" or "normal" at the older ages would be an artificially low standard for evaluating health at these ages.

Biomarkers generally show gradual age-related declines from early adulthood to the highest ages. Because of the great variability in the indications of biomarkers and in the performance of individuals at given ages, the approach described here has been considered futile as a standard for predicting the health status and longevity of individuals. Calculation of measures of biological aging at each age from combinations of biomarkers should be useful, however, in showing the average shift in the health condition of a population over the age scale.

Measures of Use and Availability of Health Services and Support Systems

In this brief section, only some of the many measures of the utilization of health resources are listed. Such measures include those relating to the availability of health services, the use of health services by patients, and use and availability of formal support services and of kinship networks that provide informal support. The information provided by most of these measures is enhanced by age and sex detail as well as by detail on other demographic and socioeconomic characteristics. Some aspects of the use of health services, such as access to care and quality of care, are discussed in Chaps. 6, 7, and 15.

Selected Measures of Use of Health Providers

The first group of measures is concerned with visits to health providers, such as physicians and dentists, visits to health facilities such as hospital out-patient units and emergency departments, and home health visits. Illustrative measures are:

Physician visits in a year per 100 midyear population
Dental visits in a year per 100 midyear population

Percent distribution of the population by number of physician visits in a year (1, 2, 3, 4, 5+)
Percent distribution of the population by number of physician hours (under 1, 1–5, 5–10, 10+), for age groups
Hospital outpatient visits in a year per 100 midyear population
Emergency department visits in a year per 100 midyear population
Percent of adults (children) with hospital out-patient visits in the past 12 months by number of visits
Percent of adults (children) with emergency department visits in the past 12 months by number of visits
Percent distribution of the population by total health-care visits during a year by number of visits
Percent of adults (children) without a usual source of health care
Percent of adults (children) without a health care visit to an office or clinic in the past 12 months

Another group of measures is concerned with the use of hospitals. Some illustrative measures are:

Hospital admissions in a year per 1,000 midyear population
Percent distribution of the population by number of hospital stays during the year
Discharges during the year per 1,000 midyear population
Days of hospital care during the year per 1,000 midyear population
Average length of hospital stay in days for the year
Discharges during the year per 10,000 midyear population by procedure (i.e., treatment)

Formal and Informal Support

Some measures of informal support are population measures representing the ratio of groups in the population who are "potentially available" to support "potentially dependent" groups. Other measures of informal support, and measures of formal support, represent the actual amounts of care delivered. Some illustrative measures are:

Percent of persons aged 65 years and over who have living children
Percent of persons aged 50 years and over who have living parents 65 years and over
Days of care delivered per week by caregivers per 100 dependent persons
Average number of days of care delivered per week by each caregiver
Percent distribution of nursing-home residents by functional status
Nursing-home residents 65 years and over (85 years and over) per 1,000 midyear population 65 years and over (85 years and over)
Nursing home occupancy rate per 100 nursing home beds

Appendix 5.1 Method of Calculating the Historical Trend in the Median Age of Disablement

To obtain the trend of the mean or median ages of onset of disability over several 5-year periods from data on numbers of persons disabled by age, the basic formula is

$$D_i = (D_{2p} - D_{1p}{}^{*}s) \div \sqrt{s}. \tag{5.14}$$

The steps here are given for 5-year age groups and time intervals, but 10-year groups and 10-year intervals could be employed as well. This formula assumes that disabled persons and nondisabled persons have the same survival rates within the 5-year age period. The steps are:

1. Secure from surveys or censuses the numbers of disabled persons in the conventional 5-year age groups 55-years and over at 5-year time intervals, so that the 5-year data can be linked to one another as cohort data (D_{2_p}, D_{1_p}). If necessary, interpolate the data into a cohort-table format.
2. Apply 5-year survival rates (s) to the initial population (D_{1_p}) to determine the preliminary number of disabled survivors 5-years older 5-years later (D_{1_pS}). The survival rates for the general population may not be appropriate because the disabled population is at higher risk; lower survival rates could be estimated and applied.
3. Subtract the survivors in (2) from the corresponding number of disabled persons in the terminal age group (D_{2_p}), to derive the preliminary numbers of net disablements for the 5-year cohort for the 5-year time period. The assumption is that there are few reverse disablements and that these reversals are temporary.
4. Divide the result in (3) by the square root of the 5-year survival rate ($\sqrt{s}$). The results are the estimates of net disablements for the 5-year cohorts for the 5-year time periods.
5. Apply third-difference Karup-King osculatory interpolation to the results in (4) to redistribute the cohort data on number of disablements into the conventional 5-year age groups.
6. Calculate the median age of disablements from the array by 5-year age groups for each 5-year time period from the results in (5).
7. Examine the time trend of median ages for each 5-year time period for reasonableness and continuity.

An alternative formula can be used when the original data are given in the form of age-specific disability ratios:

$$D_i = \sqrt{s}[1 - (d_{r2} \div d_{r1})]D_{1p} \tag{5.15}$$

where d_r represents the disability ratios for the cohort at the initial (d_{r1}) or terminal (d_{r2}) dates of the 5-year periods, D_{1_p} represents the number of disabled persons at the start of each 5-year period, and *s* represents the 5-year survival rates.

References and Suggested Readings

Sources and Classification Systems

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*, DSM-IV. Washington, DC: American Psychiatric Association.

Ferlay, J., Bray, F., Pisani, P., & Parkin, D. M. (2001). *GLOBOCAN 2000*: *Cancer incidence, mortality and prevalence worldwide*, Version 1.0 IARC (International Association of Cancer Registries) CancerBase No. 5. Lyons, France: IARC Press, Limited version available online at: http://www-dep.iarc.fr/globocan/globocan.htm.

Mausner, J. S., & Kramer, S. (1985). *Epidemiology: An introductory text*. Philadelphia: W. B. Saunders.

World Health Organization (WHO). (1980). *International classification of impairments, disabilities, and handicaps* (ICIDH). Geneva, Switzerland: World Health Organization.

World Health Organization. (1984). *International classification of impairments, disabilities, and handicaps: A manual of classification relating to the consequences of disease*. Geneva, Switzerland: World Health Organization.

World Health Organization. (1992, 1993, 1994). *International statistical classification of diseases and related problems*. Vol. I: ICD-10, Tabular List (1992); Vol. II: Instructional manual (1993); Vol. III: Alphabetical Index (1994). Geneva, Switzerland: World Health Organization.

World Health Organization. (2001). *International classification of functioning, disability, and health* (ICF). Designated ICIDH-2 and available via the World Wide Web.

Self-Reported Health

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of 27 community studies. *Journal of Health and Social Behavior, 38,* 21–37.

Idler, E., & Kasl, S. (1991). Health perceptions and survival: Do global evaluations of health status really predict mortality? *Journal of Gerontology: Social Sciences, 46B,* S55–S65.

Rahman, M. O., & Barsky, A. J. (2003). Self-reported health among older Bangladeshis: How good a health indicator is it? *Gerontologist, 43(6),* 856–863.

Wolinsky, F. & Johnson, R. (1992). Perceived health status and mortality among older men and women. *Journal of Gerontology: Social Sciences, 47(6),* S304–S312.

Disease, Injury, Disability, and Frailty

Amella, E. J. (2003). Presentation of illness in older adults. *American Journal of Nursing, 104(10),* 40–51.

Ansello, E. F., & Eustis, N. N. (Eds.). (1992). *Aging and disabilities: Seeking common ground*. Amityville, NY: Baywood Publishing Company.

Bassuk, S. S. (2000). Cognitive impairment and mortality in the community-dwelling elderly. *American Journal of Epidemiology, 151,* 676–688.

Bergman, H., Ferrucci, L., Guralnik, J., et al. (2007). Frailty, an emerging research and clinical paradigm: Issues and controversies. *Journal of Gerontology: Medical Sciences, 62A(7),* 731–737.

Bortz, W. M. (2002). A conceptual framework of frailty: A review. *Journals of Gerontology: Medical Sciences, 57A,* M283–M288.

Branch, L. G., Katz, S., Kniepmann, K., & Papsidero, J.A. (1984) A prospective study of functional status among community elders. American *Journal of Public Health, 74(3)*, 266–268.

Constantino, J. P., Gail, M. H., Pee, D., et al. (1999). Validation studies for models projecting the risk of invasive and total breast cancer incidence. *Journal of the National Cancer Institute, 91(18),* 1541–1548.

Dye, C., Scheele, S., Dolin, P., Pathania, V., & Raviglione, M. C. (1999). Global burden of tuberculosis: Estimated incidence, prevalence, and mortality by country. *Journal of the American Medical Association, 282,* 677–686.

Ferrucci, L., Windham, B. G., & Fried, L. P. (2005). Frailty in older persons. *Genus, LXI*(1), 39–53.

Freedman, V. A., Martin, L. G., & Schoeni, R. F. (2002). Recent trends in disability and functioning among older adults in the United States: a systematic review. *Journal of the American Medical Association, 288(24),* 3137–3146.

Fried, L. P., Tangen, C. M., Walston, J., et al. (2001). Frailty in older adults: Evidence for a phenotype. *Journal of Gerontology: Medical Sciences, 56A(3)*, M146–M156.

Fries, J. F. (2003). Measuring and monitoring success in compressing morbidity. *Annals of Internal Medicine, 139*(5, Pt. 2), 455–459.

Gail, M. H., Brinton, L. A., Byar, D. P., et al. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute, 81(24),* 1879–1886.

Guralnik, J. M., Simonsick, E. M., et al. (1994). A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology: Medical Sciences, 49(2),* M85–M94.

Haddon, W. (1980). Advances in the epidemiology of injuries as a basis for *public policy. Public Health Reports, 95(5),* 411–421.

Humber, J. A., & Almeder, F. E. (Eds.). (1997). *What is disease?* Totowa, NJ: Humana Press.

Katz, S., Branch, L. G., Branson, M. H., Papsidero, J. A., et al. (1983). Active life expectancy. *New England Journal of Medicine, 309,* 1218–1221.

Knopman, D. S. (1989) A verbal memory test with high predictive accuracy. *Archives of Neurology, 46,* 141–145.

Lavery, L. L. (2005). The clock drawing test is an independent predictor of incident use of 24-hour care in a retirement community. *Journal of Gerontology Series A: Biological and Medical Sciences, 7,* 928–932.

Manton, K., Gu, X., & Lamb, V. (2006). Change in chronic disability from 1982 to 2004/05 as measured by long-term changes in function and health in the U.S. elderly population. *Proceedings of the National Academy of Sciences, 103(48),* 18374–18379.

Nagi, S. (1976). An epidemiology of disability among adults in the United States. *Milbank Memorial Fund Quarterly/Health and Society, 54,* 439–467.

Parker, M. G., & Thorsland, M. (2007). Health trends in the elderly population: Getting better and getting worse. *Gerontologist, 47(2),* 150–158.

Ravaud, J. -F., Letourney, A., & Ville, I. (2002). Identifying the population with disability: The approach of an INSEE survey on daily life. *Population-E, 57(3),* 29–554.

Robertson, L.S. (1998). *Injury epidemiology.* (2nd ed.). New York: Oxford University Press.

Robine, J. -M., Jagger, C., & Romieu, I. (2001). Disability-free life expectancies in the European Union countries: calculation and comparisons. *Genus, LVII(2),* 89–102.

Rockwood, K., Andrew, M., & Minitski, A. (2007). Frailty. A comparison of two approaches to measuring frailty in older people. *Journal of Gerontology: Medical Sciences, 62A(7)*, 738–743.

Rockwood, K., & Minitski, A. (2007). Frailty in relation to the accompensation of deficts. *Journal of Gerontology: Medical Sciences, 62A(7)*, 722–727.

Rosow, I., Breslau, N., & Guttman, A. (1966). A health scale for the aged. *Journal of Gerontology, 21,* 556–559.

Royall, D. R., Mulroy, A.R., Choido, L.K., & Folk, M.J. (1999). Clock drawing is sensitive to executive control: A comparison of six methods. *Journal of Gerontology, 54,* 328–333.
Ruzicka, L. T., Choi, C. Y., & Sadkowsky, K. (2004). Co-morbidity of suicides in the light of multiple cause of death reporting. *Genus, LX(3–4),* 143–160.
Shankie, W. R., Romney, A. K., Hara, J., Fortier, D., et al. (2005). Methods to improve the detection of mild cognitive improvement. *Proceedings of the National Academy of Sciences, 102(13)*, 1919–1924.
Stallard, E. (2002, January 17–18). *Underlying and multiple cause mortality at advanced ages: United States,* 1980–1998. In CD of Proceedings of the Society of Actuaries, International Symposium: Living to 100 and Beyond: Mortality at Advanced Ages, Lake Buena Vista, FL.
UNICEF (2001). *The state of the will's children 2001.* New York, United Nations.
U.S. National Center for Health Statistics. (1997). Negative mood and urban vs. rural residence: Using proximity to Metropolitan Statistical Areas as an alternative measure of residence. By B. S. Jones & R. W. Wilson. *Advance Data No.* 281, *Vital and Health Statistics.*
U.S. National Center for Health Statistics. (2003, April). *Proceedings of the international collaborative effort on injury statistics, Vol. IV,* Paris Meeting. Hyattsville, MD: U.S. National Center for Health Statistics. See pp. 2–1 to 2–8.
Verbrugge, L. & Jette, A. (1994). The disablement process. *Social Science and Medicine, 38(1),* 1–14.
Walston, J., Hadley, E. C., Ferrucci, L., et al. (2006). Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Frailty in Older Adults. *Journal of the American Geriatrics Society, 54(6),* 991–1001.
Whitson, H. E., Purser, J. L., & Cohen, H. J. (2007). Frailty thy name is . . . phrailty. *Journal of Gerontology: Medical Sciences, 62A(7)*, 728–730.
Wolf, D. A., Mendes de Leon, C. F., & Glass, T. A. (2007). Trends in rates of onset of and recovery from disability at older ages: 1982–1994. *Journal of Gerontology: Social Sciences, 62B(1)*, S3–S10.

Age-Period-Cohort Analysis

Brady, H. E., & Elms, L. (1999). *Age-period-cohort analysis with noisy, lumpy data*. Paper presented at the annual meeting of the Political Methodology Group of the American Political Science Association, College Station, TX.
Carstensen, B. (2005). Demography and epidemiology: Age-period-cohort models in the computer age. www.pubhealth.ku.dk/~bxc
Cook, N. R., & Ware, J. H. (1983). Design and analysis methods for longitudinal research. *Annual Review of Public Health, 4,* 1–23.
Dorn, H. F., & Cutler, S. J. (1959). *Morbidity from cancer in the United States* (Public Health Monograph No. 56). Washington, DC: U.S, Government Printing Office.
Finch, C. E., & Crimmins, E. M. (2004). Inflammatory exposure and historical changes in human life-spans. *Science, 305,* 1736–1739.
Firebaugh, G. (1997). *Analyzing repeated surveys*. Sage University Paper Series No. 07-115. Thousand Oaks, CA
Frost, W. H. (1939). The age selection of mortality from tuberculosis in successive decades. *American Journal of Hygiene, 30(3A),* 91–96.
Glenn, N. D. (1976). Cohort Analysis' futile quest: Statistical attempts to separate age-period-cohort effects? *American Sociological Review, 41*(October), 900–904.
Glenn, N. D. (2005). *Cohort analysis. Quantitative applications in the social sciences* (2nd ed.). Thousand Oaks, CA: Sage.

Halli, S. S., & Rao, K. V. (1992). Age, period, and cohort effects in demography. In *Advanced Techniques of Population Analysis* (Chap. 3). New York: Plenum Press.

Hobcraft, J., Menken, J., & Preston, S. (1982). Age, period, and cohort effects in demography: A review. *Population Index, 48*(Spring), 4–43.

Jones, H. (1956). *Advances in biological and medical physics* (Vol. 4, p. 281). New York: Academic.

Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934a). Death rates in Great Britain and Sweden: Some general regularities and their significance. *Lancet, 1,* 698–703.

Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934b). Death rates in Great Britain and Sweden: Expression of specific mortality rates as products of two factors, and some consequences thereof. *Journal of Hygiene, 34,* 433–457.

Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). Age, period, and cohort effects. In *Epidemiologic research: Principles and quantitative methods* (Sect. 7.3). Belmont, CA: Lifetime Learning Publishers.

Mason, W. M., & Fienberg, S. E. (Eds.). (1985). *Cohort analysis in social research: Beyond the identification problem*. New York: Springer. See esp. papers by W.M. Mason and S. E. Fienberg, Introduction, beyond the identification problem; S.E. Fienberg and W.M. Mason, Specification and implementation of age, period, and cohort models; and W. M. Mason and H.L. Smith, Age-period-cohort analysis and the study of deaths from pulmonary tuberculosis.

Palmore, E. (1978). When can age, period, and cohort be separated? *Social Forces, 57*(September), 282–295.

Siegel, J.S. (1993). *A generation of change: A profile of America's older population*. New York: Russell Sage Foundation.

Yang, Y., Land, K. C., & Fu, W. J. (2006, August 11). *The intrinsic estimator for age-period-cohort analysis and how to use it*. Paper presented at the annual meeting of the American Sociological Association, Montreal, Canada.

Biomarkers and Allostatic Load

Baker III, G. T., & Sprott, R. L. (1988). Biomarkers of aging. *Experimental Gerontology, 23,* 223–239.

Goldman, N., Turra, C. M., Glei, D. A., Seplaki, C. M., et al. (2006). Predicting mortality from clinical and nonclinical biomarkers. *Journal of Gerontology: Medical Sciences, 61A(10),* 1070–1074.

Goodwin, J. S. (2003). Embracing complexity: A consideration of hypertension in the very old. *Journal of Gerontology: Medical Sciences, 58A(7),* 653–658. See also commentaries on this paper by W. S. Aronow, S. Denson, R. R. Hajjar, T. B. Harris, D. T. Lowenthal, J. -P. Michel et al., A. B. Newman, & D. T. Thomas.

Grundy, S. M., Brewer, H. B., Cleeman, J. I., Smith, S. C., & Lenfant, C. (2004). Definition of metabolic syndrome. Report of the National, Heart, Lung, and Blood Institute/American Heart Association Conference on scientific issues related to definition. *Circulation, 109(2),* 433–438.

Ingram, D. K. (2004). Biomarkers of aging: From marking time to moving forward. Workshop in Biomarkers of Aging. *Alliance for Aging Research, October 20,* 2004. Baltimore, MD: Gerontology Research Center, National Institute on Aging, National Institutes of Health.

Ingram, D. K., Nakamura, E., Smucny, D., et al. (2001). Strategies for identifying biomarkers of aging in long-lived species. *Experimental Gerontology, 36,* 1025–1035.

Karasik, D., Demissie, S., Cupples, L. A., & Kiel, D. P. (2005). Disentangling the genetic determinants of human aging: Biological age as an alternative to the use of survival measures. *Journal of Gerontology: Biological Sciences, 60A(5),* 574–587.

Karlamangla, A. S., Singer, B. H., McEwen, B. S., et al. (2002). Allostatic load as a predictor of functional decline: MacArthur studies of successful aging. *Journal of Clinical Epidemiology, 55,* 696–710.

McEwen, B. S., & Stellar, E. (1993). Stress and the individual mechanisms leading to disease. *Archives of Internal Medicine, 153,* 2093–2101.

Nakamura, E., & Miyao, K. (2007). A method for identifying biomarkers of aging and constructing an index of biological age in humans. *Journal of Gerontology: Biological Sciences, 62A(10),* 1096–1105.

Seeman, T. E., Singer, B., Rowe, J. W., et al. (1997). Price of adaptation – Allostatic load and its health consequences: McArthur studies of successful aging. *Archives of Internal Medicine, 157,* 2259–2268.

Seeman, T. E., McEwen, B. S., Rowe, J. W., & Singer, B. H. (2001). Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. *Proceedings of the National Academy of Sciences, 98(8),* 4770–4775.

Seplaki, C. L., Goldman, N., Weinstein, M., & Lin, Y. H. (2004). How are biomarkers related to physical and mental well-being? *Journal of Gerontology: Medical Sciences, 59A,* 201–207.

Sterling P., & Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. In S. Fisher & J. Reason (Eds.), *Handbook of life stress, cognition, and health* (pp. 631–651). New York: Wiley.

Part III
Mortality and Morbidity Considered Jointly and in Relation to Fertility, Migration, and Population Age-Sex Structure

Part III takes us from the discussion of concepts and basic tools, covered in Parts I and II, to the more substantive material of health demography and epidemiology. This part focuses on the trends of mortality and morbidity, the factors affecting them, and the relations of mortality and morbidity to the other demographic factors of change, namely fertility and migration, and to a population's age-sex structure. Chapter 6 considers the trends in mortality and morbidity, mainly in the United States and the other more developed countries; their variations by age and cause; and the factors associated with the trends and variations. The following chapter, Chap. 7, considers the variations in mortality and morbidity in relation to a wide range of demographic and socioeconomic characteristics other than age, such as sex, race, ethnicity, income, education, religion, and geographic region, again mainly for the United States and other more developed countries. In Chap. 8 I return to methodological issues for a discussion of methods of analyzing mortality and morbidity jointly and discuss the substantive findings of studies that have linked these two events.

The interrelations of health and fertility, or reproductive health, are covered in Chap. 9, and the interrelations of health and migration are discussed in Chap. 10. The special health issues in the less developed countries, and their determinants and demographic consequences, are the subject of Chap. 11. Part III of the book closes with a discussion in Chap. 12 of the interrelations of health, mortality, and age-sex population structure.

In any discussion of population from an international point of view, it is convenient to group the countries that are similar with respect to economic development, political structure, geographic location, history, and culture. The leading classification has been promulgated by the United Nations (UN). It groups the countries of the world essentially on economic grounds into the More Developed Countries (MDC) and the Less Developed Countries (LDC). The Less Developed Countries have been further subdivided into the Least Developed Countries and the other Less Developed Countries. With the dissolution of the Soviet Union in 1989, the UN assigned Russia, the former republics of the Soviet Union in Europe and

Asia, and the Soviet bloc countries in Europe to the MDC or the LDC depending on their location in Europe or Asia:

More developed countries (including Russia and the former Socialist economies in Europe)
Less developed countries (including the former Soviet republics in Asia)

Least developed countries
Other less developed countries

In the report on the *Global Burden of Disease*, Murray and Lopez (1996) employed a different type of classification. They defined eight demographic regions in the world, and placed two of these, namely the Established Market Economies (largely the OECD countries) and the Formerly Socialist Economies of Europe, in the more developed world, and the remaining six in the less developed world:

More developed countries

Established market economies
Formerly Socialist economies of Europe

Less developed countries (separated in six regions)

India
China
Other Asia and Islands (including the former Soviet republics)
Latin America and the Caribbean
Middle Eastern Crescent
Sub-Saharan Africa

The former Socialist economies of Europe and the former soviet republics of Asia have been grouped together as the Economies in Transition in some studies.

The demographic regions of the world, divided into the More Developed and the Less Developed regions, are listed in Exhibit A4.1 in Appendix 4. A list of the Economies in Transition is shown in Exhibit A4.2 in Appendix 4. A list of Least Developed Countries is shown in Exhibit A4.3 in Appendix 4. Exhibit A4.4 in Appendix 4 lists the Low- and Middle-Income regions and the High-Income regions, according to the World Health Organization as of 2004.

Chapter 6
Health Inequalities, General Trends in Mortality and Morbidity, and Associated Factors

Measurement of Inequality of Mortality and Health

Introductory Methodological Notes

All measures of health status are ultimately derived from observations of individuals. At the field level we have such measures as self-assessed health status, report of a specific disease, record of a particular death, or an individual's test on a biomarker, such as blood pressure or serum cholesterol. The observations for individuals are combined and summarized to represent subnational geographic areas, demographic or socioeconomic groups within countries, or national populations. The summary measures, whether they are percentages, averages, or rates, apply to groups. A problem arises when the measures that are based on groups are assumed to represent individuals. The analysis becomes especially problematic when the units analyzed are geographic areas and inferences are being made about individuals from the analysis for these geographic areas.

Individuals differ from one another in their health status, and they tend to be disproportionately clustered in one health status or another. A population described by health-status variables then tends to be characterized by varying degrees of health inequality. There is health inequality among population groups within countries and subnational geographic and residence areas, and among countries and world regions. The basic methods of measuring inequality of the health status of groups within countries are generally the same as the methods used to measure health inequality among whole countries and world regions.

Some simpler measures of health inequality describe single populations independently, which may then be compared with the results of the same group measure for other populations; or are designed to link two populations directly in the computation of the measure. Another group of more complex measures of health inequality links health status or health events with the demographic and socioeconomic characteristics of persons such as age, sex, race, ethnicity, marital status, educational level, and income class, and seeks to describe the extent of the association and even

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_6, © Springer Science+Business Media B.V. 2012

the causal relation between health status and these characteristics. Most measures of health inequality were developed in order to analyze differences in health status for such socioeconomic subgroups in national populations. The first group of measures, extended to include age, is described in the present chapter, and the second group of measures is described in Chap. 7. It is important to distinguish between measures of inequality in the distribution of a health measure, such as among ages, geographic areas, and general population groups, and measures of the differences in a health measure among various socioeconomic groups in a population for the purpose of determining the degree of causation between them. The latter may involve not only different measures but also causal inferences, value judgments, and differences in interpretation. While most recent research has focused on variations in the health status of socioeconomic groups within countries, in the discussion that follows in this chapter, I focus mainly on variations in health status among countries and among general populations. As is usual in demographic or epidemiological analysis, the conclusions drawn about variations and trends in health inequality among populations may differ depending on the measure used.

Measures of Inequality Within Countries

Measures That Describe Individual Populations

The reader is familiar with the summary measures of central tendency and dispersion that are used for analyzing the mortality and health characteristics of countries and comparing a group of countries with regard to these characteristics. They include the mean (x_m), median (x_{md}), and mode (x_{mo}), the standard deviation (σ) and interquartile range (IQR), and the standardized versions of the latter two measures, namely the coefficient of variation (CV) and the relative interquartile range (RIQR). The mean, median, and mode are used to compare the level of the distributions and the other measures are used to compare the pattern of the distributions.

Formulas for these measures were given in Chaps. 3 and 4; only the measures of dispersion are repeated here. For the interquartile range and the relative interquartile range:

$$\mathrm{IQR} = x_{.25} - x_{.75} \tag{6.1}$$

where $x_{.25}$ and $x_{.75}$ refer to the value of the variable (e.g., age at death) at the first and third quartile, and

$$\mathrm{RIQR} = (\mathrm{IQR} \div x_{md}) * 100 \tag{6.2}$$

where the IQR is divided by the median, or second quartile, so as to adjust for the level of the distribution and "isolate" the degree of dispersion. For the standard deviation and the coefficient of variation,

$$\sigma = \sqrt{\left[\left(\sum x_i^2 \div n\right) - x_m^2\right]} \tag{6.3}$$

$$CV = (\sigma \div x_m) * 100 \tag{6.4}$$

where, to derive the coefficient of variation, the standard deviation is divided by the mean. (To simplify the computations, the mean deviation ($= \sum_i (x_i - x_m) \div n$) may be used instead of the standard deviation; its standardized form is derived by dividing it by the mean of the distribution.) As noted earlier, the relative interquartile range and the coefficient of variation may be used to evaluate the differences in the shapes of the different distributions. It is important to use relative measures in comparing distributions that have very different levels of mortality or morbidity, such as distributions for two very widely separate dates or for different animal species.

Some measures adjust for the difference in the level of distributions being compared in a different way. Distributions can be "scaled" for comparison of their patterns by dividing or multiplying the elements in each distribution by a representative constant derived from the same distribution, such as its sum, mean, or median. After such scaling, the total, mean, or median of each distribution is reduced to a value of one. The same parameter, i.e., total, mean, or median, must be used for each distribution being compared. An alternative method is to multiply the elements of each distribution by the ratio of the mean or median of one of the distributions ("the standard distribution") to the mean or median of each of the other distributions. In effect, the "standard" distribution is left unchanged. Now all the distributions being compared have the "same" general level, scaled to the mean or median of the standard distribution.

Scaling is most appropriately applied in comparisons of age distributions of different species, where the lengths of life of the species being compared are often sharply different, e.g., deaths of humans and deaths of mice. Scaling may arguably be applied in comparisons of age distributions of two human population groups with sharply different life expectancies.

Comparing absolute and relative differences of health measures. In this and earlier chapters I describe many summary measures of the health status of population groups, e.g., the percents in categories of self-reported health, mean age at death, age-adjusted death rate, disease prevalence ratio, and life expectancy. In comparing such measures for two or more population groups, there are several choices. First, one can take the absolute difference between the measures. For this purpose one group must be selected as a standard or basis of comparison. One can select as the standard the value for the "leading" or "best" group (e.g., age-adjusted death rate in State A), to be compared with the value for the lagging group (e.g., age-adjusted death rate in State B). The absolute difference between the leading value and the lagging value provides an indication of the distance the lagging group has to go in order to close the gap between the two groups. Alternatively, the leading value may

be a future target value (e.g., life expectancy in 2015) and the other value may be the figure for the current date (e.g., life expectancy in 2005). The absolute difference between the target value and the current value provides an indication of the progress required to reach the target.

Often the choice of the target level or the leading group is arbitrary. The classification system used for the groups being compared (e.g., races) affects the choice of the leading group. For example, the amount of progress in reducing the differences between the age-adjusted death rates of the races would vary by the degree of disaggregation of the races (e.g., Asian-Americans as a group vs. Japanese, Chinese, Koreans, etc.). With the passage of time, the levels achieved by the various groups may change, the identity of the leading group may change, and the form of the computation may have to be changed.

Differences between groups can also be measured in relative terms (although some analysts object to taking ratios of rates in this way). The deviations from the leading value or target value can be expressed as a ratio of the baseline value or leading value:

$$\frac{\text{Target value} - \text{Baseline value}}{\text{Base line value}} * 100 \tag{6.5}$$

For example, given a life expectancy for black males in 2002 of 68.8, and a target value in 2010 of 75.5, the required progress is

$$\frac{75.5 - 68.8}{68.8} * 100 = 9.7\%$$

If the figure for black males in 2002 (68.8) is being compared with the figure for white males (75.1) in 2002, the percent shortfall of the black male figure is 8.4%. Here the base is the white male figure:

$$\frac{68.8 - 75.1}{75.1} * 100 = 8.4\%$$

Another measure, the progress quotient, shows the percent of the targeted change that has been achieved; that is, the difference between the current value and the baseline value is expressed as a proportion of the difference between the targeted value and the baseline value:

$$\frac{\text{Current value} - \text{baseline value}}{\text{Target value} - \text{baseline value}} * 100 \tag{6.6}$$

For example, given a life expectancy for black males in 1998 of 67.6, a current figure in 2002 of 68.8, and a target value in 2010 of 75.5, the progress quotient is:

$$\frac{68.8 - 67.6}{75.5 - 67.6} * 100 = 15\%$$

Normally this measure ranges from 0.0% to 100.0%, the first figure implying no progress and the second, achievement of the target. The result would be negative, however, if the current value moves away from the target. Because of the often small size of the changes, the progress quotient may change irregularly from year to year, and if the target value is only slightly greater than the baseline value, the instability in the measure gives the results an uncertain meaning. The current value may already have exceeded the target value, and in this case the result would exceed 100%.

Absolute and relative measures of differences may lead to different conclusions about changes over time. That is, equal relative progress for different groups does not correspond to equal absolute progress for them. The absolute difference between two groups in some health measure may decline while the relative difference may increase, depending on how the denominator is selected (as in the example below). In the calculation of the relative measure, the base chosen for calculating the percent difference at some date is usually the value for the group taken as the standard but this does not have to be the case. It may be better to use a more neutral and a more stable figure in the denominator, such as the value for all groups combined that are being compared. The percent differences in such a case would vary with the base, resulting in a difference in the measured health of the two populations:

A. Comparing the absolute and relative differences between absolute changes in percents:

$$21.5 - 4.2 = 17.3$$

$$10.8 - 1.9 = 8.9$$

$$\text{Absolute difference} = +8.4 \text{ percentage points}$$

$$\text{Relative difference} = +94\%$$

B. Computing the relative difference between the relative changes in percents:

$$\frac{21.5 - 4.2}{4.2} = +312\% \quad \frac{10.8 - 1.9}{1.9} = +368\%$$

$$\text{Relative difference} = 312/368 = -15\%$$

Using the average for all groups in the denominator:

$$\frac{21.5 - 4.2}{6.8} = +154\% \quad \frac{10.8 - 1.9}{3.2} = +178\%$$

$$\text{Relative difference} = 154/278 = -45\%$$

Here the relative measure reverses the direction of the absolute measure and the shift in the base of the percents changes the measured relative difference from −15% to −45%. Usually in making comparisons of general summary rates or ratios, the rates or ratios are age-adjusted before the comparison in order to eliminate the effect of differences in age composition from distorting the interpretation of the difference between the rates or ratios.

Measures That Link Two Populations

The above measures are computed for single populations, which are then compared with one another in order to describe differences in their levels of health. Some measures directly describe the differences in health between two populations, being computed from data linking the populations. The populations being compared may be segments of the population of countries or whole countries. The three most widely applied measures of health inequality that link two populations are quintile analysis, the index of dissimilarity, and the Gini Ratio, with its graphic counterpart, the Lorenz curve. The index of dissimilarity, the Gini Ratio, and the Lorenz curve are mathematically related to one another and so can be considered a single package of measures of inequality.

Quintile analysis. In quintile analysis, the question posed is, what percent of the aggregate of the health characteristic or percent of the population with the health characteristic is associated with each fifth of the population? As a hypothetical example, the "lowest" quintile of the population 25 years and over may be associated with 24% of the aggregate of the body mass indexes (BMI, a measure of body weight in relation to height) of the members of that population in a given year while the "highest" quintile of that population may be associated with only 16% of the BMI aggregate. In other words, the health indicator BMI is not distributed evenly in the adult population, i.e., 20% of the "weight" for each fifth of the population. One could also use quintile analysis to characterize a population in terms of the distribution of blood pressure, serum cholesterol, or other quantitative health variable that can be ranked numerically, that is, interval variables.

To do a quintile analysis relating to weight, one would rank each individual in the population according to their body mass index, and count off the top 20% of the population, then the 2nd 20% of the population, and then the 3rd, 4th, and 5th 20%s. One would then take the percent of the total aggregate of the BMIs for each fifth of the population ranked according to BMIs. The relation of population and aggregate BMIs in percents could then be stated for each fifth of the population for a series of years and represented in a 100% surface chart with each of the four graphed lines in the chart distinguishing each quintile's shares over time (Fig. 6.1).

Index of dissimilarity. The index of dissimilarity measures the inequality of two distributions with respect to some variable. The formula of the index of dissimilarity is

$$\mathrm{ID} = 1/2 \sum (\mathrm{p_i} - \mathrm{p_a}) \tag{6.7}$$

where p_i and p_a represent the percent of the health variable (e.g., obese persons and diabetic persons) in relation to their sum (all obese persons and all diabetic persons) for each of two distributions representing categories of subnational areas (e.g., census tracts of a city, states) or some demographic or socioeconomic variable (e.g., age classes).

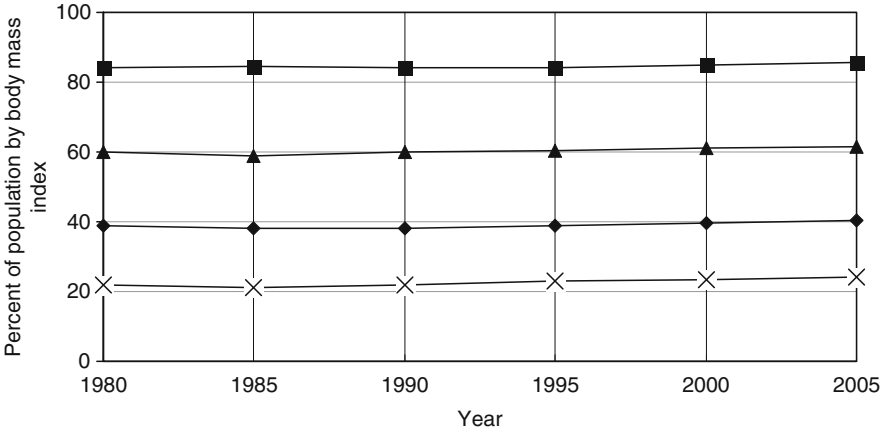


Fig. 6.1 Chart illustrating quintile analysis of body mass index in a population 18 years of age and over (Assumes that the body mass index of the members of a population or a representative sample of them is known every fifth year over 25 years. The population is distributed into fifths according to body mass index, one fifth with the lowest BMIs, one fifth with the next highest BMIs, and so on until the fifth of the population with the highest BMIs. The proportion of the BMIs total corresponding to each fifth of the population is then plotted for each year. These are hypothetical data)

Examples of such distributions that may be compared are numerous: Endogenous deaths by age vs. exogenous deaths by age; overweight persons by age vs. non-overweight persons by age; diabetic persons by age vs. nondiabetic persons by age; HIV/AIDS cases in States vs. single adult males without HIV/AIDS in States; and self-reported health (e.g., poor or fair vs. good, very good, or excellent) for census tracts within a city. The index of dissimilarity in this last example is calculated to determine the degree of inequality in the geographic distribution of healthy vs. unhealthy persons in a city. Indexes of dissimilarity may then be calculated for other cities to compare the degree of geographic concentration of persons with good or bad health in a group of various cities. This measure can be calculated for any variables, including non-quantitative, "non-rankable" attributes, i.e., nominal variables.

To derive the index, the absolute numbers in each distribution (e.g., obese persons, diabetic persons) are converted to percents of their totals, the differences between percents for corresponding categories (e.g., census tracts) are taken, these differences are summed over all areas, and then the sum is reduced by half. Table 6.1 provides an illustration of the calculation of the index of dissimilarity. It compares, with hypothetical data, the distributions of adult obesity and diabetes for the census tracts in a county. The index is low because obesity and diabetes tend to be distributed geographically in a similar fashion.

The index of dissimilarity (ID) varies between zero and one, zero representing perfect equality between the distributions and one representing perfect inequality between them. With perfect equality every geographic area or every age has the

Table 6.1 Calculation of the index of dissimilarity between census tracts, comparing the distribution of the number of diabetes and the number of cases of adult obesity: Hypothetical data

Tract	Cases of diabetes (1)	Cases of adult obesity (2)	Percent of total Diabetes (3)	Adult obesity (4)	Absolute difference (5) (3)–(4)
All tracts	562	1760	100.0	100.0	–
A	75	293	13.3	16.6	3.3
B	143	495	25.4	28.1	2.7
C	125	361	22.2	20.5	1.7
D	98	357	17.4	20.3	2.9
E	121	254	21.5	14.4	7.1
Sum (without regard to signs)					17.7
Index of dissimilarity = 1/2 sum					8.8

identical percent for the two variables being compared; for example, the share of obese persons in each county of a state is the same as the share of diabetics in each county of the state, or the proportion of persons in excellent, very good, or good health is the same as the proportion of persons in fair or poor health in every census tract of a county. Perfect inequality requires extreme conditions, such as that all the healthy persons live in one census tract and all the unhealthy persons in other census tracts of the county. The value of ID indicates the amount by which the proportions in the first distribution would have to be shifted upwards for negative differences and downwards for positive differences to make the two distributions equal.

The index of dissimilarity has a number of limitations, such as being affected by the number of categories in the distribution and the choice of classes of the variable being compared (e.g., 5-year age groups vs. 10-year age groups; census tracts vs. health zones), and being limited to comparing only two variables at a time (e.g., healthy vs. unhealthy persons).[1]

Gini Ratio and Lorenz curve. The Gini Ratio (Gini index of concentration/Gini coefficient) is a summary measure of the inequality of distributions similar to the

[1]We may wish to compare distributions of rates by age (e.g., death rates) rather than absolute numbers (e.g., deaths). The index of dissimilarity may be adapted for this purpose. We simply treat the age-specific rates as absolute numbers in applying the formula; i.e., sum the rates, compute the percent distribution of the rates, and take the differences between the percents at each age. Converting the rates to percents tends to dampen the variation in the original data considerably. In deriving the percent distribution of the rates, the rates are multiplied in effect by the reciprocal of the total of the distribution.

In an alternative handling of age-specific death rates, they may be weighted by population before the sums and percent distributions are calculated. This is to allow for the fact that the rates at the highest ages dominate the distribution. The rates are in effect converted to absolute numbers of deaths by this weighting process. The weighting process will modify the resulting indexes of dissimilarity greatly.

index of dissimilarity and related to it. The Gini ratio like the ID ranges from zero to one. A ratio of one indicates perfect inequality and a ratio of zero indicates perfect equality.

The calculation of the Gini ratio may best be described geometrically. First, calculate the cumulative shares of an aggregate health measure corresponding to cumulative shares of persons. Next, plot the results on a rectangular grid, with persons from 0 to 100 (percent) on the x-axis and with the aggregate health measure from 0 to 100 (percent) on the y-axis.[2] This line is designated the line of inequality, or Lorenz curve. Third, draw a straight (diagonal) line from (0, 0) to (100, 100), representing a perfectly equal distribution of population and health. The Gini ratio represents the proportion of the total (triangular) area under the diagonal line that lies in the area between the diagonal line and the Lorenz curve (Fig. 6.2). In the

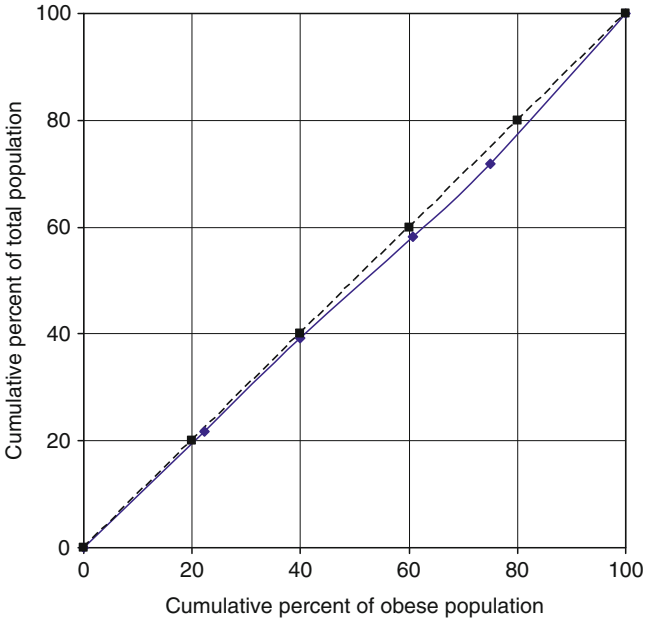


Fig. 6.2 Lorenz curve relating the income distribution of the population 18 years of age and over and the income distribution of the obese population 18 years and over: United States 2006. (The cumulative percent of the population is paired with the cumulative percent of obese persons, ranked according to household income, and plotted on a square grid. A diagonal line is drawn from (0, 0) to (100.0, 100.0), representing an equality of the distribution of population and obese persons according to income. The deviation of the line of inequality from the diagonal line represents the degree of inequality between the general population and obese persons according to income status. The figure shows a minimal degree of inequality) (Source: Adaptation by author of NHIS data for 2006)

[2] Adapted from Siegel, J. S. (2002). *Applied demography, applications to business, government, law, and public policy* (pp. 26–27), San Diego, CA: Academic Press (Fig. 1.5, p. 27).

figure, if B is assumed to represent the area under the Lorenz curve and A is assumed to represent the area between the diagonal line and the Lorenz curve, the Gini ratio is calculated as the share that A constitutes of (A + B). The maximum vertical distance between the line of equality and the Lorenz curve is equivalent to the index of dissimilarity.

The computing formula for the Gini ratio is,

$$\text{Gini Ratio} = \sum_{i} P_i Q_{i+1} - \sum_{i} P_{i+1} Q_i \tag{6.7}$$

where P_i and Q_i represent the cumulative proportions of individuals whose distributions with respect to the health variables are being compared.[3] The health variable can be a non-quantitative, nominal variable. Examples of such health variables are healthy persons, age at death, number of cases of a chronic illness, waist-hip ratio, body mass index, and blood pressure. A specific example would be a comparison of the distribution of high blood pressure of the general population and high blood pressure of children 5–14.

Measurement of International Inequality

Inequality of international health refers to the unequal distribution of good and bad health among countries as units. A variety of measures of inequality are used for measuring international health inequality similar to those used for intranational inequality in health and mortality. Goesling and Firebaugh (2004) also suggest that we could develop a world measure of inequality in population health by combining a global measure of inequality in population health within countries and a global measure of inequality in population health between countries or regions. World health inequality (I_w) would then be represented by the sum of the intracountry variation (I_i) and the intercountry variation (I_b):

$$I_w = I_b + I_i \tag{6.8}$$

There are many methods of analysis and many indicators that could be used to represent population health for international comparisons. I follow Goesling and Firebaugh (2004) in selecting life expectancy as the health variable and four measures of health inequality to illustrate the measurement of levels, differences, and trends in international health inequality. Life expectancy is a proxy for many aspects of health and data for it are virtually universally available. The illustration used here defines health inequality as the uneven distribution of life expectancy among world regions considered as units (with China and Japan listed separately).

[3] An illustration of the computation of the Gini Ratio is given in Siegel and Swanson (2004), Table 6.6, p. 117.

The four measures of health inequality employed here are the Gini coefficient, the Theil index, the mean logarithmic deviation, and the squared coefficient of variation. These measures are usually employed to measure inequality in income and other socioeconomic variables. They can be used to measure inequality for any variables that can be expressed on an interval scale, such as morbidity prevalence ratios, bed-days of activity limitation, and percent dying from a given disease. The formulas for all four indexes, using the notation of Goesling and Firebaugh, are structured as weighted averages of some function of the variable of interest. Here the variable of interest is the life expectancy ratio, defined as the ratio of life expectancy in a particular area (x_i) to the weighted average of life expectancies over all areas ($\sum_i p_i x_i$), the weights being the relative size (or percentage) of the population in each area (pi):

$$\text{Life expectancy ratio} = r_i = x_i \div \sum_i p_i\, x_i \tag{6.9}$$

Note that $\sum_i p_i = 1$. In calculating the various inequality indexes, we have to transform the life expectancy ratio into a specified mathematical function of the ratio and then average the transformed values of the life expectancy ratios by the populations in each area. That is:

$$\text{Measure of inequality} = \sum_i p_i\, f(r_i) \tag{6.10}$$

where r_i is the life expectancy ratio for a country or region and f is some mathematical function of it.

The Gini coefficient is defined as

$$GC = \sum_i p_i\, r_i (q_i - Q_i) \tag{6.11}$$

where q_i is the proportion of the total population over all areas in which the value of life expectancy is lower than in area i and Q_i is the proportion of the total population over all areas in which the value of life expectancy is higher than in area i.

$$TI = \sum_i p_i\ r_i\ \ln(r_i) \tag{6.12}$$

where the symbols r_i and p_i have the same meanings as above and the function of r_i is the product of r_i and the natural logarithm of r_i.

The mean logarithmic deviation is defined as

$$MLD = \sum_i p_i\ \ln(1/r_i) \tag{6.13}$$

where the logarithm of the reciprocal of r_i is the function of r_i used.

The squared coefficient of variation is defined as

$$CV^2 = \sum_i p_i \, (r_i - 1)^2 \tag{6.14}$$

where the function of r_i becomes $(r_i - 1)^2$.

If the x_i (i.e., life expectancies) are equal, $r_i = 1$. All of these indexes are standardized at zero when life expectancy is the same for all areas. Each index converges to zero as the life expectancy ratio approaches one ($r_i \rightarrow 1$); in this case there is a greater degree of equality among the areas studied. As the value of the index moves toward 1.0, there is a greater degree of inequality among the areas studied.

Table 6.2 shows the results of evaluating the four formulas for the geographic regions of the world with respect to inequality in longevity in 1980, 1990, and 2000, as calculated by Goesling and Firebaugh. Each measure implies a somewhat different degree of inequality and change in inequality over time. The Gini coefficient gives the least indication of inequality and the other three indexes are rather consistent in showing a major degree of inequality and change in inequality over the last few decades.

The mean logarithmic deviation showed a 21% decline in inequality between 1980 and 1990 and an increase of 26% between 1990 and 2000 (Table 6.2). The other three measures showed similar reversals in direction over the two decades, although the Gini coefficient indicated shifts in inequality from only −13% to +6%. Interregional inequality increased in the 1990–2000 decade mainly because life expectancy declined in sub-Saharan Africa (by 7 percent) as a result of the HIV/AIDS epidemic, and in the Transition Economies (by 2%) as a result of the disorder following the dissolution of the former Soviet Union. Accordingly, while global life expectancy continued to rise during both decades, the global rise in the 1990–2000 decade was small.

Principal Trends and Causes of Mortality

Trends in the United States

Age Changes

Life expectancy at birth in the United States has risen by about 9 years in the last half century and by about 30 years since 1900. Life expectancy at birth was 47 years in 1900, 68 in 1950, and 77 in 2000 (Table 6.3; Fig. 6.3). These numbers reflect an annual average absolute gain of 0.3 year and an annual average rate of gain of 0.5% over the century. The two halves of the century can be distinguished by their very different gains in life expectancy. Major shifts also occurred in the age incidence and

Table 6.2 World and regional trends in life expectancy at birth and in health inequality between regions: 1980–2000

	(Both sexes combined)				
	Life expectancy			Percent change	
Region[a]	1980	1990	2000	1980–1990	1990–2000
World	62.5	65.2	66.4	+4.3	+1.9
Western Europe	73.9	76.1	78.0	+3.1	+2.4
Transition economies	68.1	69.3	68.0	+1.8	−1.9
Western offshoots	73.8	75.5	77.4	+2.3	+2.4
Latin America and the Caribbean	64.7	68.0	70.4	+5.1	+3.6
Middle East and North Africa	59.2	65.0	68.5	+9.7	+5.4
Sub-Saharan Africa	47.6	50.0	46.5	+5.0	−7.0
South Asia	53.4	58.3	62.1	+9.1	+6.5
East Asia (excl. China and Japan)	59.7	65.1	68.5	+8.9	+5.2
Japan	76.1	78.8	80.7	+3.6	+2.4
China	66.8	68.9	70.3	+3.0	+2.0
Index of health inequality[b]					
Gini coefficient	.797	.691	.730	−13.3	+5.6
Theil index	.103	.080	.099	−22.3	+23.8
Mean logarithmic deviation	.107	.084	.106	−21.5	+26.2
Squared coefficient of variation	.200	.155	.185	−22.5	+19.4

Source: World Bank (2002). *World Development Indicators Online Database*: www.worldbank.org/data/onlinebs/onlinebases.htm. Secondary source: Goesling and Firebaugh (2004), Table 6.2, p. 137. Copyright © John Wiley & Sons. Reprinted with permission

Transition economies (22 countries): Former member countries of Soviet Union in Asia and socialist economies of Eastern Europe

Western offshoots (4 countries): United States, Australia, Canada, and New Zealand

Latin America and the Caribbean (29 countries)

Middle East and North Africa (20 countries)

Sub-Saharan Africa (45 countries)

South Asia (8 countries)

East Asia (excluding Japan and China): (13 countries)

[a]Western Europe (19 countries)

[b]See text for definitions

the causal pattern of mortality rates between the first half and the second half of the century. Prevention and control of infectious diseases explain much of the increase in life expectancy at birth in the first half of the century, and reductions in the chronic diseases of later life explain much of the increase in the second half of the century. In general, improvements in nutrition, personal hygiene, public sanitation (i.e., water supply and sewage disposal), and housing conditions account for most

Table 6.3 Life expectation at birth and at age 65, for the United States: 1900–1902 to 2005

Year	Life expectation		Increase over preceding date	
	At birth	At age 65	At birth	At age 65
1900–1902[a]	49.2	11.9	X	X
1939–1941	63.6	12.8	14.4	0.9
1949–1951	68.1	13.8	4.5	1.0
1959–1961	69.9	14.4	1.8	0.6
1969–1971	70.7	15.0	0.8	0.6
1979–1981	73.9	16.5	3.2	1.5
1989–1991	75.4	17.3	1.5	0.8
1999–2001	76.8	17.7	1.4	0.4
2005	77.8	18.7	1.0	1.0
Increase				
1900–1902 to 1949–1951			18.9	1.9
1949–1951 to 1999–2001			8.7	4.0

Source: Based on decennial life tables, 1900–1902 to 1999–2001, and annual life table for 2005 published by the U.S. National Center for Health Statistics or its predecessor agencies

[a]Original Death Registration States

X Not applicable

of the reduction in death rates during the first half of the century. The greater access to health care, medical advances, and adoption of a healthier lifestyle (e.g., reduced smoking) mainly account for the reductions in death rates during the second half of the century (Fried 2000). An element in this general reduction is the greater survival of people to age 65 in a condition of better health.

As a result of the timing and types of socioeconomic and medical developments during the last century, the distribution of the gain in life expectancy over the century at age 65 was quite different from the distribution of the gain in life expectancy at birth. While life expectancy at birth increased sharply early in the century, life expectancy at age 65 improved mainly after 1950 (Table 6.3; Fig. 6.3). Life expectancy at age 65 increased by 4.1 years during the second half of the twentieth century and only 1.9 years during the first half of the century. These absolute gains seem modest, but the increase of life expectancy at age 65 in relative terms over the whole century (51 percent) was almost as great as the relative increase in life expectancy at birth (56%).

Table 6.4 also illustrates this trend with data on changes in survival ratios and age-bounded life expectancies (i.e., average years lived in an age interval). The average years lived in the interval from age 65 to age 80 increased from 11.1 years to 12.5 years (out of a total of 15 possible years), that is, by 1.4 years, in the half century from 1950 to 2000, whereas in the previous half century the increase in average years lived in this interval amounted to only 1.0 year. In the birth-to-65-year interval, the comparable shifts were in the opposite direction, from 14.3 years in the earlier period to 3.5 years in the later period.

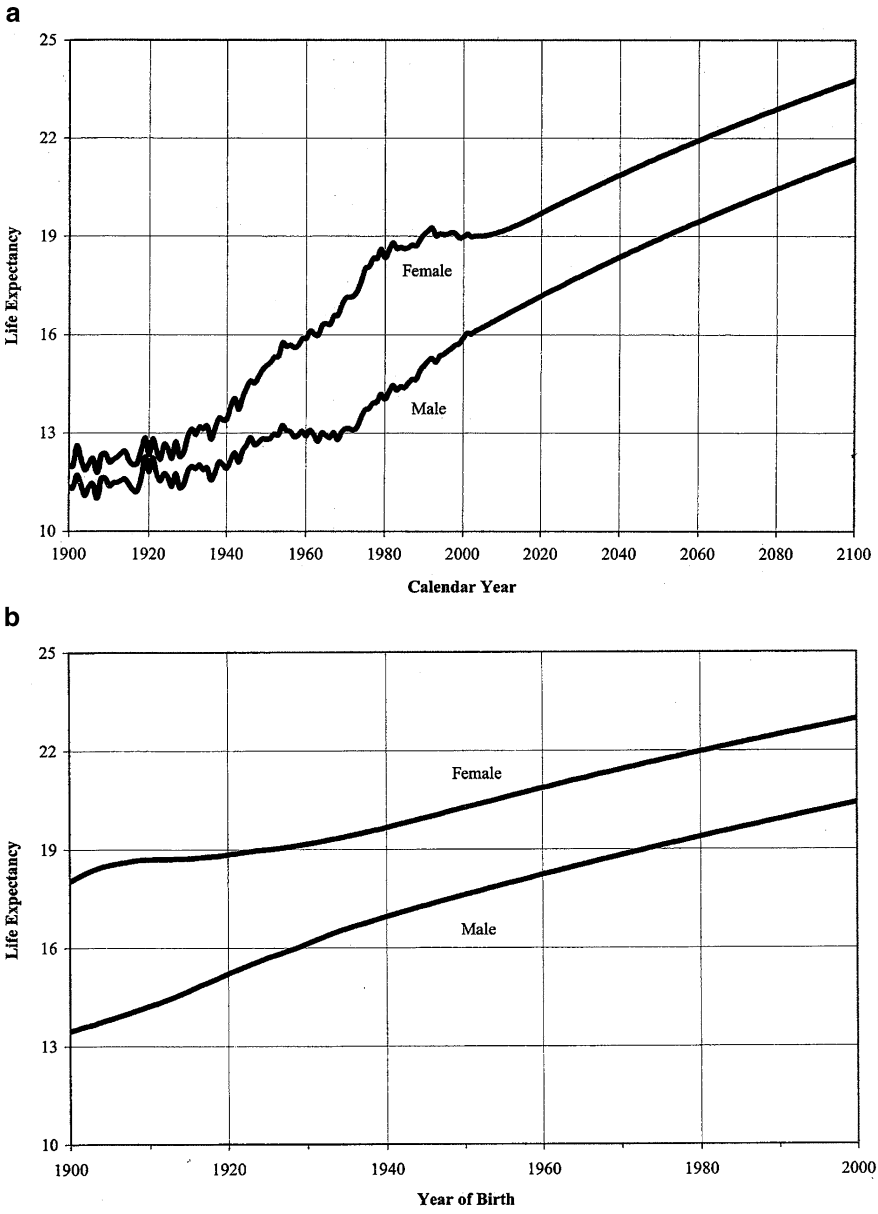


Fig. 6.3 Trend of life expectancy at age 65, by sex: U.S. Social Security Area

The reductions in death rates and the increases in survival ratios at the older ages (ages 65–80) have also increased greatly in relation to the corresponding changes for the younger age groups (ages under 65). Table 6.4 shows survival changes for the

Table 6.4 Percent surviving and age-bounded life expectancies, for the United States: 1900–1902 to 2005

	Percent surviving			Average years lived in interval			
	Birth to 65	Age 65–80	Age 80–95	Birth to 65	65–80	80–95	95+
1900–1902[a]	40.9	3.1	2.5	44.4	10.1	5.2	2.2
1939–1941	60.4	37.9	3.7	55.9	10.6	5.6	2.6
1949–1951	67.6	43.4	5.2	58.7	11.1	6.2	2.5
1959–1961	71.1	47.2	4.5	59.7	11.4	6.3	2.4
1969–1971	71.9	49.1	7.9	60.0	11.5	6.9	3.1
1979–1981	77.1	56.0	11.7	61.1	12.0	7.6	3.3
1989–1991	79.5	59.2	13.3	61.6	12.3	78.0	3.3
2000	82.1	62.1	14.2	62.2	12.5	8.1	3.5
2005	83.1	65.2	17.5	62.3	12.8	8.5	3.6
Increase							
1900–1950	26.7	10.3	2.7	14.3	1.0	1.0	0.3
1950–2000	14.5	18.7	9.0	3.5	1.4	1.9	1.0

Source: Based on decennial life tables, 1900–1902 to 1989–1991, and annual life tables for 2000 and 2005 published by the U.S. National Center for Health Statistics or its predecessor agencies
[a]Original Death Registration States

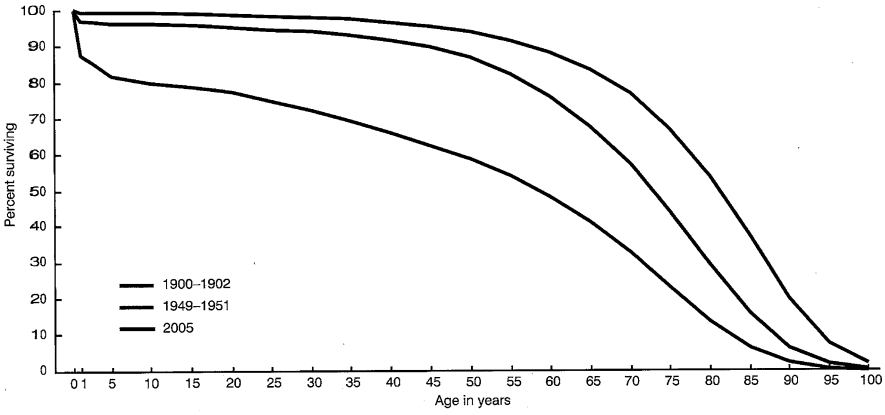


Fig. 6.4 Life table survival curves for the Original Death Registration States, 1900–1902, and the United States, 1949–1951 and 2005 (Source: U.S. National Center for Health Statistics, *National Vital Statistics*)

older and younger age groups during the periods 1900–1950 and 1950–2000. The change at the older ages during the earlier period was far smaller than at the younger ages (10 vs. 27% points), but during the later period the change at the older ages was much greater than at the younger ages (19 vs.14% points). (See also Fig. 6.4).

Leading Causes of Death

The four leading causes of death currently in the United States, accounting for two-thirds of all deaths, as reported by NCHS, are, in rank order, heart disease, malignant neoplasms (cancer), cerebrovascular disease (stroke), and chronic lower respiratory diseases. The next six leading causes are accidents, diabetes, Alzheimer's disease, influenza and pneumonia, kidney disease, and septicemia (i.e., blood poisoning). (See Table 6.5) A century earlier the top four ranks were held by influenza and pneumonia, tuberculosis, heart disease, and diarrhea in that order, and the next six leading causes included cerebrovascular disease, kidney disease, and malignant neoplasms (Linder and Grove 1947). Two conclusions can be drawn from these lists. First, at a time when the general level of mortality was far higher than it is today, infectious diseases headed the list of causes while today the list is headed by the chronic diseases of later life. Second, both then and now several chronic diseases, particularly heart disease, stroke, kidney disease, and cancer, are in the top ten. Because most of the leading causes of death are endogenous causes now, these same causes figure among the top 10 causes at ages 65 years and over currently (2003):

Cause[a]	Rate[b]
All causes	5023.4
1. Heart diseases	1568.5
2. Malignant neoplasms	1082.7
3 Cerebrovascular diseases	384.6
4. Chronic lower respiratory diseases	303.8
5. Alzheimer's disease	174.9
6. Influenza and pneumonia	160.6
7. Diabetes mellitus	152.9
8. Nephritis, nephrotic syndrome, and nephrosis	96.1
9. Accidents (unintentional injuries)	95.6
10. Septicemia	73.6
All other causes (residual)	928.1

Source: U.S. NCHS (2007)
[a]Cause of death based on *International Classification of Diseases, Tenth Revision,* (1992)
[b]Rate per 100,000 population 65 years and over

The combination of heart disease and cerebrovascular disease, the main cardiovascular diseases, accounts for over one-third of all of the deaths currently. This is true even though there has been a tremendous decrease – about 31% – in the death rate from this group of causes between 1970 and 2004 (U.S. NCHS 2007a). The reduction in the death rates from these and some other major chronic diseases of later life in the United States in this period was spectacularly large, unanticipated, and unprecedented. Mainly for this reason, overall mortality at the older ages declined by over one-third between 1970 and 2004. On the other hand, the pace

Table 6.5 Numbers and Rates for the 15 Leading Causes of Death, by Rank, for the Total Population of the United States: 2004

Rank[a]	Cause of death[b]	Code	Number	Percent of total deaths	Death rate[c]	
					Crude	Age-adjusted[d]
X	All causes	X	2,397,615	100.0	816.5	800.8
1	Diseases of heart	I00-I09, I11, I13, I20-I51	652,486	27.2	222.2	217.0
2	Malignant neoplasms	C00-C97	553,888	23.1	188.6	185.8
3	Cerebrovascular diseases	I60-I69	150,074	6.3	51.1	50.0
4	Chronic lower respiratory diseases	J40-J47	121,987	5.1	41.5	41.1
5	Accidents (unintentional injuries)	V01-X59, Y85-Y86	112,012	4.7	38.1	37.7
6	Diabetes mellitus	E10-E14	73,138	3.1	24.9	24.5
7	Alzheimer's disease	G30	65,965	2.8	22.5	21.8
8	Influenza and pneumonia	J10-J18	59,664	2.5	20.3	19.8
9	Nephritis, nephrotic syndrome, and nephrosis	N00-N07, N17-N19, N25-N27	42,480	1.8	14.5	14.2
10	Septicemia	A40-A41	33,373	1.4	11.4	11.2
11	Intentional self-harm (suicide)	X60-X84, Y87.0	32,439	1.4	11.0	10.9
12	Chronic liver disease and cirrhosis	K70, K73-K74	27,013	1.1	9.2	9.0
13	Essential (primary) hypertension and hypertensive renal disease	I10-I12	23,076	1.0	7.9	7.7
14	Parkinson's disease	G20-G21	17,989	0.8	6.1	6.1
15	Assault (homicide)	X85-Y09, Y87.1	17,357	0.7	5.9	5.9
X	All other causes	Residual	414,674	17.3	141.2	X

Source: U.S. NCHS/Miñino et al. (2007a), Table C, p. 8

X Category not applicable

[a]Rank based on number of deaths

[b]Based on the Tenth Revision, International Classification of Diseases, 1992

[c]Deaths per 100,000 population

[d]Based on the year 2000 standard population, i.e., total population in 2000 in 10-year age groups

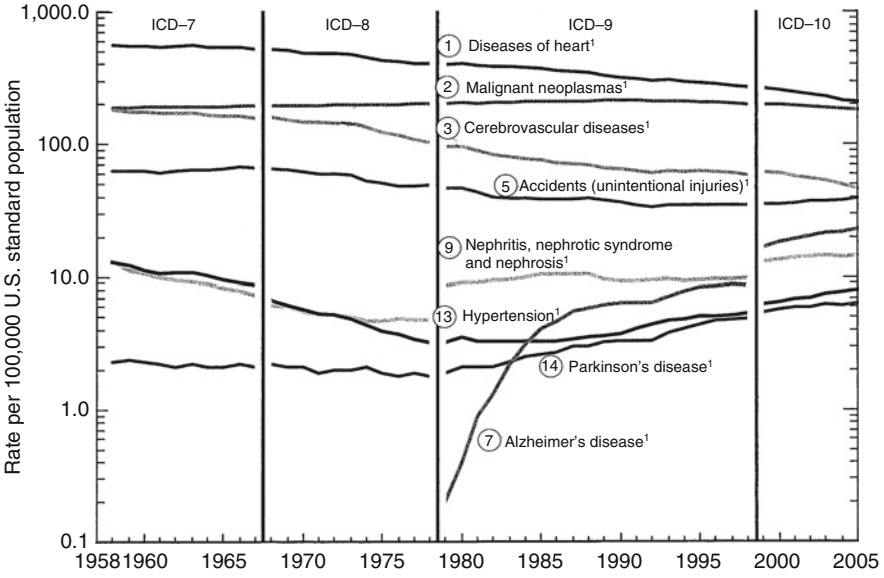


Fig. 6.5 Trend of age-adjusted death rates for selected leading causes of death, for the United States: 1958–2005 (Notes: *Circled numbers* indicate rank of the cause of death. Age-adjusted rates expressed per 100,000 standard population; Source: U.S. National Center for Health Statistics. (2008). Deaths: Final data for 2005. *National Vital Statistics Reports*, *56*(10), 2008, Fig. 5, p. 8)

of the decline of the major cardiovascular diseases has been slowing during these decades. The rate of decline in the death rate from these causes was 15% in the 1970s, 13% in the 1980s, and 8% in the 1990s.

During the same 34-year period the death rates for several other leading diseases have been increasing or changing little. The (age-adjusted) death rate from cancer has shown zero net change in this period, first increasing in the 1970s and 1980s and then decreasing in the 1990s (Fig. 6.5). The (age-adjusted) death rates from Alzheimer's disease, chronic obstructive pulmonary disease (COPD), influenza and pneumonia, diabetes, and a few other leading chronic diseases have risen in this period. Diabetes is reaching near-epidemic proportions (U.S. NCHS 2007a).

On the basis of the anticipated aging of the population, medical developments, lifestyle changes, and other factors, we can expect a shift in the future distribution of deaths by cause in the United States. Even if current cause-specific death rates do not change, expected population changes alone would result in some reordering of the causes. For example, if a hypothetical distribution of deaths by cause was obtained as the product of the middle series of projections of population for 2050 published by the U.S. Census Bureau and recent recorded age-sex-cause-specific death rates, all causes except heart disease, the other cardiovascular diseases, and diabetes would lose ground, relative to these three causes, between 2000 and 2050 (Sonnenschein and Brody 2005). To determine the shift in the pattern of causes resulting from changes both in age structure and in cause-specific death rates, we

need to combine population projections for 2050 with projections of mortality rates by cause-categories, age, and sex. Such data can be derived from the files of the U.S. Social Security Administration, which prepares its U.S. population projections by making assumptions about future cause-specific mortality. In the industrialized countries a few causes of death now dominate the cause-pattern of mortality. Some notes on these leading causes are set forth below, drawing mainly from the lists of the 15 leading causes of death at each age in the United States published annually by the NCHS (Table 6.5; U.S. NCHS 2007b).

Heart disease. Heart disease is a leading cause of death at every age, but at the older ages this cause is consistently the front-runner. Over one-quarter of the deaths in any year is due to heart disease and one out of two to three persons will eventually die from the condition. Heart disease is often associated with and results from several other chronic diseases such as hypertension, atherosclerosis, and diabetes, but at a more basic level, factors such as stress, unhealthful diet, poor lifestyle, and adverse environments, contribute to it and to the other chronic diseases associated with it.

Cancer. Cancer, the second most common cause of death, is a leading cause at every age, but it afflicts persons in the late middle ages and early old age particularly often. At the most advanced ages the rate slacks off somewhat and is greatly exceeded by the death rate for heart disease. Nearly one-quarter of the deaths in any year is due to cancer and one out of four or five persons will eventually die of this disease. Cancer results from a variety of causes, not a single cause, but all cancers have a common characteristic, the uncontrolled proliferation of cells. Uncontrolled growth may result from a single defective or missing gene or, more commonly, from cumulative mutations that affect numerous genes and that ultimately destroy the integrity and functioning of cells.

Cerebrovascular disease ("stroke"). Stroke is the third most common cause of death. A stroke is caused by the rupture of a blood vessel in the brain as a result of an aneurysm (i.e., a bulge in an artery wall) or by the blockage of a blood vessel in the brain as a result of a blood clot. It may cause paralysis of part of the body, with loss of muscular control, aphasia (i.e., inability to speak), and memory loss. Most stroke victims are over 65 years of age. It is a leading cause of death at nearly all ages (except 1–4), even though its rate is quite low at the ages below 65 (e.g., 5.5 per 100,000 population at ages 35–44). The chance of ever dying from stroke is about 1 out of 10 and about 6% of all deaths are due to stroke.

Chronic lower respiratory diseases. Chronic lower respiratory diseases represent another group of endogenous (or intrinsic) causes of death. They have proven extremely difficult to reduce, if not eliminate. This group of causes of death is also referred to as COPD, or chronic obstructive pulmonary disease, and includes bronchitis, emphysema, asthma, and other chronic lower respiratory diseases. It appears in the list of the 10 leading causes at most ages (excluding the ages 25–44 years). High in infancy, it falls low on the list or disappears in the intermediate ages and then moves up again in later years.

Accidents. Unintentional injuries, or accidents, is one of the 10 leading causes at every age but, with advancing age, its position falls farther and farther back in the list. It still accounts for nearly 5% of all deaths at ages 65 years and over, however. The accident rate has been gradually increasing in the decade 1995–2005 after a period of substantial decrease.

Diabetes. Diabetes is now the sixth leading cause of death in the United States. Diabetes appears in the list of the 10 leading causes as early as ages 20–24, its rate rises rapidly after these ages, and it remains among the leading causes at all higher ages (U.S. NCHS 2007b). Nearly one in three people in the United States will develop diabetes in their lifetimes and one in 50 persons will die from it. Type 1 diabetes results when the body cannot produce insulin, a hormone needed to convert food into energy, and first appears usually among children. Type 2 diabetes, which covers 90% of all cases, results from insulin resistance or deficiency. Levels of diabetes are increasing in the United States and around the world. The greater prevalence of diabetes is associated with the tremendous increase in obesity in the last several decades. A family history of obesity or diabetes, that is, having parents or siblings with these conditions, is a risk factor for diabetes both in childhood and adult life. Diabetes is associated with and is cause of several other chronic conditions, including heart disease, stroke, blindness, and kidney disease.

International Trends

Variations in Life Expectancy

Wide disparities exist among the mortality levels of individual regions and countries, as noted earlier. With a figure of 78 years for life expectancy at birth in 2005, United States was far from first in rank according to this measure of health (Table 6.6). Japan, Australia, Canada, Costa Rica, Israel, and much of Northern, Western, and Southern Europe (e.g., Austria, France, Italy, Spain, Sweden, Netherlands, and Norway) are in front, leading the United States by a year or more. Japan leads the countries of the world in life expectancy with a figure of 82 years and hence surpasses the United States by 4 years. This has been achieved in part by its low levels of heart disease and cerebrovascular disease (Yanagishita and Guralnik 1988).

The five countries with the highest and lowest life expectancies at birth in 2005 are:

Highest		Lowest	
Japan	82	Botswana	35
Iceland	81	Lesotho	35
Sweden	81	Swaziland	35
Australia	80	Zambia	37
Canada	80	Angola	40

Table 6.6 Life expectancy at birth and at age 65, by sex and by rank, for selected countries: 1980 and 2002

Country	Male			Female		
At birth	1980	2002	Rank	1980	2002	Rank
Australia	71.0	77.4	6	78.1	82.6	7
Bulgaria	68.5	68.9	34	73.9	75.6	35
Costa Rica	71.9	76.2	12	77.0	81.0	18
Cuba	72.2	74.7	25	NA	79.2	28
Finland	69.2	74.9	23	77.6	81.5	11
Hungary	65.5	68.4	35	72.7	76.7	34
Israel	72.2	77.5	5	75.8	81.4	13
Japan	73.4	78.3	2[a]	78.8	85.2	1
Norway	72.3	76.4	10	79.2	81.5	11
Russian Fed.	61.4	58.9	37	73.0	72.0	37
Slovakia	66.8	68.9	33	74.3	77.8	33
United States	70.0	74.5	26	78.8	79.9	26
At age 65						
Australia	13.7	17.4	4	17.9	20.8	5
Bulgaria	68.5	68.9	34	12.7	13.1	33
Costa Rica	16.1	17.8	2	18.1	20.5	8
Cuba	NA	16.8	10	NA	19.3	19
Finland	12.5	15.8	22	16.5	19.6	17
Hungary	11.6	13.1	33	14.6	17.0	32
Israel	14.4	17.3	6	15.8	19.7	12
Japan	14.6	18.0	1	17.7	23.0	1
Norway	14.3	16.2	18	18.0	19.7	12
Russian Fed.	11.6	10.9	36	15.6	15.1	37
Slovakia	12.3	13.3	32	15.4	17.0	32
United States	14.1	16.6	13	18.3	19.5	18

Source: U.S. National Center for Health Statistics (2006d), Table 26. For primary sources, see this publication

Note: Ranks are from highest to lowest life expectancy for the most recent year available. Since calculation of life expectancy varies among countries, comparisons among them should be made with caution. Countries with the same life expectancy receive the same rank. The country with the next lower life expectancy is assigned the rank it would have received had the higher-ranked countries not been tied

NA Not available

[a]Hong Kong holds rank 1

Then, among the highest, come France, Italy, Norway, Spain, and Switzerland, all with a life expectancy of 80, and among the lowest, Sierra Leone with 40, Zimbabwe, with 41, and Afghanistan, Liberia, and Mozambique with 42 (Population Reference Bureau 2005). The populations of the Newly Independent States (NIS) of the former Soviet Union in Eastern Europe and Asia (i.e., the Economies in Transition) have much lower life expectancies than the countries of Western Europe. In 2000, life expectancy in these countries averaged 68 years in comparison with 78 years in the West (Table 6.2; World Health Organization 2002).

Similarly wide variations in mortality exist in Latin America. In Latin America as a whole, life expectancy is about 72 years. Haiti has the lowest life expectancy, with 52 years, and Guatemala, Bolivia, and Guyana are at 66 years or less. Cuba, Costa Rica, Chile, Uruguay, Mexico, French Guiana, and Panama have the highest expectancies at 75 years or higher.

The region of greatest concern is sub-Saharan Africa. At its best in 1990 this region lagged considerably behind the other regions of the world (Table 6.2). The HIV/AIDS epidemic of the decades since the 1980s has pushed these countries farther behind the world average. Its life expectancy in 2005 was only 48 and some countries in the region have lost as much as 10 years of their previous progress (e.g., Kenya). In 2005 the gap in life expectancy between sub-Saharan Africa and Japan was 34 years as compared with 29 years in 1980.

In the early nineteenth century the range of life expectancy among countries was small in absolute terms (Goesling and Firebaugh 2004). Life expectancy at birth ranged from the low 20s to the low 40s. Large gains in life expectancy then occurred in areas where life expectancy was highest and, as a result, inequality in life expectancy increased. Inequality in regional life expectancy appears to have peaked between the two World Wars. At its peak, life expectancy was about twice as great in the West as in sub-Saharan Africa, South Asia, and China. Then, in a reversal, inequality sharply declined in the second half of the century, as the countries with the lowest life expectancy made the greatest gains. In recent decades, overall life expectancy in Europe has been increasing, but Eastern Europe has experienced a decline – a sharp decline in some countries. As a result, the gap among countries of the European continent widened in the 1990s after converging in the previous decade. As noted above, the drop in life expectancy in Russia and in other countries of Eastern Europe is a result of the social and economic disorder in these countries following the dissolution of the Soviet Union.

Great inequality in life expectancy among the countries of the world remains, but the variation is much smaller now than a half century ago. The divergence is reflected now in two main clusters, one consisting of the world's poorer countries converging around 45–50 years, and another consisting of the world's richer countries converging around 75–80 years (Goesling and Firebaugh 2004). As the countries of the European Union forge a common identity and establish a common economic union, we may expect to see a further convergence of the mortality levels of these countries. United Nations projections of life expectancy at birth for the major regions of the world for the first half of this century posit a considerable convergence of this measure during this period (Fig. 6.6).

Different combinations of demographic, socioeconomic, and environmental factors account for the rise in life expectancy in different countries during a particular period, even from the same initial level of mortality, and in the same country at different times. With the same apparent influences, mortality changes may vary across countries and across time at different rates. As noted, the declines in mortality that occurred in the LDC in the second half of the twentieth century were more rapid than those experienced in the MDC in the first half of the century.

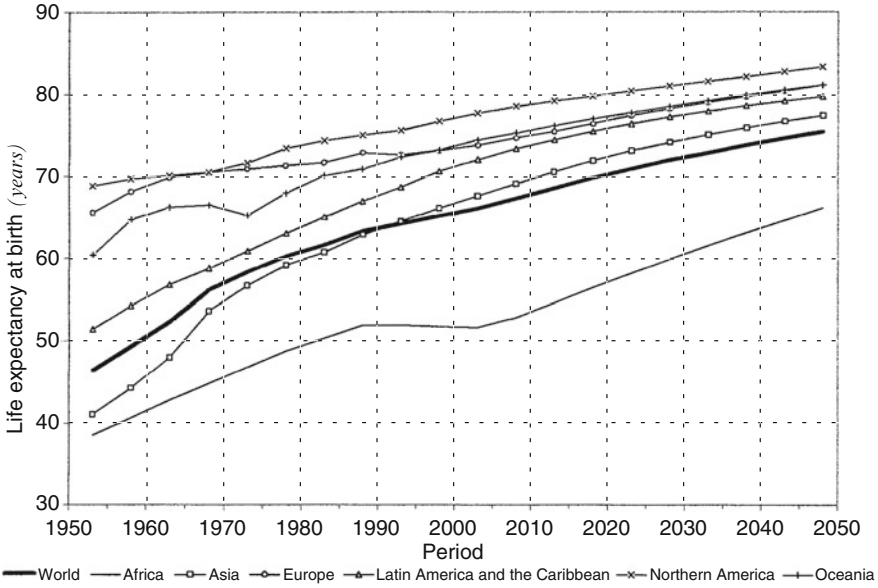


Fig. 6.6 Life expectancy at birth for the world and its major areas: 1950–2050 (Source: United Nations (2007))

The differences in mortality levels among the countries of the world are reflected in even greater variations in infant mortality. Infant mortality at 27 per 1,000 live births for Latin America is still high by the standards of the best countries, but far better than the average achieved by the less developed countries as a group (59) and by sub-Saharan Africa (94). These figures may be compared with the figure of 6.6 per 1,000 live births for the United States and 5.4 for Canada (Population Reference Bureau 2005).

Leading Causes of Death

Chronic degenerative diseases are no longer afflictions only of the affluent, developed societies, but have become global problems. Chronic diseases, especially the cardiovascular diseases, cancer, diabetes, and the chronic lower respiratory diseases, account for more than half of all deaths in the world (Yach et al. 2005). These diseases are leading causes of death in both the more developed countries (MDC) and the less developed countries (LDC). In spite of the much greater role of the infectious and parasitic diseases in the latter countries than in the former, the chronic degenerative diseases are now quite common in the LDC and their prevalence there is increasing rapidly. Yach et al. report that about three million deaths occur annually from cardiovascular diseases in both India and China.

Consumer practices and activity patterns have been changing in the LDC. Increasing use of tobacco, consumption of non-nutritious food, and pursuance of

less active lives have been elevating the risk of chronic disease, and hence the incidence of deaths from these conditions. One million tobacco-related deaths occur annually in China and 700,000 such deaths occur annually in India. As explained further below, an unhealthful lifestyle, including use of tobacco, poor diet, and lack of exercise, is a major risk factor contributing to the rise of chronic diseases, both in the LDC and MDC.

Most efforts at reduction of mortality in the world, such as the Millennium Development Goals of the United Nations, focus on selected infectious diseases and child health. However, a World Bank analysis of the most effective means to improve health in Eastern Europe and Central Asia concluded that measures to control cardiovascular diseases would contribute more to life expectancy than would measures implementing the Millennium Development Goals. According to Yach et al. (2005) this finding should probably be generalized to apply to many of the four billion people living in low- and middle-income countries.

Principal Causes and Trends of Morbidity

The principal causes of morbidity in the MDC are the major mental diseases (major depression, bipolar disease, and schizophrenia); the major neurological diseases of later life (Alzheimer's disease and Parkinson's disease); and the leading chronic physical diseases of later life (e.g., heart disease, cancer, cerebrovascular disease, hypertension, chronic lower respiratory disease, diabetes, arthritis, osteoporosis, and hepatitis). Neuropsychiatric (i.e., mental and neurological) disorders are the leading causes of disability from noncommunicable diseases in the world (World Health Organization 2002, 2003; Lopez et al. 2006). They account for 11.5% of the total disability load or disease burden (i.e., a weighted combination of years of life lost and years of disability). The United States experience follows a similar pattern. In the section below I consider several of these diseases individually with respect to their characteristics, prevalence, and associated conditions.

The principal causes of morbidity and disability in the LDC are HIV/AIDS, malaria, and tuberculosis. I mention these conditions only in passing in this chapter and discuss them more fully in Chap. 11, which treats special issues of health in the LDC.

Mental Diseases

The major mental disorders are schizophrenia, bipolar disorder, and major depression although others, such as post-traumatic stress disorder, autism, and eating disorders, are also common and serious. An estimated 13 million American adults (approximately 1 in 17 persons) have a seriously debilitating mental illness (Kessler et al. 2005). Perhaps 20 percent of the population of the United States suffers

from some type of mental illness and two-thirds of those afflicted go untreated. According to the World Health Organization (2003), mental health disorders are the leading cause of disability in the United States and Canada, accounting for 30% of all disability-adjusted life years (i.e., years lost due to disability and premature mortality).

The frequency of these disorders and the lack of treatment for most victims in part explains the high suicide rates for youth and all higher ages. In the United States suicide is the eleventh leading cause of death. It is the third leading cause of death among youths 15–24 years of age, following accidents and homicides, with a death rate of 10 per 100,000. The suicide rate rises generally with increasing age. The rate for ages 85 and over is three-quarters larger than the rate for ages 15–24 years of age. Adults 65 and over account for 17% of all deaths from suicide.

The major psychiatric disorders "run in families." The increased risk for developing the same disease among first-degree relatives varies from 3 to 10 times (autism, 150 times). Mental illnesses are generally now viewed as due to gene defects resulting in chemical imbalances in the brain. Multiple genes, where some are protective and others are pathogenic, and where some interact with other genes and some interact with the environment, appear to be involved.

Major Depression

Major depression is characterized by a feeling of dark gloom, sadness, inability to concentrate, inability to function in a disciplined way, and sleeplessness. About 10% of the U.S. population experiences depression, either chronic or episodic depression, severe enough to seek medical attention. The U.S. National Health Interview Survey reported that in 2004 14% of those 65 years old and over who had Medicare-only health-care coverage experienced feelings of sadness for all, most, or some of the time during the 30 days prior to the interview. This figure compares with 10% of those who had both Medicare and private health insurance. Among adults under 65, 17% of those without health insurance and 8% of those with private insurance reported such feelings. These figures probably greatly understate the facts, but we can say that a substantial share of the population suffers from depression and we can surmise that most of the affected individuals are going untreated (Table 6.7). Major depression ranks at the top of the neuropsychiatric disorders among the causes of disability in the world and is the first-ranked cause of disability in the Americas (WHO and World Bank 1996, 2003).

Depression can result from a familial tendency, side-effects of certain drugs or physical illnesses, changes in hormone levels, other mental disorders, and emotionally traumatic events, although there may be no apparent cause. To some extent major depression is familial, if not genetic, inasmuch as a person has three times the risk of being afflicted with the condition if a close relative has the illness. Depression increases the risks for a variety of other illnesses, including particularly heart disease, and is a principal cause of suicide.

Table 6.7 Age-adjusted percentages of persons 18 years old and over, in poor or fair health or with selected diseases, by sex and by age, for the United States: 2004

	Sex			Age group				
Disease	Total	Male	Female	18–44	45–64	65–74	75+	85+[a]
Fair or poor health[a]	26.0[b]	26.4[b]	25.7[b]		19.6[c]	22.9	28.5[d]	33.6
Heart disease	11.6	12.5	10.9	4.5	12.3	27.3	37.6	38.5
Coronary	6.4	8.3	4.9	1.1	6.9	18.4	26.1	
Hypertension	22.0	21.9	21.9	7.4	30.9	49.8	55.4	50.5
Stroke	2.6	2.8	2.4	0.5	2.5	6.9	12.4	
Emphysema	1.7	1.9	1.5	0.3	2.0	4.9	6.0	
Asthma[e]	6.7	5.0	8.2	6.4	7.0	7.5	6.6	
Chronic bronchitis	4.2	2.7	5.6	3.2	4.9	6.1	6.3	
Cancer, total	7.0	6.9	7.4	1.9	7.9	18.7	24.7	
Breast cancer	1.2	–	2.2	0.1	1.6	3.0	4.8	
Cervical cancer	1.0	–	1.0	0.9	1.2	0.8	0.9	
Prostate cancer	1.9	1.9	–	–	1.1	5.8	13.0	
Diabetes	7.1	7.6	6.6	2.0	10.1	18.9	16.4	11.0
Ulcers	6.9	7.1	6.8	4.5	8.2	10.9	13.3	
Kidney disease	1.7	1.5	1.9	0.9	1.8	3.4	4.9	
Liver disease	1.3	1.2	1.4	0.8	2.1	1.4	1.7	
Arthritis	21.6	18.5	24.2	8.0	28.8	46.5	55.8	
Chronic joint symptoms	26.8	25.3	28.1	15.7	35.0	44.7	49.8	
Migraines or severe headaches	15.3	9.7	20.2	18.4	15.0	7.1	5.1	
Pain in neck	14.6	12.1	17.0	12.4	18.7	13.9	15.0	
Pain in lower back	27.1	25.0	29.0	23.9	30.8	28.5	32.5	
Hearing trouble	16.4	20.1	13.2	7.7	18.5	31.7	48.9	58.0
Vision trouble	8.8	7.6	10.0	5.1	10.9	14.1	19.9	30.3
Edentulous[f]	7.9	7.7	8.1	1.8	8.5	21.3	30.7	40.2
Feelings of sadness[g]	3.3	2.7	3.9	3.1	3.8	2.5	3.5	
Feelings of hopelessness[g]	2.0	1.5	2.4	1.9	2.5	1.2	1.7	
Feelings of worthlessness[g]	1.8	1.5	2.1	1.6	2.1	1.2	2.5	
Feeling everything is an effort[g]	5.7	5.1	6.3	5.8	5.8	4.4	6.2	

Source: U.S. National Center for Health Statistics 2006a,b
Standard errors are given in the source tables
Blank spaces represent Not Available
– Not applicable or too small to be shown
[a]Annual average, 2000–2003
[b]65 years and over
[c]55–64 years of age
[d]75–84 years of age
[e]Currently has asthma
[f]Lacking all natural teeth
[g]All or most of the time

Bipolar (Manic-Depressive) Disorder

In bipolar disease a person's mood swings unpredictably between mania (uncontrolled excitement, unrealistic euphoria, and a sense of unlimited power) and depression (uncontrollable gloom). About 2% of the U.S. adult population, or more than 5 million people, suffer from bipolar disorder in the United States. Bipolar disorder has a largely genetic basis, with several susceptible genes being involved. With bipolar disorder there is a 7–10 times increased risk of developing the disease among first-degree relatives. With fraternal twins, if one twin has bipolar disorder, there is a 20% chance the other twin has it. With identical twins, this risk rises to 80%.

Schizophrenia

Schizophrenia is characterized by hallucinations or false sensory perceptions, delusions or false belief systems, disordered thinking, difficulty in functioning in interpersonal and occupational relations, and diminished motivation and drive. Schizophrenia ranks ninth worldwide among all the causes of disability and fourth worldwide among the major neuropsychiatric causes of disability. One percent of a general population tends to have schizophrenia. For persons with a sibling having the condition, the risk goes up to 10%; and for identical twins when one twin has the disease, the probability for the other twin is 50% (Malaspina 2002). The average age of onset of schizophrenia is 18 for men and 25 for women. An inherited genetic defect is one cause of schizophrenia but so also are acquired genetic defects resulting from accumulated mutations of genes in sperm cells, as demonstrated by the higher rate of schizophrenia among children of older fathers. Other contributing factors for schizophrenia include prenatal malnutrition, obstetric complications, and infections of the mother.

Neurological Conditions

Alzheimer's disease and Parkinson's disease are the most common, serious neurodegenerative diseases. They result from the deaths of particular groups of neurons in the brain, but the causes of these changes remain unresolved.

Alzheimer's Disease

Alzheimer's disease (AD) is characterized by progressive memory loss, cognitive deterioration, and dementia. It is a disease of later life and its prevalence rises sharply with advancing age after age 65. It is estimated that about 4.5 million people now have Alzheimer's disease in the United States (Herbert et al. 2003).

Some persons younger than age 65 show cognitive impairment but by ages 65–74 the number begins to rise sharply. About 5% of the population 65–74 years of age, 17% of those 75–84 years of age, and nearly half (46%) of the population 85 years and over suffer from it. The numbers are expected to grow dramatically as the population ages, especially among persons 85 years and over. Herbert et al. have projected that some 13.2 million older Americans (their middle projection series) will have Alzheimer's disease in 2050 unless new ways are found to prevent or treat the disease. Of these, 8.0 million will be 85 years old and over.

A death rate from AD of 2.0 (per 100,000 population) is now reported at ages 55–64. The reported rate rises sharply with advancing age. AD first appears among the 10 leading causes of death at ages 75–84, where it is sixth in order. Up to the early 1980s, AD hardly appeared as a cause of death. Apparently because of the growing number of cases of Alzheimer's disease in the United States and a growing public and medical awareness of the condition, it has become an increasingly popular diagnosis as an underlying cause of death reported on death certificates. Accordingly, the reported number and rate for AD have steadily risen in the United States over the last few decades. In 2006 72,400 persons were reported as dying from AD and only a decade earlier the number was 21,400. Its true rank as a cause of death in relation to the other causes is subject to question and needs further investigation. Calculations made by Ewbank (2004) would place it at a much higher rank – third among all causes over all ages – than the NCHS-reported rank of seventh. On the other hand, some neurologists would question the role of AD as an underlying cause of death.

At the present time the etiology of Alzheimer' disease is unknown. There is evidence that the disease has a genetic basis.[4] A healthy lifestyle and healthy behaviors, such as a healthy diet, regular exercise, a supportive social network, and intellectual pursuits, may lower one's risk of developing Alzheimer's disease. It has been shown too that the risk factors for heart disease and stroke (i.e., high readings on blood pressure, serum cholesterol, and homocysteine) are also risk factors for Alzheimer's disease. If the risk factors for stroke and AD are the same, common approaches are suggested for the prevention and treatment of vascular cognitive impairment and of Alzheimer's disease.

[4]Three mutated genes – amyloid precursor protein (APP), presenilin I, and presenilin 2—are known causes of inherited early-onset Alzheimer's disease. A gene for late onset AD (i.e., APOE) has been identified; it comes in three forms, conveying different degrees of vulnerability to the disease. The underlying biochemical changes associated with Alzheimer's disease are the accumulation of beta-amyloid peptide (i.e., a gummy protein that accumulates outside of nerve cells as plaque) and hyperphosphorylated tau protein (i.e., neurofibrillary tangles inside brain cells). These proteins build up in and around neurons in the neocortex and hippocampus, parts of the brain that control memory. When these neurons die, individuals lose their capacity to remember and their ability to do everyday tasks. As of now, the diagnosis of Alzheimer's disease is based on vague clinical criteria, such as simple memory tests and tests of efficiency in performing activities of daily living. A diagnosis can be confirmed only by biopsy or autopsy. Testing with brain imaging devices (MRIs, PET scans) and spinal taps are in the early stages.

Parkinson's Disease

Parkinson's disease is a neurodegenerative disease characterized by muscular tremors, impaired motor control, and muscular rigidity. It results from destruction of the neurons in brain cells that produce the neurotransmitter dopamine, but the underlying cause is unknown. These neurons die off in the basal ganglia, an area of the brain that controls body movements. As a result, the person can no longer control the movements of his or her body. It is much less common than Alzheimer's disease and appears 14th on the list of the 15 leading causes of death in the United States. Because it is concentrated among the older ages, its prevalence may be expected to grow rapidly in the next several decades.

Chronic Physical Diseases of Later Life

The chronic physical diseases of later life in combination with the neuropsychiatric illnesses described above account for a considerable share of the morbidity in the United States. According to the National Health Interview Survey, the lethal diseases among them alone account for a substantial share of the total for the chronic physical diseases. It reported that in 2004 12% of the U.S. population 18 years and over had heart disease, 22% had hypertension, and 3% have had a stroke (Table 6.7). Seven percent had some type of cancer and a similar percent had diabetes. Twelve percent had chronic lower respiratory diseases such as emphysema, asthma, and chronic bronchitis. In every case the prevalence ratio for the condition rises with advancing age. For example, for heart disease, at ages 18–44 it is 5%, but at ages 75 and over it is 38%. Over half (55%) of the population 75 years and over has hypertension, one quarter (25%) has some type of cancer, and one-sixth (16%) has diabetes. Many persons have two or more chronic conditions – so-called comorbidities – including most persons 80 years and over. (See Fig. 6.7.)

The prevalence of some of these diseases has been increasing in recent decades in the United States. Among them are diabetes, influenza and pneumonia, and chronic lower respiratory diseases. In particular, the prevalence ratio of diabetes has increased sharply in the 50 years since 1955.

The leading nonlethal chronic physical conditions are arthritis, osteoporosis, urinary incontinence, overactive bladder syndrome, and benign hyperplasia of prostate. Some notes on these conditions are given in the following paragraphs.

Osteoarthritis

Osteoarthritis is a chronic degenerative disorder in which the joints and bones erode because the cartilage separating the bones has deteriorated. It results from excessive wear or abuse, or an accident affecting the joints. Rheumatoid arthritis is a related but distinct disease. It is a chronic autoimmune disorder in which the joints erode

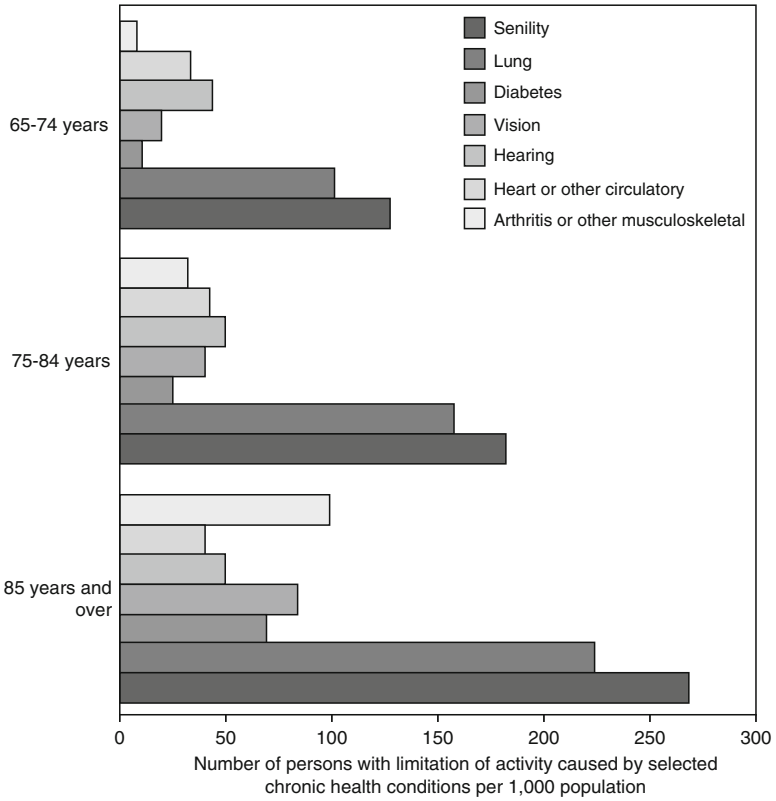


Fig. 6.7 Selected chronic health conditions causing limitation of activity among older adults by age, for the United States: 2003–2004 (Notes: Data are for the civilian noninstitutional population. Persons may report more than one health condition as the cause of their limitation; Source: NCHS (2006d). Primary source: National Health Interview Survey)

as a result of the immune system's attacking the body. Osteoarthritis is the more common form of arthritis.

While arthritis is not a cause of death, it results in inflammation and severe long-term pain and disability. It is the leading cause of disability. About one in every five persons 18 years and over and half the persons 75 years and over have arthritis (Table 6.7; U.S. NCHS 2006a,b). An estimated 47 million Americans 18 years and over, mostly middle-aged and older, have either osteoarthritis, rheumatoid arthritis, or other diseases of muscles, bones, or joints, such as gout, lupus, or fibromyalgia.

Osteoporosis

Osteoporosis is characterized by bone thinning and bone loss. Bones become more porous and fragile as well as thinner in later life. They continue to grow in later life

but bone loss begins to exceed bone growth; as a result, there is a net loss of bone. The change leaves bones vulnerable to fractures. Hip (femur), wrist (radius), and spinal (collapsed vertebrae) fractures are frequent consequences of osteoporosis. These fractures are painful and can impair one's ability to perform the activities of daily life. Bone loss is more typical for women than men because women have less bone mass at the start and bone loss begins at an earlier age for women. For women, the principal cause of osteoporosis is a deficiency of estrogen, associated with the postmenopausal years. Apart from age-related musculoskeletal changes, poor bone health usually results from lack of exercise, particularly weight-bearing exercise, and from inadequate amounts of calcium and vitamin D.

The combined facts that older women exercise less than older men and that they are also far more likely to be affected by osteoporosis and bone fractures than men are of considerable official concern. A Surgeon General's report on *Bone Health and Osteoporosis* published in 2004 warns that by 2020 half of all Americans over age 50 will be at risk of fracturing a bone because of osteoporosis.

Detection of osteoporosis and determining one's risk of incurring a bone fracture are uncomplicated. Inasmuch as the risk depends on bone mass and density, newer imaging methods can be employed to measure it easily. Examination of the skeleton by computer automated tomography (CAT scan) and other imaging methods can detect bone thinning and loss with minimal error.

Urinary Incontinence

Urinary continence is the unintended loss of urine to a degree that is significant enough to make it difficult to maintain good hygiene and carry on one's regular public activities. Its direct consequences are quite disruptive. It interferes with such daily activities as participation in social events and is responsible for sleep deprivation, mood disorders, and increased risks of mobility impairment and falls. It is a common cause of (personal) disability and hence a common reason for the transfer of older persons from home care to nursing-home care. An estimated 25 million adults in the United States have urinary incontinence.

Overactive Bladder Syndrome (OBS) and Benign Prostatic Hyperplasia (BPH)

An estimated 17 million Americans, mostly women, suffer from overactive bladder syndrome (OAB). Most are under the age of 55. Many do not treat their ailment, and as a result, limit their activities on the basis of the available bathroom facilities. Benign prostatic hyperplasia (BPH) is a noncancerous enlargement of the prostate gland in men, resulting in frequent and urgent need to urinate, difficulty in starting urination, and a weak urine stream. It is very common in men over the age of 50. Nocturia, the urgent need to urinate during sleep at night, is a frequent consequence of BPH. In disrupting sleep, it contributes to daytime fatigue and stress.

Common Infectious Diseases

Apart from the common childhood infectious diseases, many other infectious diseases are present in the MDC. They include HIV/AIDS, hepatitis, septicemia, and tuberculosis – diseases that are extremely common in the LDC. According to Hotez (2010), several parasitic diseases – toxocariasis, trichomoniasis, Chagas disease, and cysticercosis – flourish in the United States; they are found predominantly among the rural and urban poor. Brief notes on HIV/AIDS and hepatitis are given here and additional material on these and other infectious diseases is given in Chap. 11.

HIV/AIDS

HIV is the disease that is the precursor of the lethal illness AIDS. Some think of HIV as merely a risk factor for disease since many persons appear to live "disease-free" with HIV for many years. About two million persons (or less than 2/3 of one percent of the population) in the United States and around 33 million persons in the world are now infected with HIV/AIDS. In the United States men are more likely to be infected than women, and non-Hispanic blacks are more likely to be HIV-positive than all other race-ethnic groups. In 2006 12,100 persons died of HIV/AIDS in the United States, considerably fewer than a decade earlier, when 31,100 died. Two-thirds of the deaths are at ages 35–54. According to the United Nations, the HIV/AIDS epidemic is responsible for the deaths of three million people in the world each year, with the African continent being particularly stricken.

The basis of acquiring AIDS has been quite different for men and women. In the United States nearly half (48%) of the men with AIDS acquired it through male homosexual contact; 26% acquired it through infected needles; 6% through heterosexual contact not related to intravenous drug use; and 10% through sexual relations with a partner and use of intravenous drugs. Of the women with AIDS, 60% acquired it through intravenous drug use and 22% from unprotected sex with an infected partner. (See U.S. CDC 1998) The ways of acquiring AIDS in the less developed areas are quite different from the ways it is acquired in the West. In the LDC most cases are acquired by heterosexual contact.

Until recent developments in treatment modalities, nearly all HIV cases resulted in death about a decade or more after infection. Now effective treatment with a battery of retroviral drugs has converted it into a chronic infectious disease that may not end in death for several additional decades after infection. Since HIV/AIDS is a sexually transmitted disease, it has affected patterns of sexual behavior, marriage, family structure, and social dependency (e.g., frequency of orphanhood).

Hepatitis

There are three types of infectious hepatitis – A, B, and C. The various types are genetically distinct and are spread by different microbes. Hepatitis A is spread by an enteric pathogen, hepatitis A virus (HAV); and hepatitis B and hepatitis C are blood-borne, sexually transmitted diseases. Hepatitis A is the most widespread form of the disease (affecting 37% of adults 20–59 in 1988–1994), but hepatitis C (HCV) is of more concern because it has more serious health consequences. Hepatitis C is a potentially lethal virus that hits young men in the prime of life. In 1988–1994, 2.0 to 3.3 million Americans 20–59 years of age were infected by this virus and most did not know it (U.S. NCHS 2005).

Hepatitis C is spread by use of reusable shared syringes, accidental needle sticks with the blood of infected patients, and receipt of blood from an infected blood bank. About 80% of HIV patients are infected with hepatitis C, 87% of hemophilia patients treated before 1987 are infected, and 10% of dialysis patients are infected. Inoculation programs among former soldiers or the practice of tattooing may have spread the virus. Untreated, it can lead to cirrhosis and death. Current safeguards with blood banks, disposal of used medical paraphernalia, and other protections have reduced the risk to a negligible level in the United States and other Western countries; in 2006 only 7,250 persons were reported as dying of viral hepatitis in the United States. IV drug abuse, however, continues to pose a problem, as does the receipt of blood before 1992. The reported numbers of deaths and the death rate from hepatitis may be artificially low in the United States currently because some of the deaths due to hepatitis may be reported as deaths from cirrhosis.

Epidemiological and Health Transitions

The Epidemiological Transition

Over the course of the last few centuries the western industrial countries have experienced vast changes in the relative contributions of fertility and mortality to their population growth. These changes are collectively termed the demographic transition. They are encapsulated in three stages. In the first stage birth and death rates are high and population growth is low; in the second stage population growth is rapid as death rates fall but fertility remains high; and in the third stage, fertility falls and once again growth is low. These changes occurred largely in the eighteenth, the nineteenth, and the first half of the twentieth centuries. Many of the less developed countries have been moving through a somewhat similar demographic transition, but the transition began more recently, there is considerable variation among these countries in their experience of the stages, and the changes do not always conform to the classic pattern.

Paralleling the primary demographic transition in the more developed countries, but with a substantial lag, a pronounced shift has occurred in the pattern of the causes of mortality and morbidity. It began in the latter part of the nineteenth century. Extending the notion of a demographic transition, Omran used the phrase "epidemiological transition" to characterize the shifts in the cause-of-death and cause-of-morbidity patterns that occurred mainly in the last century. The cause patterns shifted from a predominance of acute, infectious and parasitic diseases to a predominance of chronic, degenerative diseases of later life, and accident-related and "self-imposed" conditions (Omran, 1971, 1977; Siegel 1993; Olshansky et al. 1997). The chronic degenerative diseases of later life are similar to the endogenous (intrinsic) diseases, which, as we may recall from Chap. 3, comprise such diseases as heart disease, cancer, diabetes, cerebrovascular disease, kidney diseases, and emphysema. The shift in concern for the acute infectious diseases to the chronic degenerative diseases has been accompanied by a shift in attention from the transmission of pathogenic agents to concerns, first, about the deterioration in physiological systems of the body and, then, about healthy behavior, healthy lifestyles, and health promotion. A similar change has been occurring in many less developed countries since the third quarter of the twentieth century (Omran 1971; Gribble and Preston 1993).

In Omran's design, the stage of the predominance of endogenous/chronic-degenerative diseases is really the third stage of the epidemiological transition. The first is the "age of pestilence and famine" during which mortality is high and variable and life expectation is under 30 years. In the second stage, termed the "age of receding pandemics," mortality falls sharply, life expectancy rises to about 50 years, and population grows briskly. In the last stage, mortality decline has slowed, communicable diseases virtually disappear, and the degenerative diseases rise to prominence.

Health Transitions

Viewing this three-stage model today, it is readily evident that some modifications and extensions are needed. In the MDC we have recently seen a sharp decline in several chronic, degenerative diseases, the resurgence of several infectious diseases, and the emergence of many new infectious diseases. Moreover, there are variations in the application of the Omran model, especially to the LDC. Accordingly, Olshansky and Ault (1986) proposed a fourth stage of the transition, the "age of delayed degenerative diseases," in which the ages for the onset of mortality from some of the leading degenerative diseases move upward as mortality declines. In this fourth stage of the transition rapid progress has occurred in the United States and other Western countries in the reduction of the rates for several chronic diseases of later life, especially heart disease and cerebrovascular disease. The spectacular and unanticipated declines in these rates that began in the late1960s have led to notable declines in the overall death rates at the higher ages and an acceleration in the increases in life expectancy at these ages.

We are readily able to document the rapid progress in the reduction of the rates for various causes of death, but documentation of the incidence and prevalence of the corresponding diseases is more difficult. Various national health surveys in the United States have revealed an increase in the prevalence of severe disability in the 1970s and decreases in the more recent decades. Whether these changes in disability levels correspond to delays in the onset of the degenerative diseases or an expansion of the period of morbidity in later life is not clear. One finding that questions U.S. progress in health comes from the Health and Retirement Study. Persons born between 1948 and 1953 (about 55–59 years of age in 2008) reported poorer health than the older cohorts, those born between 1936 and 1941 and 1942 and 1947. Their overall self-assessed health was not only poorer but they reported more difficulties in mobility and physical strength, more psychiatric problems, and higher levels on biomarkers such as blood pressure and cholesterol.

A fifth stage in the epidemiological transition appears to be developing with the resurgence of some of the infectious and parasitic diseases of a prior era, the increased incidence of others, and the emergence of new infectious diseases not known previously in the western world. To facilitate understanding of these infectious agents, Exhibit 6.1 sets forth a simplified classification of pathogenic microbes.

Emergence of New Infectious Diseases, Zoonoses

West Nile virus, ebola virus, monkeypox virus, severe acute respiratory syndrome (SARS), avian flu, and HIV/AIDS figure among the newly emerging diseases. Some of these diseases represent transplants from regions in Africa and Asia, areas where infectious agents have often been transmitted to humans from animals. Diseases that are transmitted to humans from animals are called zoonoses, They use a variety of vector paths and methods to accomplish this transfer.

The methods of transmission, with illustrations of diseases (not necessarily new diseases) transmitted from animals to humans, include:

1. Direct contact: Anthrax
2. Ingestion: Giardiasis
3. Inhalation: Q fever
4. Anthropod (e.g., tick) vector: Rocky Mountain spotted fever
5. Animal (e.g., raccoon) bite: Rabies

Many diseases are transmitted in more than one way. Most cannot be transmitted from human to human, only animal to human. A very general pattern of movement of the diseases from animals to humans consists of their becoming widespread or endemic among feral animals, then their transfer to domestic animals, including those domesticated animals that live indoors and other animals that live among humans (e.g., raccoons, bats), and finally their transfer to humans (Weinberg 2006).

Living microbes:

Prokaryotes: Unicellular microorganisms containing no nuclear membrane

Bacteria: Causes cholera, typhoid fever, tuberculosis, bacterial pneumonia, Lyme's disease, syphilis

Subtypes: Rickettsiae: Causes typhus, Rocky Mountain spotted fever

Ehrlichiae: Tick-born infections causing fever, headaches, and malaise

Archaea: A type of prokaryote that usually thrives on extreme temperatures and is classified as a separate domain in taxonomic systems on the basis of similarities of DNA sequences.

Eukaryotes: Microorganisms whose cells contain a membrane-bound nucleus

Subtype: Protozoa: unicellular organisms such as amoebas.

Non-living microbes:

Prions: Single protein molecule, containing no nucleic acid or DNA or RNA; that is, it lacks genetic material needed to replicate itself. Considered to be the infectious agent in certain diseases of the nervous system, e.g., Creutzfeldt- Jakob disease.

Viruses: Noncellular submicroscopic entities consisting of a core of RNA or DNA surrounded by a protein coat. Causes smallpox, measles, influenza, viral pneumonia, viral hepatitis; dengue, ebola virus, Marburg virus, West Nile virus, monkeypox virus; herpes virus, chickenpox, herpes simplex; human papilloma virus, avian flu. Some distinctive subtypes of viruses are:

Hantavirus: Causes hantavirus infection

Corona virus: Causes SARS

Retrovirus: HIV/AIDS virus: Converts its RNA into DNA by means of the enzyme reverse transcriptase. Causes HIV/AIDS.

Echovirus: Causes gastrointestinal tract diseases

Rotavirus: Causes gastroenteritis

Arbovirus: Causes viral encephalitis and dengue

Arenavirus: Causes meningitis

Rhinovirus: Causes common cold

Exhibit 6.1 Classification of pathogenic microbes, with illustrative associated diseases[a]

[a]Infections caused by parasites (e.g., cryptosporidiosis, giardiasis, schistosomiasis, onchocerciasis, and malaria) and fungi (*e.g.*, candidiasis) are not included

Another general pattern of transfer is from animals that fly, to terrestrial animals, and finally to people (e.g., SARS corona).[5]

A feature of the current, or fifth, epidemiological stage is the appearance of a new class of infectious pathogens called retroviruses. Among the retroviruses is human

[5]The distribution of zoonoses is affected by geoclimatic conditions (e.g., warming of waters), the availability of avian and aquatic hosts, the presence of anthropod vectors, migration patterns of animals and humans (including recreational and adventure travel), global trade involving the transfer of feed, food, animals (e.g., exotic pets), and inert conveyors, and the general influence of eco-and bio-systems. The risk of acquiring these diseases is increased for immune-compromised individuals who travel in areas where the disease is prevalent, persons who purchase exotic pets, and persons in areas where anthropod vectors abound. The problem of control is likely to become more difficult in the future because of people's pet fads, resistance of the newer bacteria to existing antibiotic drugs, the increase in the number of compromised persons, global warming, the invasion and clearing of previously undeveloped regions, and the increasing international migration of things, animals, and people (Weinberg 2006).

immunodeficiency virus, or HIV, the precursor to AIDS. The HIV/AIDS epidemic has resulted in a rise in the death rates in some regions, particularly sub-Saharan Africa.

Resurgence of Old Infectious Diseases and Rise in Existing Infectious Diseases

Malaria, whooping cough, typhoid, measles, and tuberculosis figure among the reemerging diseases in the industrial countries. At the same time the incidence of some existing infectious diseases, for example syphilis and influenza, has been rising. The syphilis rate rose in the current decade after it had declined to very low levels in the last quarter of the last century. This may reflect a new complacency with respect to protection against sexually transmitted diseases, and/or a phase of a disease cycle wherein immunity has diminished and the pathogenic agent has grown more powerful. The influenza rate has also been sharply rising in the last few decades.

One reason for the resurgence of bacterial infections in the industrial countries is the gradual impotence of the antibiotics that have been used to control them. Overuse of these antibiotics has encouraged the growth of microbes that are resistant to the control of antibiotics – so-called superbugs or superresistent microbes. Two such superbugs are C. diff. (*clostridium difficile*) and MRSA.[6] Few antibiotics are available to deal with this problem. Since the development of new antibiotics by pharmaceutical firms is not viewed as cost-effective, as compared with other medicinals that are used repetitively by the same individual, the companies are not generally engaged in developing them. Hence, new types of lethal bacterial infections are appearing for which there are no effective control agents.

Implications for the Trend of Mortality and Morbidity

An important implication of the epidemiological transition and the later health transitions, especially the new developments in infectious diseases, the rise in obesity and in diabetes, and the impact of smoking, is that a monotonic downward trend in mortality can no longer be taken for granted, even in the western world. This view of steady progress has been typical of both professional and lay thinking in the last half century, but the reality has been different. First, the historical record reflects pronounced fluctuations in the rate of decline of mortality. As we may recall, life expectancy in the United States remained virtually unchanged between 1954 and 1968, life expectancy among blacks in the United States moved downward in various years during the 1980s, and the rates for several chronic diseases of later life moved upward in the last few decades even while others moved sharply and unexpectedly downward.

[6]MRSA is the abbreviation for Methicillin-Resistant Staphylococcus Aureus, a strain of staphylococcus resistant to nearly all antibiotics and common in hospitals.

Not only have there been a number of major epidemics but they are continuing to occur. The influenza epidemic of 1918–1919 was responsible for the deaths of 40 million people around the world (Golini 2002) and about half a million people in the United States. Lesser influenza epidemics have occurred in the world and the United States since then. Tuberculosis, once considered under control, now kills 1.7 million persons in the world and several hundred persons in the United States each year.

A review of these developments has led to considerable questioning of the original epidemiological-transition model and to the proposal to rename the entire process the "health transition process." This term would suggest not a further stage in the epidemiological transition but the variety of changes in the pattern of causes of death and sickness that are occurring. For example, the term health transition has been used to refer to the "social, behavioral, cultural, and ideational changes that accompany mortality decline and improved health" (Caldwell and Santow 1989). More specifically, it has been used to refer to the recent interest of demographers in the inequalities in health levels, and the social and economic factors underlying these inequalities, especially how such factors as race and ethnicity, education, income, and marital status are related to health outcomes. The term, health outcomes, is defined broadly in this context to include death and disease, mental health as well as physical health, and wellness as well as sickness.

Summary Note on Developments in Mortality and Morbidity

It is apparent that the nature of mortality and morbidity is vastly different among today's population than among previous populations, among the old than among the young, and among the more advanced industrial populations than among the less developed countries. In the advanced industrial societies the epidemiological transition has replaced the major infectious diseases, usually acute illnesses of childhood and youth, with the chronic diseases of later life as the principal causes of sickness and death. As a result, mortality in the western countries is almost entirely due to chronic, degenerative diseases. This shift in causal patterns of sickness and death are beginning to characterize the less developed countries as well.

The old are likely to have multiple concurrent chronic diseases, or comorbidities, and with very advanced age, one or more functional limitations, or codysfunctionalities. Comorbidities lead to impaired functionality, decrease in biological homeostasis (i.e., the ability of an organism to maintain equilibrium of its physiological systems), and ultimately to death. Unlike the infectious diseases causing death at younger ages, the multiple morbidities of later life do not run their courses independently of one another. Rather, they are likely to have an interdependent and cause-and-effect relation with one another that intensifies the risks of dysfunctionality and mortality of the affected individuals.

Achieving acceptable levels of health both in the MDC and LDC appears to depend more on successful functioning and stability of the public health and sanitary

systems and the social and economic order than on new medical developments, just as did the shift from the first to the second stage of the demographic and epidemiological transitions. Yet, technological developments and medical advances are a major characteristic of the health systems of the MDC and are expected to grow in importance.

Factors Associated with Mortality and Morbidity

General Overview: Biological, Social, Individual, and Stochastic factors

Many statements have been put forward as to the way the risk factors accounting for diseases and deaths should be apportioned between genetic and other factors. For example, the National Research Council's, Institute of Medicine (2000) concluded, on the basis of available studies, that half of all deaths and the majority of diseases and disabilities in the United States are linked to behavioral and social factors (Goldman 2002). The National Institute on Aging of the National Institutes of Health announced that, while the relative roles of genetic and environmental factors in exceptional longevity are not known, life span studies indicate that genetic factors account for about 30% and environmental (i.e., nongenetic) factors for about 70% of the variation in longevity (Hodes 2005). Hjelmsberg et al. (2006) have reported that human family studies suggest that a similarly modest amount of the overall variation in adult life span (approximately 20–30%) can be accounted for by genetic factors.

A genetic component in human longevity can be immediately inferred from the widely different life spans of the different animal species, but the complexity of the issue becomes evident from the considerable variability in longevity within each species, especially humans. This variation implies that nongenetic (i.e., environmental and lifestyle) factors can be quite important as well. The roles of genetic, social (i.e, environmental), individual (i.e., behavioral and lifestyle), and stochastic (i.e., chance) factors in the causation of disease and mortality are complementary and interactive. These factors do not usually act independently but in innumerable complex combinations and variations, with different degrees of influence under different circumstances. Their relative impact differs for different parts of the age cycle, different groups in the population (e.g., sex, race, ethnic, marital, and socioeconomic groups), and different life experiences. Early life experiences affect health and mortality in later life as well as later-life experiences, greatly augmenting the role of environmental and behavioral factors in later life. The genetic influence on human life span over the generations varies with the genetic characteristics of the parents and their antecedents and the tendency of individuals to marry within or outside their racial/ethnic/socioeconomic group (i.e., homogamy/exogamy).

Biological Factors

Age is the leading biological risk factor for disease and death. For this reason the role of age in disease and death is considered often in this and other chapters. Among adults, with advancing age the risks of disease and death rise steadily. Like other ascribed characteristics, age (equivalently, date of birth) is a personal characteristic that does not change or that changes in a completely predetermined way, although the health consequences of age changes are usually modifiable to a substantial extent and they vary greatly from individual to individual. Age has important social dimensions in its relation to health and mortality as well, as when persons of different ages have different degrees of access to the health-care system or when persons at the older ages are treated differently by the health-care system than younger persons. Sex and race, like age, are also ascribed characteristics that have important biological and social dimensions. For the present work, I have grouped sex and race with such social characteristics as marital status and socioeconomic status, for the discussion of group variations in the next chapter.

Various types of biological influences in addition to the compositional factors of age, sex, and race affect morbidity and mortality. First is the influence of biological evolution (the "species" influence). Evolution has imposed some broad protections and restrictions on the processes of development of all humans that affect their health and longevity. At the same time, each family passes along a particular set of genetic characteristics ("family inheritance") that differentiate the genetic potential of its progeny from that of other families. Some of these familial genetic influences are essentially determinative in their consequences while others have a weaker potential whose expression in the progeny depends on the chance reactions of internal metabolic processes, environmental experiences, and life-style. In practice, it is difficult to distinguish the various types of biological influences in health outcomes but it is useful and important to try to do so.

Evolutionary (or "Species") Influences

Some evolutionary pressures, such as the increasingly limited role of natural selection after the peak of reproduction (i.e., late teen ages and early twenties), operate throughout the animal kingdom.[7] The probability of the occurrence of many adverse somatic conditions rises steadily with advancing age, especially after the reproductive period, because evolution does not work as consistently to protect against these risks as it does earlier in life. One could say that these risks are, in effect, "hard-wired" by evolution into the genes of humans. Carnes et al. (1996) and Carnes and Olshansky (1993) have described the stark difference in the course of the mortality rate for females, the effectiveness of natural selection, and the course

[7]This statement has some rare exceptions. They are described as cases of negative senescence.

of cumulative reproductive success. They note that evolution "designed" various genetic mechanisms to increase the likelihood of survival to reproductive age even under conditions of duress, and to maintain health and vigor long enough for the individual to reproduce. Evolution has not, however, designed any genetic program for aging or death in humans. "Aging is not a genetically controlled mechanism. It is an inadvertent byproduct of bodies surviving beyond their warranty period, i.e., the time required for reproduction, nurturing, and grandparenting" (Carnes 2005). Our genes do affect our longevity, however, working singly or in combination, to control (i.e., support or suppress) various metabolic and hormonal processes directly influencing our state of health.

Family Inheritance and Genetic Influences

Some health conditions are purely genetic in origin, being passed on to offspring through the specific genetic characteristics of the mother or father. Genetic diseases of this kind are likely to manifest themselves early in life although there are important exceptions. The stronger the role of genes in a disease, the earlier in life will its effects lead to disability and death. The risk of incurring some diseases and dying from them rises to about age 50 and then declines with increasing age. Cystic fibrosis, genetic dyslipidemic hypertension, and genetically determined cancers of the breast and lung are manifest by age 40 (Manton and Stallard 1994). Two of the genes that carry the mutation for breast cancer, called BRCA1 and BRCA2, account for early-onset cancer and are responsible for 5–10% of breast cancer cases. These genes are also implicated in hereditary ovarian cancer and prostate cancer. Genetically determined breast cancer is thought to be inherited as a dominant trait; that is, only one parent has to transmit it.[8] Sometimes, genes are inherited as recessive traits, that is, an affected individual has inherited one mutated (or non-functioning) gene for the disorder from each parent.[9] Cystic fibrosis and Gaucher's disease type I are recessive traits. There are many illustrations of familial genetic conditions that are passed on to progeny. Some examples of inherited genetic diseases are given in Exhibit 6.2.

Genetic disorders inherited from fathers. Cumulative genetic mutations of male germ cells account for a number of inherited genetic disorders. Male germ cells continually divide and errors accumulate during spermatogenesis as men age. These mutations are then passed on to their children. As a result, older fathers are far more

[8] A woman can inherit a BRCA mutation from her mother or father but she also has a 50% chance of not inheriting it at all.

[9] A person who has a mutation in only one gene of the pair of genes for a given condition inherited from the parents is a carrier. Carriers are healthy persons who do not develop the disorder but are at risk for passing on the mutated gene to their children. If both parents are carriers of the gene for the same condition, there is a 25% chance that their child will be affected with the condition, a 50% chance that the child will be a carrier, and a 25% chance that the child will be neither affected nor a carrier.

1. Phenylketonuria (PKU) is a rare genetic disease that causes mental retardation, deafness, and seizures. Immediate initiation of dietary therapy virtually eliminates PKU as a cause of mental retardation; continuing dietary therapy before and during pregnancy can prevent the problems associated with maternal PKU, allowing normal development of infants of affected mothers.
2. Non-classical congenital adrenal hyperplasia (CAH) is the most common single-gene disorder in the population. It causes a broad spectrum of disorders. There is a prenatal treatment that prevents the manifestations of CAH.
3. Congenital hypothyroidism is a metabolic condition that results from the loss of thryoid function and, if left untreated, causes irreversible mental retardation. Where newborns are screened for this condition, it can be eliminated as a cause of mental retardation.
4. Fragile X syndrome is the leading cause of inherited mental retardation. Its cause is a mutation in the Fragile X Mental Retardation 1 (FMR) gene found on the X chromosome. The mutation is an inheritance mechanism called a "Atriplet repeat," which gets larger in each generation and, when large enough, disrupts gene function.
5. Cystinosis is a rare genetic error of metabolism that can cause kidney failure and neurological damage. As a result of an error in the gene for Rett syndrome, infant girls gradually lose their language capabilities, mental functioning, and ability to interact with others. Taking the drug cysteamine bitartrate regularly completely prevents any manifestation of the condition.
6. Tay-Sachs disease is a fatal hereditary disease in which an enzyme deficiency leads to the accumulation of gangliosides in nervous tissue. It affects young children of eastern European Jewish descent almost exclusively.
7. Sickle-cell anemia is a hereditary form of chronic anemia caused by a mutant hemoglobin gene that makes the red blood cells become crescent in shape. The disease is characterized by fever, leg ulcers, severe joint pain, and jaundice, and affects black persons almost exclusively.
8. Hemophilia A is a bleeding disorder in which there is a deficiency of clotting factor VIII. The gene abnormality is sex-linked, that is, inherited through the mother and expressed in the male.
9. Polycystic kidney disease is a disorder in which many cysts form in both kidneys; as a result, the kidneys enlarge, some kidney tissue is destroyed, and functioning kidney tissue is lost.
10. Cystic fibrosis is a hereditary disease characterized by impaired respiratory and pancreatic function as a result of the production of abnormal secretions from many exocrine glands (*i.e.*, glands that secrete fluids into a duct). It is a recessive genetic disorder that disproportionately affects white persons as compared with other races.

Exhibit 6.2 Examples of inherited genetic diseases

likely than younger fathers to have children with inherited genetic disorders. Some of the disorders resulting from inherited genetic mutations associated with paternal age are as follows (*Narsad Research Newsletter* 2002):

Achondroplastic dwarfism: Improper development of cartilage at the ends of the long bones, resulting in a form of congenital dwarfism.

Marfan syndrome: A disorder principally affecting the connective tissue, manifested by excessive bone elongation and joint flexibility, and resulting in weakened tissues and abnormalities of the heart, blood vessels, lungs, bones, joints, and eyes.

Neurofibromatosis: Disease characterized by the formation of neurofibromas (i.e., soft fleshy benign growths of nerve tissue), sometimes accompanied by physical deformity (e.g., curved spine, rib deformities).

Osteogenesis imperfecta: Defective formation and development of bony tissue.

Advanced age of the father is an important risk factor for schizophrenia in children. Malaspina (2002) analyzed birth records that showed a strong escalation in the risk of schizophrenia in the child as the age of the father increased. Records of 92,000 pregnant women in Jerusalem were studied, including a linkage to the records of the psychiatric case registry. She found that the risk for children of fathers in their forties was double that for children of fathers under 25 years old, and that the risk was triple for children of fathers over 50. One of every 50 children of fathers over 50 suffered from the condition. Father's age alone accounted for one-fourth of all cases of schizophrenia in the population studied. Several other studies have yielded similar results.

Randomly Acquired Genetic Defects

Many more diseases result from randomly acquired genetic defects. During the long period from fertilization to old age, genetic damage to DNA and somatic cells occur and accumulate as a result of hostile internal and environmental influences and copying errors in the subdivision of the cells. These may manifest themselves as one of the chronic diseases of older age such as colon cancer or heart disease. Further, one's chances of developing such diseases double if one's parents or siblings are affected by these diseases. The risk is even higher for people with two affected family members.

As suggested earlier, most purely genetic defects are expressed by the time of the menopause and are lethal. By the older ages, the genetic conditions determining most health outcomes have been greatly modified by a "lifetime" of environmental and lifestyle influences as well as by internal metabolic, hormonal, and stochastic processes. Hence, it is reasonable to accept the conclusions of the National Institute of Health and others cited earlier that only a modest amount of the overall variation of the adult life span is accounted for by genetic factors.

Genetic Influences at the Very Oldest Ages

Genetics appear to play a different role in health risks at the very oldest ages than at the adult post-reproductive ages, however. Genetic factors may become increasingly important for survival at these ages. On the basis of studies of extinct cohorts of Danish, Swedish, and Finnish twins (20,502 individuals), Hjelmborg et al. (2006) found that having a twin survive to old age substantially and significantly "increased" the chances of reaching the same old age and that this chance is higher for monozygotic (MZ) twins than for dizygotic (DZ) twins. Specifically, the relative risk of reaching age 92 is 4.8 for MZ males and 1.8 for DZ males. Hjelmborg et al. concluded that genetic influences on life span are minimal prior to age 60 but increase thereafter, and interpreted this finding as support for the search for genes affecting longevity in humans, especially at advanced ages. (Further research on this issue is discussed in Chap.13.)

Biological Aging

Biogerontologists differentiate superficial bodily changes that occur with age, such as graying and thinning of the hair and wrinkling of the skin, from those changes that increase the risk of disease, disability, or death. They use the term biological aging to describe all these changes, especially the second group of changes. When such age-related vulnerabilities manifest themselves as disease or disability, biogerontologists call the process senescence. Alternatively, senescence may be defined as the progressive deterioration of bodily functions over time. This loss of function is associated with decreased fecundity (i.e., physiological ability to parent a child) and increased risk of morbidity and mortality as an individual gets older. The rate and progression of the process of somatic deterioration varies greatly from person to person and from bodily organ to bodily organ for any given person, but generally over time every organ of the body is affected. Since individuals differ greatly in the rate and progression of these deleterious processes, a group of persons of the same chronological age may differ greatly in their biological age.

Biogerontologists disagree as to when senescence begins. Some maintain that it begins at birth, others that it begins after the peak reproductive years, and still others that it begins at various other ages before age 45. In any case, the process leads to, or represents, a loss of bodily functions, increasing the probability of death and ending ultimately in death itself.

Aging as "cause" of disease. Of all the factors one can identify as possible contenders for the role of leading "cause" of the chronic diseases of later life, aging comes first. One cannot reflect on these diseases independent of the role of aging. Most biogerontologists maintain, however, that aging is not a disease even though aging is associated with many bodily changes that are considered diseases. They draw a distinction between normal aging and pathological aging, but blur the distinction by identifying numerous age-related conditions that compromise health as part of normal aging. Most "normal" (that is, usual or typical) age-related changes shift gradually with advancing age to pathological levels – that is, levels that are immediate threats to, or limit, functioning.

There is no easy basis for setting a dividing line between normal and pathological age-related somatic changes, even if measures of physiological performance or biomarkers are taken into account. Ling et al. (2007) conclude that state-of-the-art imaging and immunologic and biochemical methods are needed to analyze the process by which a chronic disease (e.g., osteoarthritis) develops and progresses as an individual ages, and to understand how complex the determination of the line between aging and disease is. While an important heuristic and philosophical distinction can be made between normal and pathological aging, it may not be productive to debate when, and even whether, certain gradual degenerative changes should be called normal aging or pathological aging. I discuss this topic in more detail in Chap. 13.

The principal physical changes that gradually occur in humans with age over an adult lifetime have been the object of study by the Baltimore Longitudinal Study

1. Thoracic area: Lung tissue loses much of its elasticity; muscles of the rib cage shrink. Maximum breathing (vital) capacity diminishes progressively beginning about age 20 and may decline by about two-fifths between the ages of 20 and 70.
2. Heart and vascular system: Blood vessels accumulate fatty deposits and lose much of their flexibility. With the resulting atherosclerosis, the heart needs to supply more force to propel the blood forward through the less elastic arteries. The heart muscle thickens; maximum oxygen consumption declines by about 10% for men and by about 7.5% for women with each decade in adult life. The heart's maximum pumping rate and the body's ability to extract oxygen from the blood decline.
3. Gastrointestinal system: Production of digestive enzymes diminishes and tissues lose much of their ability to catabolize and absorb foods properly.
4. Reproductive system: In women, production of vaginal fluid decreases and sexual tissues atrophy. In men, sperm production decreases and the prostate gland enlarges. Libido in both sexes declines gradually.
5. Urinary system: Kidneys gradually become less efficient at extracting wastes from the blood; bladder capacity declines. Individuals are at greater risk of urinary incontinence.
6. Musculoskeletal system: Without adequate exercise muscle mass declines over one-fifth between the ages of 30 and 70. Bone mineral is lost and replaced simultaneously until the loss begins to outstrip replacement around age 35; the loss is especially rapid for women after menopause (unless offset by regular weight-bearing exercise).
7. Neurological system:
   Brain: The brain loses some of the structures (axons) that connect nerve cells (neurons) to each other, and the ability of neurons to function may diminish. The flow of blood and the supply of oxygen diminish. Some types of memory, the ability to shift between tasks, reaction speed, and attention deteriorate. Other types of memory (*e.g.*, expert knowledge and access to cognitive "templates"), automatic functions, and the ability to focus on a task and resist distractions are improved in older age, and the brain's control of the senses is maintained rather well. Senses of vision and hearing: Various aspects of vision begin to decline at later ages: Focusing up close (in the forties) and distinguishing fine details (in the seventies); adjustment for glare, seeing at low levels of illumination, and detecting moving targets (in the fifties). Hearing acuity declines with age: Ability to hear higher frequencies is reduced; and understanding speech is more difficult, especially in situations where there is background noise.

   Other senses: Acuity of the other senses – smell, taste, touch, heat, pain, kinesthesia, proprioception – similarly declines in older age. A reduced "sense" of balance results from the aging of the vestibular system (inner ear including the nerve receptors in the semicircular canals). This sense also depends on the proper functioning of the eyes, ears, and the central nervous system.
8. Metabolic changes: Absorption of food is reduced, weight falls, and the individual tends to lose total body fat. However, fat is redistributed in the body to the hip-waist area and from just beneath the skin of the abdomen to deeper parts of the body. For women the shift is to the thighs and hips; for men the shift is to the abdomen.

Exhibit 6.3 Principal physical changes with advancing age

of Aging. They are enumerated in Exhibit 6.3. As noted there, every physiological system experiences a major loss of efficiency from youth to old age. Underlying these pervasive system changes is the fact that, on average, by age 75 or 80 the number of cells in the human body has decreased by 30% or more.

Environmental and Ecological Factors

There are many ways to set forth the nongenetic, or social and individual factors, affecting the risk of disease and death. Any attempt to do this on the basis of an assessment of their relative importance, except perhaps in particular cases of disease or death, is fraught with grave difficulties. They work collectively to influence health outcomes. Many are related to one another, some in a temporal or causal sequence. Some have a sort of imperial power, like smoking, obesity, or high blood pressure, to effect negative outcomes. Presumably these conditions could be put at the head of the list. Yet, for making positive gains in longevity, regular exercise comes up shining. Many of us were brought up to observe five valid health rules, prescribed in no particular order: Do not smoke, check your blood pressure, eat properly, get sufficient exercise, and choose your parents carefully. I do not attempt any ranking of the nongenetic risk factors here, only a simple enumeration of them, with brief notes regarding their role in disease in the United States and elsewhere today.

One group of influences affecting health, the environmental factors, are outside the control of the individual except as individuals are part of a group, act collectively, or experience them in common. They include community health programs, air and noise pollution, workplace conditions, food and water quality, exposure to infectious agents and radiation, and availability and quality of health services. I consider these environmental factors in more detail next.

Community Health Programs

Public health risks can be greatly influenced by the efforts of private and public agencies to inform and guide the public in its health choices and practices. Community agencies can sponsor public health information programs that promote and support responsible health choices and that are designed to reduce the burden of preventable disease. Such programs can motivate people, especially young people and older Americans, to make healthful choices regarding diet, nutrition, and exercise; and can help remove the stigma from reporting and treating certain health conditions, especially mental illness and sexually transmitted diseases. Community organizations can set up programs to train individuals to examine themselves for high blood-pressure, breast cancer, sexually transmitted diseases, and skin cancer. They can set up community clinics to carry out these examinations and can sponsor programs of training in prenatal care.

Air and Noise Pollution, Workplace Conditions, and Radiation Exposure

Air pollutants, such as sulphur dioxide, nitrogen dioxide, and mercury, emitted mostly by power plants and automobiles but by other sources as well, contribute to the formation of smog. These emissions react with sunlight to form ground-level

ozone, also part of smog. A correlation has been found between ground-level ozone and death rates in cities. More than 150 million Americans live in areas with officially excessive levels of ozone.

The workplace may expose workers to excessive noise; air pollution, including dust, smoke, chemical fumes and gases, and radiation; hazardous work conditions, with excessive risk of injury; and other health risks, such as occupational stress. The work environment may put the worker's health directly at risk for respiratory and lung diseases, such as asthma, chronic bronchitis, and lung cancer. Work in occupations that involve repetitive, boring tasks, low rewards with low prestige, and minimal control over one's work life is associated with higher rates of musculoskeletal conditions, cardiovascular diseases, and emotional disorders. Persons in manual occupations report poorer health than persons in white-collar work, even allowing for education and income (Case and Deaton 2003).

Individuals are constantly exposed to various degrees of radiation – in the home, public places, the workplace, and the health center (from medical diagnostic and therapeutic equipment). The radiation absorbed by the body is negligible in most cases (e.g., dental x-ray), but in other cases it is substantial (e.g., chest x-ray, whole body CT scan, and overseas air flight). It is cumulative over an individual's lifetime, and workers who are constantly or frequently exposed to radiation and who fail to take adequate protective measures live with an excessive risk of incurring various types of cancer.

Food and Water Quality, and Exposure to Infectious Diseases

Impure water supplies and sanitation not only are responsible for illness and death among children, but they also stunt the growth of those who survive. Pesticides used in food cultivation may be absorbed by the growing crops, and then have toxic effects on the animals and humans who eat them. The infectious diseases that ravished childhood early in the last century have been largely eradicated, but some old ones and other new ones, such as HIV/AIDS, are creating increased risks for the population. Hospital stays contribute their own health risks because of nosocomial (i.e., hospital-caused) infections as well as common systems errors, including particularly errors in drug treatment and slow response to in-house heart attacks.

Housing Quality

It is well known that the home can be the source of many types of health risks, especially as a common milieu of accidents. These risks may emanate from the physical structure itself, the characteristics of the occupants in relation to the structure, or the surrounding environment. The physical structure may be a toxic environment for the occupants if it contains lead pipes, asbestos in the building materials, or radon in its subterranean areas. The U.S. Annual Housing Survey inquires about signs of rats

or mice, holes in floors, open cracks or holes, exposed wiring, and water leakage. Homes occupied by elderly persons may not be physically adapted for optimum safety to the limitations of its occupants. The lighting may be inadequate, protective safety devices, such as grab bars, may be missing, or the stairs may impose a special risk of falls. The environment in which the house is located may include special health risks such as excessive street or other noise, neighborhood crime and inadequate police protection, and trash, litter, and junk on the streets.

A summary measure useful for relating housing conditions to health is the crowding index. The crowding index is the number of residents in an occupied housing unit divided by the number of rooms in the housing unit. An index of 1.01 persons or more per room is usually considered an indication of crowded housing conditions. The observations may be grouped as less than 0.5 person per room, 0.5–0.99 person per room, 1.0–1.5 persons per room, and 1.51 persons or more per room. The use of this measure is illustrated in a report based on the National Health and Nutrition Examination Survey showing a gradation in the crowding index for persons having one or more of several infectious diseases, including hepatitis A, B, and C, and herpes simplex virus (U.S. NCHS/Ogden et al. 2005).

Quality of Care, Access to Care, and Utilization of Care

Individuals' health risks differ because of differences in their opportunity to access health care and in the quality of their health care. The risks arising from the variations in the quality of health care vary depending on the appropriateness, safety, timeliness, and equity of the health care (U.S. AHRQ 2005). Equitable care is care that does not vary in quality because of the personal characteristics of the patient, such as age, gender, race/ethnicity, and socioeconomic status. To minimize health risks, an individual should be able to gain access to the health care system easily, as suggested by the following indicators: Availability of health insurance and of an ongoing source of care; convenient transportation; no difficulties in scheduling appointments; little problem in getting referrals to specialists, in securing necessary health screening tests, childhood immunizations, and flu vaccinations; and ready access to prenatal care.

For some, the lack of communication and transportation facilities complicates the problem of securing health care when needed. Many persons lack telephone or public transportation services. About 5% of the households in the United States lack public transportation and the same percentage lack telephone service.

The United States is unique among the industrialized countries in lacking a system of universal health insurance. As a result, a massive number of persons do not have health insurance (about 47 million in 2007), and many more millions who have health insurance have coverage that is severely limited or is tenuous because of the conditions of their employment. Moreover, the U.S health care "system" fails to offer meaningful mental health insurance, dental insurance, and eye-care and hearing-care insurance. Many persons cannot afford to take advantage of the available health services because of the high costs of care. Physicians, particularly

specialists, are scarce in many areas and some refuse to participate in government-sponsored health insurance programs, such as Medicaid and Medicare, because of the inadequate remuneration for their services. The problem of access to health care is discussed in more detail in Chap. 15.

Lifestyle and Behavior

A variety of lifestyle practices and behaviors are associated with specific diseases and are considered to be risk factors for these diseases. The list of behavioral risk factors includes some conditions that could be considered symptoms of disease or diseases, and not merely risk factors. Thus, a clear dividing line between risk factors for disease, symptoms of disease, and disease cannot always be drawn. Consider a few examples. Obesity is a risk factor for several serious diseases, but it is sometimes considered a disease in itself, especially in the extreme form labeled morbid obesity. Hypertension is always considered a disease, even if it is asymptomatic, but it is also considered a risk factor for disease because of its close linkage with, and role as precursor to, other cardiovascular diseases. Persistent pain is a symptom of many diseases. Often, as in the case of arthritis, pain is a symptom of the disease and it is treated as a symptom. Pain may be viewed as a disease in itself, however, if it is unexplained pain or there is no easy modality for treating it.

Accordingly, in this section I commingle lifestyle practices and behaviors that are risk factors for disease with some health conditions that are symptoms of disease and some that are diseases in their own right. Among the lifestyle practices and behaviors examined for their potential health impact are tobacco use, alcohol consumption, sleep, exercise/inactivity, nutrition and diet, immunization history, and social support and stress. Among the symptoms or indicators of disease considered are overweight/obesity, hypertension, high serum cholesterol levels, C-reactive protein and homocysteine levels, premature birth and low birth weight, and sexually transmitted diseases. The reader will recognize some of these items as biomarkers discussed in Chap. 5.

The data collection systems that provide much of the basic data on behavioral risk factors discussed below are the Behavioral Risk Factor Surveillance System (BRFSS), sponsored by the Centers for Disease Control and Prevention (CDC), and the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES), sponsored by the National Center for Health Statistics (NCHS). Data on five important health-related behaviors (alcohol use, cigarette smoking, leisure-time physical activity, body-weight status, and sleep), based on the National Health Interview Survey, are given in the NCHS reports, *Health Behavior of Adults: United States, 2002–2004* and *Health, United States, 2006*. Separate reports on some of these risk factors have also been issued. Inasmuch as health behaviors are self-reported in the National Health Interview Survey, it is reasonable to expect underreporting of the unhealthy behaviors.

Tobacco Use

Smoking among U.S. adults 18 years and over fell sharply in the last third of the last century, from 42% in 1965 to 21% in 2004 (U.S. NCHS/Adams and Schoenborn 2006c; NCHS 2006d). On the other hand, there was an increase in youths' smoking through much of the 1990s that has been essentially maintained over the early years of this century. In addition to those who were current cigarette smokers (i.e., smoked every day or some days), 22% of U.S. adults were former smokers. Hence, well over half (57%) of U.S. adults had never smoked.

The shares of both men and women who smoke are still high. The age-adjusted percentages of men and women 18 years and over who smoked in 2004 were 23 and 19 (U.S. NCHS 2006d). The gender gap was much smaller than a generation earlier but is still sizeable. In 1965 51% of men and 34% of women 18 years and over smoked. Clearly a larger decline occurred among men than women. This is reflected indirectly in the slower rates of decline in death rates from lung cancer among women than men.

Cigarette smoking is the single most important preventable cause of premature mortality in the United States. The CDC estimates that about 400,000 deaths in the United States each year are smoking-related; that is, smoking "causes" one in five deaths each year. The validity of this estimate depends on the selection of causes of death imputed to smoking. Some analysts view it as an overestimate. Smoking substantially raises one's risk of developing cancer, various cardiovascular diseases, and various respiratory diseases, and so it has the potential to reduce life expectancy by several years. It is responsible for almost a third of all deaths from cancer. Smoking during pregnancy increases the risk of complications of pregnancy, fetal losses, premature births, low-birth-weight infants, and infant mortality. It is teratogenic; that is, it tends to produce defects in the fetus *in utero*. Young infants exposed to second-hand smoke are at increased risk of Sudden Infant Death Syndrome (SIDS), asthma, and other respiratory conditions.

Illustrative calculations show that smoking takes a tremendous toll in years of expected life. Rogers et al. (2005) calculated life expectancies at age 20 for persons with different smoking histories and different social and lifestyle histories, using data in the NHIS-MCD file for 1990–1997 (Table 6.8). The figures they obtained cover a tremendous range. They indicate essentially that smoking reduces expected life sharply for both men and women, especially for current smokers who have certain other unfavorable characteristics. Current smokers whose lives are characterized by unmarried status, low income, less than some college education, stressful lives, relative physical inactivity, and overweight have only about half the life expectancy of the average U.S. male (54 years) and the average U.S. female (60 years) at age 20. The figures for persons with these characteristics are only 27 and 34.[10]

[10]Stress, having a parent with only a high-school, or more limited, education, growing up in a one-parent household, drinking alcohol, and having a poor academic record increase the chances of becoming a smoker. These are some of the characteristics that contribute to a relatively low life expectancy of smokers as well as nonsmokers.

Table 6.8 Life expectation at age 20 for men and women according to smoking status and other social and behavioral characteristics: 1990–1997

	Average		Best case[a]		Worst case[b]	
Smoking status	Male	Female	Male	Female	Male	Female
Never smoked	63.4	72.4	75.2	92.4[c]	37.1	48.1
Current smoker	51.5	53.3	59.8	62.8	27.2	34.0
Former smoker	62.4	68.1	73.8	84.8[d]	36.4	45.5
<1 pack a day						

Source: Rogers et al. (2005). Copyright © John Wiley & Sons. Reprinted with permission
[a]Married, income of $50,000+, some college education, drinks occasionally, uses seatbelt, does not experience a lot of stress, more active than peers, normal weight
[b]Not married, income below $50,000, no college education, drinks excessively, does not use seatbelts, experiences a lot of stress, less active than peers, weight too low or too high
[c]Figure implies a life expectancy at birth of about 111 years
[d]Figure implies life expectancy at birth of about 104 years

Consumption of Alcohol

About 55% of U.S. adults were current drinkers in 2004; that is, they had at least 12 drinks in a lifetime and at least one drink in the last year (U.S. NCHS 2006d). One in six adults were infrequent drinkers (i.e., current drinkers who had fewer than 12 drinks in the past year) and about two in five were regular drinkers (i.e., current drinkers who had 12 or more drinks in the past year). Of current drinkers seven percent were "heavier" drinkers (i.e., more than seven drinks per week for women; more than 14 drinks per week for men). Since 1997 the share of heavier drinkers among current drinkers has fallen only slightly (from 7.9%). About one in three adults were lifetime abstainers.

A more favorable picture is presented by the results of NHANES for 1999–2004 (U.S. NCHS/Wright et al. 2009). It was reported that 84% of U.S. adults used alcohol moderately or not at all, on the basis of the definition of moderation as 1 drink or less per day for women and 2 drinks or less per day for men. The NHIS figure for 2004 was 76%.

In general, the abuse of alcohol and of a variety of other addictive drugs is considered inimical to health, but research findings differ as to the amount of drugs that would have to be ingested to be harmful. Alcohol consumption in small amounts each day has been found to be helpful in retarding cardiovascular diseases. On the other hand, no one is advised to begin drinking alcohol for this purpose and consumption of several drinks a day is considered unhealthful. Where, as in Russia, excessive alcohol consumption is widespread, it has been found to be a major factor in the recent rise in death rates and the decline in life expectancy (Nicholson et al. 2004).

Sleep

Humans are "hard-wired" to require about 7–8 h of sleep each night. To secure this amount of sleep, they need to use about the same clock-hours each night. Millions of Americans suffer from sleep problems, sleeping too few hours or irregularly because of overcrowding of the daily schedule, irregular work shifts, a sleep abnormality, genetic influences, a bad lifestyle, and especially, mental illness.

About 63% of the U.S. adult population usually sleeps 7–8 h according to the National Health Interview Survey (U.S. NCHS 2006c). The survey also shows that 29% of the U.S. adult population and 24% of the population 65 years and over usually sleeps 6 h or less each night. There may be a genetic basis for an individual's sleep requirements. Some persons have a gene mutation that reduces their sleep needs to less than 6 h per night and they recover more quickly from sleep deprivation.

On the basis of such data, the AARP Foundation and ILC-USA (2003) reported that sleeplessness regularly affects about one-third of older Americans. Most persons aged 65 years and over report at least one chronic sleep problem; insomnia occurs frequently after age 70. On the basis of the Alameda County study it was concluded that having less than about $7^1/_2$ h of sleep per night has an adverse effect on longevity. Cappucchio (2007) maintains that consistently sleeping about 7 h per night is optimal for good health and a sustained failure to get this amount of sleep may predispose to ill health. Other analysts have also reported that lack of adequate sleep is associated with poor health and susceptibility to disease, particularly in older men and women (AARP Foundation and ILC-USA 2003; Miles and Dement 1980).

Lack of sufficient sleep may cause an increased risk of falling, disturbances in hormone production and metabolism, compromised immunity, and destructive changes in the nervous and cardiovascular systems. It may also result in cognitive impairment, high blood pressure, reduced sensory acuity, shortness of attention span, impairment of attention, memory, and comprehension, an increase in reaction time, a depressed mood, and, in extreme cases, symptoms of psychosis. Too little sleep is also associated with weight gain and damage to the body's ability to regulate blood-sugar levels (increasing the risk for type 2 diabetes).

Experiments with rats indicate that mild, chronic sleep restriction may have long-term destructive effects on neurogenesis (i.e, ability to generate new brain cells) and neural function (Hairston et al. 2005). Hairston et al. maintain that the brain needs sleep more than any other part of the body and lack of sleep undermines the rejuvenating effect of new learning on the brain. On the other hand, too much sleep is associated with depression and cancer-related fatigue.

As with many of the risk factors discussed, a formal measure of adequacy has been developed. The Pittsburgh Sleep Quality Index has been devised for assessing the quantity and quality of sleep among older adults. It consists of 18 questions relating to the seven areas in which sleep problems occur.

Exercise/Inactivity

Every field study seeking to identify the factors contributing to health and longevity concludes with a recommendation for regular exercise suitable to the health condition of the person (e.g., Alameda County, Framingham Heart Study, Baltimore Longitudinal Study). Regular physical activity has been shown to have many positive effects, such as reducing falls, strengthening bones, increasing muscle mass, improving balance, aiding in sleep, enhancing musculoskeletal functioning, reducing stress and anxiety, boosting immunity, and reducing excessive weight and obesity. In addition, regular physical activity improves cardiac output, retards atherosclerosis, lowers blood pressure, and reduces the risk of heart disease and several types of cancers (especially breast and colon cancer), diabetes, and depression. Exercise contributes to one's recovery from cardiovascular illness or cancer. It may even improve cognitive functioning and lower the risk of acquiring Alzheimer's disease. This list suggests that the benefits of exercise may go far beyond its strength- and conditioning effects on health. Exercise seems like the universal health potion.

Exercise is the only intervention for humans for which there is some evidence that disability is prevented or the onset of disability is delayed. It has a positive effect on all the modulatory systems of the body. Exercise increases the level of hormones such as testosterone and DHEA, substantially reduces the levels of inflammation shown by biomarkers, stimulates the production of free radical scavengers, and in some instances improves the function of the autonomic nervous system.[11] Of course, an alternative explanation of the role of exercise is that people who exercise may simply be less ill.

In spite of the pervasive public admonitions about the need for regular exercise to maintain and enhance health, less than one out of three adult Americans engaged in regular leisure-time physical activity in 2004 (U.S. NCHS 2006d). NHANES for 1999–2002 reported that 36% of the civilian noninstitutional population engaged in moderate or vigorous leisure-time physical activity for 150 min or more per week. In addition, one out of three adults engaged in some, but not regular, leisure-time physical activity. The shares of men and women who are regularly active were within a few percentage points of one another in 2004, but as recently as 1998 there was a substantial gap in favor of men. There is a strong inverse relation between age and the percentage of persons who engage in physical activity; for example, 67% of persons 18–24 years old exercise regularly or some of the time and only 38% of persons 75 years and over do so. Older women engage in regular physical exercise much less than older men.

[11]The biological mechanism(s) explaining the beneficial effects of exercise are not completely known. In the case of some types of cancers (colorectal cancer and breast cancer), exercise may help lower levels of insulin, insulin-like growth factor (IGF), and estrogen, which are associated with tumor growth. On the other hand, exercise may help only these cancers and not others because of their age and growth patterns.

While aerobic exercise is the key element in any exercise program, a complete program has four components: aerobic (cardiovascular) activity, weight (strength) exercises, stretching (flexibility) routines, and balance exercises. These varied activities are intended to counteract the tendencies of muscles, including the heart muscle, to shorten, become less flexible, and weaken with age, of bones to lose calcium, become thinner, and weaken, of the body to put on fat, and of the balance skill to deteriorate.

Nutrition and Diet

Humans need three basic categories of macronutrients, namely carbohydrates, fats, and proteins. A healthful diet calls for a mix of foods from these three categories, although in different shares. It also calls for a variety of micronutrients, that is, vitamins and minerals. The general guidelines for a healthful diet requires eating fruits, vegetables, nuts, whole grains, and low-fat dairy products, and avoiding saturated fats, trans fatty acids (partially hydrogenated oils), and processed sugars.[12] Some types of fat, specifically those from most plants and fish, are healthful foods. Poor diets are associated with a number of unhealthful outcomes, such as hypertension, overweight/obesity, bone-thinning, and high cholesterol, and ultimately with diabetes, atherosclerosis, heart disease, stroke, and some forms of cancer.

NHANES (U.S. NCHS/Wright et al. 2009) reported how well the U.S. adult population was following recommendations on several nutritional standards to reduce CVD risk in 1999–2002. These recommendations are the least likely to be followed of the lifestyle recommendations designed to reduce CVD risk. They relate to consumption of fruit, vegetables, salt, and saturated fat. On the recommendation to eat 2–4 servings of fruit per day, only 16% complied, and on the recommendation to eat 3–5 servings of vegetables per day, only 29% complied. The advice to consume 2,400 mg. or less of sodium per day was followed by only 31%, and 42% followed the recommendation to consume less than 10% of total calories from saturated fat per day. Far greater percentages complied with recommendations on smoking, alcohol use, and checking of blood cholesterol and blood pressure, with physical activity being intermediate.

Some research suggests further that certain foods and vitamin supplements may reduce the risk of Alzheimer's disease. These include fish, vitamins E and C, folates, and vitamins B-6 and B-12. The merit of some of these vitamin supplements is

[12]In 2002 the U.S. Department of Agriculture (USDA) revised its nutritional recommendations to include new "dietary reference intakes" (DRI) for proteins, carbohydrates, fats, and fiber. Its guidelines advise the eating of fruits, vegetables, whole grains, and unsaturated fats, and engaging in daily exercise. The World Health Organization endorsed these recommendations in 2003, adding the admonition to reduce salt consumption. In 2005 the USDA further revised its guidelines to urge limiting caloric intake and engaging in regular exercise, and modified its recommended food pyramid to put new emphasis on whole grains and healthful fats. Its customized food pyramid offers 12 different eating plans depending on age and level of activity.

that they are antioxidants and may therefore offset the deleterious effect of oxygen free radicals. Vitamin supplements should be taken sparingly, however; and they should not be taken as substitutes for a healthful diet. Healthy persons can obtain all their vitamins and minerals from their diet, with few exceptions, such as calcium supplements for those who need to improve their bone structure and folic acid supplements for women of childbearing age.

Malnutrition, undernutrition, overnutrition, and food insufficiency, refer to other nutritional risk factors and problems predisposing to disease. Malnutrition encompasses the other nutritional problems listed. Elderly persons are especially vulnerable to malnutrition for a variety of reasons, among them poor oral health, low income, poor appetite related to illness and medications, and insufficient knowledge of nutrition. There is no standardized test for assessing malnutrition among members of a population although a Mini Nutritional Assessment has been designed to identify older adults who have malnutrition or are at risk of developing malnutrition (DiMaria-Ghalili and Guenter 2008). It consists of 18 questions, including six questions used to screen older persons for their nutritional status and a dozen additional questions used to obtain more comprehensive information. The 18 questions include three anthropometric tests – body mass index, arm circumference, and calf circumference.

Inasmuch as the varieties of malnutrition are particularly common in the less developed areas, they are discussed further in Chap. 11.

Other Risk Factors

Many other risk factors may be grouped under the heading, lifestyle and behavioral factors. Like the other individual factors I have discussed, these are partly, sometimes even largely, under the control of the individual although there are always environmental, genetic, and stochastic influences involved. Overweight and obesity, social support and stress, hypertension and similar biomarkers, premature birth and low birth-weight, sexual practices, and immunization history are discussed in the following paragraphs.

Overweight/Obesity. In the present context obesity is treated as a risk factor predisposing to disease although some analysts treat obesity as a disease as well as a risk factor. Excessive body weight, whether in the extreme form labeled obesity or in the more moderate form labeled overweight, contributes to the risk of sickness and death. Obesity is associated with an increased risk for diabetes, gall bladder disease, some types of cancers (e.g., colon cancer, kidney cancer, postmenopausal breast cancer), and a range of cardiovascular conditions, including hypertension, coronary artery disease, congestive heart failure, and stroke. Obesity is also a contributor to osteoarthritis, chronic fatigue, low back pain, edema (i.e., swelling in feet and ankles), insomnia, indigestion, impotence, and various conditions of the hips and knees. The Framingham Heart Study found a strong link between overweight/obesity and heart failure. This risk increases steadily with increasing body weight; it is 104 percent higher for obese persons than for nonobese persons.

In short, obesity greatly increases the risk of premature death, either directly or indirectly, from a range of causes, and the more severe the obesity the greater the risk. In the United States a few hundred thousand deaths a year can be attributed to causes for which obesity is a major risk factor. Olshansky et al. (2005) have estimated that the elimination of obesity would add about 3 years to life expectancy at birth, that is, about the same number of years as the elimination of all cancers.

Anthropometric measures: Body mass index, waist-hip ratio, waist circumference. The most common measure of obesity is the body mass index (BMI). BMI is defined as the quotient of body weight in kilograms divided by height in meters squared (wt_{kg}/ht_{m}^{2}). Persons with a body mass index (BMI) of 25 or more are classified as overweight. The increase in the percent of overweight persons in recent decades in the United States has been spectacular. Between 1960–1962 and 2001–2004 the percentage of adults 20–74 years of age with a BMI of 25 or more rose from 45% to 66% (Fig. 6.8 and Table 6.9). About 74% of persons between the ages of 55 and 74 years were overweight in the more recent period. This figure is 17% points greater than the corresponding figure only a quarter century earlier (57% in 1976–1980). The percent of overweight children aged 6–11 (defined in the note in Fig. 6.8) nearly quadrupled between 1963–1965 and 1999–2002 (from 4.2% to 16%), and the percent of overweight adolescents aged 12–17 more than tripled between 1966–1970 and 1999–2002 (from 4.6% to 16%).

Obesity is defined as having a BMI equal to or greater than 30 kg/m2. The National Heart, Lung, and Blood Institute (NHLBI) has identified three classes of obesity. These correspond to three levels of risk: Class I = 30 $\leq$ BMI < 35; Class II = 35 $\leq$ BMI < 40; and Class III = BMI $\geq$ 40. Obesity increases the risk for several diseases over an overweight condition, as noted above, and the risks increase further from Class I obesity to Class II and Class III obesity. Obesity is the most common nutritional "disorder" in Western countries. As of 2001–2004, 32% of the population 20–74 was obese (U.S. NCHS 2006d). Adult women are more likely than adult men to be obese (34% vs. 30%). A substantial percentage, between 11% and 14%, of children are classified as obese (Troiano and Flegal 1998). Older adults tend to be obese more commonly than younger adults; the percentage at ages 55–64 (38%) is 12% points greater than the percentage at ages 20–34 (26%). Then, apparently because of selectively greater survival of nonobese persons to the older ages, the percentage of obese persons falls to 22% at ages 75 and over.

The prevalence of obesity has risen sharply in recent years. Between 1960–1962 and 2001–2004 the percentage of adults 20–74 years of age who were obese in the United States increased from 13% to 32%, with most of the increase occurring after 1976–1980. The high prevalence of obesity and its rise are not limited to the United States; it is observed in all the more developed countries. In addition, the incidence of obesity is rising rapidly in the less developed countries as they take on the ways of living of the more developed countries.

After age 40, without regular activity, muscle mass is steadily lost and the lost muscle mass is often replaced by fat. The health threat lies with abdominal fat, the fat that collects around the organs in the abdomen, rather than subcutaneous fat, the

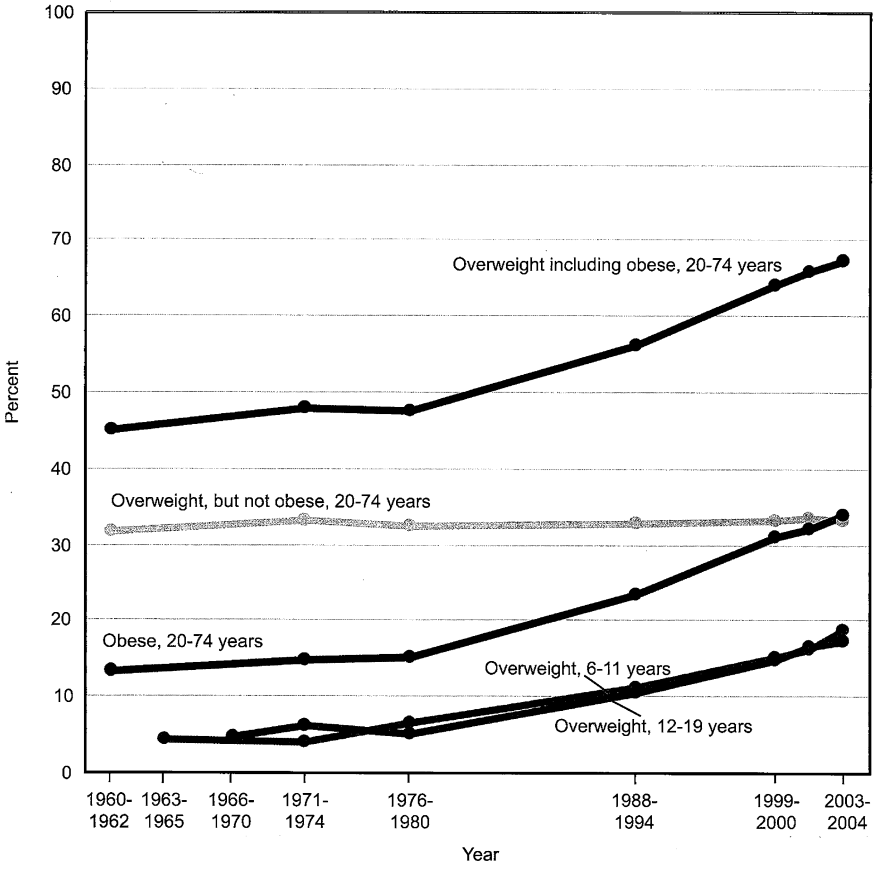


Fig. 6.8 Trend in percent overweight and percent obese for adults and in percent overweight for children: United States, 1960–2004 (Note: Percents for adults are age-adjusted. For adults, overweight including obese, BMI = />25; overweight but not obese, 25 = /< BMI < 30; obese, BMI = />30. For children, overweight = BMI at or above the sex- and age-specific 95 percentile BMI cut points from the *2000 CDC Growth Charts, United States*. Obese is not defined for children. Points are plotted at the middle years of the periods indicated for the surveys; Source: U.S. National Center for Health Statistics (2006d), Fig. 13, p. 39. Primary source: National Health Examination Survey and National Health and Nutrition Examination Survey)

fact that lies just below the skin of the abdomen. Excess weight around the belly (i.e., abdominal obesity) is common among obese adults. Obesity with visceral, or abdominal, fat, is a particularly risky type of overweight condition because it is a high risk factor for a wide range of cardiovascular diseases and diabetes. The body mass index is not a good measure for this condition since the same BMI may represent very different amounts of abdominal fat and muscle may be distributed very differently in different individuals. Abdominal obesity is more satisfactorily and more simply measured by the waist-hip ratio, and it is indicated by a larger waist

Table 6.9 Mean body mass index of the U.S. population, by age and sex: 1960–1962 and 1999–2002

Age (years)	Male		Female	
	1971–1974	1999–2002	1971–1974	1999–2002
Under 20				
6–11[a]	16.6	18.2	16.7	18.3
12–17[a]	20.4	22.5	21.1	22.8
18–19[a]	23.5	24.5	22.0	25.0
	1960–62	1999–2002	1960–1962	1999–2002
20 and over	NA	27.8	NA	28.1
20–74	25.1	27.9	24.9	28.2
20–39[b]	24.8	27.0	23.2	27.4
40–59[b]	25.6	28.6	25.8	28.9
60–74	24.9	28.6	27.2	29.2
75 and over	NA	26.8	NA	26.8

Source: U.S. NCHS (2004). Primary source: National Health Examination Survey and National Health and Nutrition Examination Survey

Note: Overweight but not obese, 25<BMI<30; Obese, BMI ≥ 30; healthy weight, 18.5< BMI <25

Body mass index is calculated as weight in kilograms divided by height in meters squared

NA not available

[a]Average of single ages

[b]Average of two 10-year groups

circumference than hip circumference. Waist circumference is intended to reflect the extent of abdominal fat, the "bad" fat, and hip circumference is supposed to reflect the extent of subcutaneous fat, the "good" fat. The standard value for the waist-hip ratio is 0.82, and the point at which the risks for health conditions begin to "take off" are 0.90 for men and 0.85 for women.

An even more direct and simple measure of the extent of abdominal fat is the waist circumference, a measure that is concerned only with abdominal fat since almost all the girth of the waist derives from abdominal fat. Determining waist circumference requires only a measuring tape. Like BMI, waist circumference is a good biomarker for heart disease, but it is a better predictor of diabetes than BMI. It has not become part of the regular medical diagnostic examination, however, perhaps because it does not seem "clinical" enough or because agreement has not been reached as to where the waist measurement should be taken. The principal recommendation is to take the measurement at navel level and to consider the cutoff point for health problems 40 inches for men and 35 inches for women. Recent research suggests, however, using even lower cutoff points, perhaps as low as 35 for men and 33 for women, and using waist circumference and BMI jointly in crossclassified scales. Some researchers prefer to employ as a proxy measure of abdominal fat the sagittal abdominal diameter, the length of a straight line through the body from slightly below the navel to the lower back, but this is not as simple to measure as waist circumference.

Fitness or low weight – which is better? The evidence shows that moderate physical activity as well as intense exercise can achieve weight loss. Because the efforts of many overweight women to reduce their weight by exercise meet with limited success, it is useful to ask, it is better to be fit than thin? The evidence shows that active women, thin or fat, are much less likely to have a heart attack or other cardiac problems than women who do not exercise. Nevertheless, weight does matter. Overweight women are still much more likely to develop diabetes than women who are not overweight, with or without exercise.

Being underweight can also be a health risk. It should be a matter of concern that persons of borderline "healthy" weight have an adequate reserve for the weight loss that accompanies many infectious and other illnesses and major surgery. A low BMI associated with unintentional weight loss is a greater mortality risk to older adults than is obesity with or without intentional weight loss (Locher et al. 2007). The curves of the BMI or the waist-hip ratio in relation to mortality are U-shaped; this reflects the fact that extremely low or extremely high values are "dangerous for one's health."

Causes of obesity. Weight gain occurs when the number of calories gained by eating (regardless of the type of food) exceeds the number of calories lost by exercise (of whatever type).[13] Hence, weight reduction can occur only by reducing the intake of calories from food or increasing physical activity. Our proclivity to eat goes beyond our desire to satisfy our hunger. Eating is a source of pleasure, a way to deal with stress, a "duty" at social events, and a convenient preoccupation to avoid other more serious concerns, given that food is so readily available. For some, eating becomes an addiction. This occurs when the need to eat more than is required for sustenance becomes compulsive behavior, is subject to decreasing rational control, and becomes associated with changes in brain structure. In general, compulsive eating and an addiction to overeating are precipitated by easy access to food and poor lifestyle choices. People who act compulsively do not think about the act; the addiction is converted into a powerful automatic response. Once excess weight has been put on, it becomes extremely difficult to discard it even if a healthy life style and a diet plan are pursued in an effort to eliminate the extra weight.

Genetics and environmental/lifestyle factors both contribute to the onset of obesity, but their relative contributions are difficult to specify. Several genes influence a person's weight but these genes have not been identified. Obesity tends to run in families but, inasmuch as the family is both a gene-transmitter and an environment for its members, the family connection complicates any attempt to determine the relative role of genetics and environment in the expression of obesity. Adult obesity, common in the postmenopausal years, is associated with obesity in childhood (Ferraro et al. 2003) and lower socioeconomic status, especially

[13]This statement is an oversimplication. There are circumstances where the volume of food eaten is effectively reduced by changes in the absorption rate, as when the immune system is compromised or a person undergoes a major surgical procedure or experiences other bodily trauma.

among women. The degree of obesity is exacerbated by a high-fat diet, alcohol consumption, a variety of medications, and physical inactivity, especially among older persons.

From the view of evolutionary biology, a substantial part of obesity is genetically caused and possibly triggered by hormones. Our body resists our efforts to follow a healthful lifestyle because it is "hard-wired" by evolutionary pressures to put on weight, keep its weight up, resist starvation, and defend itself from premature death (i.e., death prior to reproductive success) resulting from lack of food. Moreover, there is the biological tendency to store fat in the midregion of the body pending a food shortage and obviating the threat of starvation.

Social support and stress. Social support is positively associated (and social isolation is negatively associated) with physical and mental health and can reduce the stress caused by many adverse life events (Seeman et al. 2002). According to a study of the MacArthur Foundation for the Study of Successful Aging, a very strong predictor of well-being among adults is the frequency of visits with friends and the frequency of attendance at meetings of clubs and organizations (Rowe and Kahn 1998). One's social network can be supportive and protective, and aid in coping with serious illness, death, and other adverse life events.

Social support mollifies the effects of these events by reducing stress. Much research has been done to identify and rank the events that create stress. These events include death of a spouse, death of a child, chronic illness of a child, divorce, loss of a job, a serious accident, the death of a close friend, a major lawsuit against the person, and destruction of one's house by fire, flood, or hurricane. Other lifetime traumas are an unwanted pregnancy, rape, failing a school grade, divorce of parents, family violence, and witnessing a death. Death of a spouse is one of the most traumatic events; it is an event that removes from the survivor a principal source of social support. Note that some stressful events, such as birth of a child, marriage, and successfully passing a PhD preliminary examination, are potentially positive experiences and may contribute to "eustress" ("good stress") rather than "distress" ("bad stress").

A person will experience many major and minor stressful events during the life course. The effect they have on her or his physical and mental well-being will be determined by how she or he deals with them. The effect of stress on health and mortality also depends on the stage of life at which the stress occurs. Stress experienced by some persons early in life may result in their premature illness or death, even before they reach middle or old age, or in long periods of chronic disease and disability prior to death in later life; or it may endow them with a greater resilience or motivation to handle future stressful events. The perception of stress by individuals has an effect similar to the stress itself. How persons view their lives, cope with the demands made on them, and secure and use social support are important. To overcome stress, people at risk for a high level of stress can engage in exercise, Yoga, meditation, prayer, breathing exercises and other relaxation techniques, or other forms of personal stress reduction, as well as seek external social support.

Stress as a pathway to illness. We need a biological explanation of how negative lifestyle risk factors or lack of social support can increase the risk of disease, or how the availability of social support or absence of negative lifestyle risk factors can reduce the risks of disease. Although some lifestyle factors are directly responsible for disease (e.g., smoking, radiation exposure), diseases generally do not result directly from such factors. We need to find a mechanism or mechanisms by which the body is biologically altered and set on the path to pathological change. Stress may serve as that pathway. Stress has been linked to heart disease, stroke, cancer, chronic lower respiratory disease, and depression, anxiety, and ahedonia (i.e., an inability to experience pleasure). It is responsible for episodes of asthma, rheumatoid arthritis, and gastrointestinal illnesses. Other consequences of stress are headache, sleep and appetite problems, difficulty in swallowing, ringing in the ears, an urgent need to urinate, fatigue, and attention difficulties.

One theory connecting stress and a disease reaction maintains that chronic stress accelerates the aging process by shortening the life span of cells (Epel et al. 2004). In the Epel et al. study the cells of the immune systems of women under much stress (i.e., parents with chronically ill children) aged 10 years (i.e., lost life span) more than the corresponding cells of women under little stress (i.e., mothers with healthy children). Three factors in the cells of the immune system of research subjects were measured in this study: Telomeres, which cap the ends of chromosomes and become shorter as cells reproduce; telomerase, an enzyme that replenishes telomeres; and oxidative stress, which damages DNA in telomeres. The telomeres in cells of women under stress had a shorter length, telomerase showed lower activity, and the cells experienced greater oxidative stress.[14]

Some researchers suspect that hormones such as cortisol, epinephrine, and norepinephrine, which respond to stress, are responsible for the disease process. Animal and human studies link responses to stressful social relationships to the body's patterns of neuroendocrine regulation (Seeman and McEwen 1996). These patterns are part of the homeostatic regulatory processes of the body, that is, those that maintain its internal equilibrium. Nonsupportive social relationships induce the body's stress reaction, which entails enhanced responses of the hypothalamic-pituitary-adrenal (HPA) axis, the sympathetic nervous system (SNS), and other internal regulatory systems. Chronic neuroendocrine activity leads to such physiological responses as elevations in blood pressure, heart rate, and serum lipids – conditions that add to the risk for a number of chronic pathologies.

Inflammation: A possible unifying theory of major chronic diseases. The concepts relating stress, neuroendocrine regulation, and disease leads to a hypothesis linking inflammation as a physiological factor in the development of the major chronic

[14]Telomeres are DNA-protein complexes that cap the ends of chromosomes and promote genetic stability. When a cell divides, part of the DNA in telomeres is eliminated, and after many cell divisions, with the elimination of much of the telomeric DNA, the aged cell stops dividing. Telomerase is an enzyme that replaces a part of the telomeres during each cell division and protects them. Oxidative stress is a destructive metabolic process that contributes to the shortening of telomeres.

diseases. In the search for a unifying biological mechanism to complete the link between social risk factors, physiological responses to risks, and actual disease, many medical researchers point to inflammation as that mechanism. They hypothesize that inflammation is the common underlying cause of the major chronic diseases of later life – the cardiovascular diseases, diabetes, cancer, and Alzheimer's disease. How does this hypothesis tie these diseases together?

Inflammation is the body's protective reaction to an attack by disease. It is part of the immune response and hence is a beneficial immediate response to illness and injury. It results from the physical action of white blood cells and the chemicals they produce, including antibodies and cytokines. The characteristics of inflammation may be recalled by their rhyming Latin names – *rubor* (redness), *tumor* (swelling), *calor* (heat), and *dolor* (pain). Normally when the process of inflammation has done its job, neuroendocrine-regulation systems turn it off. According to Seeman and McEwen (1996), however, the stress response may persist, with "inflammatory" results. When the risk factors and stress persist, the neuroendrocrine responses and inflammation may continue and the latter become factors themselves in the causation of disease. In this way, Seeman and McEwen hypothesize, excessive activation of the HPA axis and SNS increases the risks for a number of important pathophysiologic processes, including hypertension and cardiovascular disease. As a further example, persons with excessive levels of C-reactive protein, a biomarker for inflammation, are more likely than others to develop insulin resistance, a precursor to diabetes. With this condition, cells reject insulin and do not properly metabolize glucose circulating in the blood.

Blood Pressure, Cholesterol, C-Reactive Protein, and Homocysteine

Hypertension, or high blood pressure, may be considered a disease, a cause of disease, and a risk factor for disease. High blood pressure can be a factor in heart disease, kidney disease, and stroke. According to NHLBI it is a factor in 50% of the heart attacks, two-thirds of the strokes, 90% of the cases of heart failure, and one-quarter of the cases of chronic kidney failure in the United States. High blood pressure compromises circulation and so adds to the risk of cognitive impairment and dementia.

Blood pressure is measured in two phases, diastolic blood pressure (i.e., pressure in the blood vessels between beats, or when the heart relaxes) and systolic blood pressure (i.e., pressure when the heart beats). For diastolic pressure, <80 mmHg is considered healthful and >90 mmHg is considered high. For systolic pressure, <120 mmHg is considered healthful and >140 mmHg is considered high. Blood pressure less than 120/80 mmHg is considered optimal and blood pressure over 140/90 mmHg is considered unhealthful. In 2003 NHLBI issued new guidelines for diagnosing high blood pressure. It set up a new blood pressure category, namely a "prehypertension" range for systolic pressure – 120–140 mmHg. This band covers 22% of American adults (18 years and over), or about 50 million people. A high

systolic pressure in combination with a low diastolic pressure is denominated isolated high blood pressure. This condition was formerly considered benign, but now it is viewed as a risk factor that requires treatment.

Over one-fifth of adult Americans had high blood pressure in 2004 (Table 6.7). The percentages generally move upward with increasing age, especially after age 45. The lifetime risk of developing high blood pressure at age 55 is 90%. Half the population 65–74 years of age has high blood pressure. A peak is reached at ages 75–84, with some decline thereafter. This decline is presumably a result of the selective demise of those having the disease, given that the population is heterogeneous with respect to this risk.

The specific cause of most cases of hypertension is unknown. A small share of the cases is caused by kidney disease, glandular tumors, diabetes, and some drugs (including some medicines for colds and migraine), but most cases are associated with a poor lifestyle, that is, physical inactivity, unhealthy diet, stress, smoking, and ingesting too much salt and alcohol. There is a wide array of medications for treating high blood pressure but, in treating the condition, elimination of the causes and predisposing factors enumerated above is usually considered first.

Research cardiologists differ on the merits of reducing high blood pressure at the ages over 80. Goodwin (2003) concluded from a number of research studies that higher levels of systolic and diastolic blood pressure in persons 80 years old and over are associated with increased survival, He notes that clinical trials of the treatment of hypertension for those 80 years and over result in outcomes favoring placebo over drug therapy and that for persons of this age range, high blood pressure is not associated with premature death. Goodwin cites epidemiological and clinical evidence (e.g., the Framingham study, EPESE) to support his conclusions. He measures success in terms of survival after 5 years. In several studies, both total mortality and cardiovascular mortality were substantially decreased in the treatment group under age 80 but not in the group 80 years and over. These studies also indicate that incidences of heart attack, congestive heart failure, stroke, and even cognitive impairment were reduced as a result of the use of medication even while mortality increased.

Others conclude from the available studies that there is increased survival of persons over age 80 when blood pressure is lowered with medication, or shows an insignificant rise. Aronow (2003) and Newman (2003) present this view, noting also that antihypertensive drug therapy improves the quality of life for persons 80 years and over by reducing adverse health events such as stroke even if it does not significantly lower mortality from all causes. The reason for this specific latter outcome may be that some aged persons in the studies have disproportionately low blood pressure from various health conditions, or from having their blood pressure lowered excessively by the medication, rendering them more vulnerable under conditions of sustained stress. The broader outcome may be a result of demographic selection in that these elderly populations may already have been “cleared” of those with the greatest risks. Thus the issue is whether treatment of hypertension in

persons over 80 is more important for increasing survival or reducing the incidence of several serious health conditions. In sum, there is no consensus about treating hypertension in the very old.[15]

Cholesterol. Cholesterol is a lipoprotein (i.e., a fatty organic compound) found in body structures and some foods. Excess cholesterol builds up in arteries in the form of plaque and adds to the risk of heart attack, heart failure, and stroke. Lipoproteins in the blood, measured by cholesterol tests, can be fractionated into high density cholesterol (HDL) and low density cholesterol (LDL), the former being good for cardiovascular health and the latter being bad for it. Levels of lipoproteins increase slightly with advancing age, particularly in women after menopause.

There has been a decline of a few percentage points in the cholesterol levels of the U.S. population in the last few decades. Cholesterol levels among all adults in the United States decreased from 206 mg/dL during the 1988–1994 period to 203 mg/dL during the 1999–2002 period, mainly as a result of the use of statin drugs (Carroll et al. 2005). The drop was accounted for entirely by older adults (i.e., men 60 years and over, women 50 years and over); levels for younger adults did not change. Current levels are still considered too high for good health; a total cholesterol below 200 has been recommended by the National Heart Lung and Blood Institute. Serious questions have been raised, however, about the merits of lowering cholesterol levels after age 80 using statin medications.

A person's genetic makeup affects the rate at which the body makes, uses, and disposes of fats. The statin class of drugs can be used to lower LDL and raise HDL. HDL levels can also be increased by changes in diet, exercise, and weight, and sometimes by doses of the nutrient niacin. Recent research suggests the desirability of more aggressive use of cholesterol-lowering drugs than has been the case, since both men and women at high risk of a heart attack or a stroke benefit from the drugs even if their cholesterol levels are within the healthy range. The recommendation is that high-risk persons – people who smoke, have diabetes or high blood pressure, or have suffered at least one heart attack – should lower their risk by reducing their level of LDL from the previously recommended 129 mg/dL to100 mg/dL and even 70 mg/dL.

C-reactive protein and homocysteine. Elevated levels of C-reactive protein (CRP) are found in the blood when some disease in the body causes an inflammatory condition. Inflammation in the vascular system is usually indicated by a high level of CRP in the blood, and so the level of CRP/inflammation is now considered an "etiological" factor in heart attacks and strokes. A high level of homocysteine

[15]Hajjar (2003) believes that, until studies specifically designed to detect a survival benefit in the very old are conducted, aged patients with high blood pressure should be treated actively for that condition, barring other strong reasons for avoiding such treatment. This may be a prescription designed to protect physicians from the charge of not practicing in accordance with the state of the art. It will be difficult to secure definitive evidence for the advanced aged, however. As explained in Chap. 17, few very old qualify for research protocols at NIH, mainly because the pool of applicants is quite small and their many comorbidities confound the interpretation of research results.

in the blood is another biomarker for vascular inflammation and hence another major risk factor for heart disease and stroke. High levels of homocysteine may promote the formation of blood clots, injure the lining of the arteries, and increase the risk of atheromas (i.e., deposits of lipid-containing plaques on the walls of an artery) and atherosclerosis (i.e., thickening and reduced elasticity of the walls of an artery, resulting from atheromas). Homocysteine levels increase with advancing age, particularly in women after menopause.

Prenatal Care: Low Birth Weight and Preterm/Premature Birth

Low birth weight is defined as a weight of less than 2,500 g at birth. The corresponding population measure is the birth-weight ratio (per 100), defined as the percentage of births in a given year weighing less than 2,500 g. Low birth weight is a factor in infant mortality as well as poor health in later years. It can be prevented by proper prenatal care. Prematurity is usually associated with low birth weight. A preterm or premature infant is an infant born at less than 37 weeks of uterogestation. The population measure is the percent of births with gestational age under 37 weeks. Like low birth weight, prematurity can be prevented by proper prenatal care. Another measure of the general physical condition of newborn infants is the Apgar score. It rates five characteristics of newborn infants 5 min after birth and sums the scores of the five components, A score of less than 7 out of 10 is considered poor or fair. For the United States in 2004 the birth-weight ratio, percent of premature births, and the percent of newborn infants with Apgar scores under 7 were 8.1%, 12. 5%, and 1.5%, respectively (U.S. NCHS/Martin et al. 2006e)

Sexual Behavior

There is ample evidence that regular sexual companionship is conducive to both physical and mental health. This may explain in part the advantage of the married state over the other marital states with respect to longevity and health. On the other hand, having numerous partners and promiscuity constitute risky sexual behavior, associated especially with the risk of incurring sexually transmitted diseases. It was reported in the *New York Times* of March 12, 2008, that three-quarters of American adolescents are sexually active by the age of 19 and one-quarter of teenage girls has a sexually transmitted disease. Sexually transmitted diseases, including the most serious of them, HIV/AIDS, will be discussed further in Chap. 11 because of their widespread prevalence and pervasive consequences in many less developed countries.

Immunizations

I include immunizations as a risk factor to call attention to the fact that failure to secure proper childhood and other immunizations exposes the person to the risk

of incurring many infectious diseases with serious health consequences, including death. The formulation of the germ theory of disease a century and a half ago led to the development of vaccines as treatment modalities. Many types of immunizations are well established as standard procedure in preventive public health. The DTP vaccine, the vaccine used to protect children against diptheria, tetanus, and pertussis, is one such standard vaccine. New vaccines continue to be developed for old diseases and new ones. For example, a vaccine against Haemophilus influenzae type b (Hib) was developed in 1987. Since then, Hib infections have decreased by more than 99%. The incidence of Hib meningitis, previously a leading cause of acquired mental retardation, has also decreased by more than 99%.

A vaccine against Staphylococcus Aureus, a major cause of death and infection in hospital patients, became available in 2002. S. Aureus causes illnesses ranging from minor skin infections to pneumonia, meningitis, and heart infections, particularly in persons with weakened immune systems. More recently a form of S. Aureus resistant to available antibiotics known as MRSA evolved and is contributing to a resurgence of S. Aureus in hospitals and in the community. A vaccine for typhoid fever is under development; it shows a 91.5% effectiveness rate in clinical trials. Perhaps most important of all are the current efforts to develop a vaccine for HIV/AIDS; clinical trials for such vaccines are now under way.

The Stochastic (Chance) Factor

Health and longevity are greatly influenced by stochastic processes as well as by genetics, lifestyle, and the environment. Chance is a major cause of individual differences in morbidity and mortality since random events occur at all stages of life, from fetal development and birth to the oldest ages (Finch and Kirkwood 2000). Chance events affect the physiological development of the individual and have potentially positive and negative effects on the individual's growth. As a result, chance plays a major role in determining the cause and timing of the events of sickness and death.

Considerable phenotypical variations (i.e., differences in observed traits) have been found among individuals reared in relatively homogenous conditions, partly because of the stochastic factor. Thereby, identical twins reared together show differences in health and longevity. Persons reared with different lifestyles may experience a common negative health outcome depending on the workings of chance. For example, some people die of lung cancer who do not smoke or work or live in smoke-contaminated environments. People suffer from peripheral vascular disease, coronary artery disease, and abdominal aortic aneurysms even though they have lived a healthful lifestyle and have monitored their health regularly. In sum, chance, interacting with a person's lifestyle, environmental experiences, and genetic predispositions, can account for many unexpected adverse outcomes.

On another level of analysis, the causes of death compete with one another to take the life of an individual. Death is a random event and, to a large extent, the

reported cause of death is arbitrary. As noted in Chap. 3, when individuals die, they usually suffer from an underlying cause (which is typically the reported cause), contributing causes (including the immediate cause and other conditions leading to death), and other health conditions. With such an array of health conditions, it is quite likely that, if they had not died from the reported cause, they would have died shortly thereafter from one of the contributing causes or other associated conditions, or a combination of them. According to Hayflick (2001), the assignment of a conventional cause of death is unrealistic and all intrinsic deaths result from the cumulative "molecular disorder" that characterizes the aging process in persons who live long enough to experience it.

Consider the case of the distinguished biogerontologist, Roy Walford, who died at age 79 in 2004 of amyotrophic lateral sclerosis ("Lou Gehrig's disease"). He devoted many years to research into and writing about the requisites for a long life. He religiously pursued a healthy lifestyle, which for him mainly involved exercising regularly and eating a very low-calorie nutritious diet. Such "caloric restriction" is known to contribute to an increase in the life span of several species of subhuman animals. In Walford's case the confluence of genetics and chance were dominant and intervened to take his life prematurely in spite of his efforts to "live right."

Interrelations of Genetic and Nongenetic Factors

We recognize then that the etiology of sickness and death involves complex interrelations of genetics, environment, lifestyle, and chance. All these factors are always involved and they operate interdependently. Global figures for the contribution of these factors to longevity are usually expressed, however, as a dichotomy between genetic and environmental/lifestyle factors; that is, the structuring of the question neglects the role of chance. Even so, experts disagree as to the relative importance of these influences on human disease. Furthermore, this way of structuring the problem can be misleading in that it conceals a considerable degree of variability. The share contributed by each factor varies not only with the cause of illness or death but also the age, sex, and race of the subject. For example, lifestyle risk factors dominate in the teenage years and youth as direct and relatively independent causes of illness and death (i.e., accidents and other deaths of violence). At the later ages genetic and nongenetic causes play more complex interactive roles. Cumulative chance mutations account for changes in the genes, and environmental "hits" result in numerous changes in them as immunosenescence weakens the body's ability to ward off internal and external "enemies."

At the ages under 50, inherited genetic diseases may also be important, although they are uncommon and confined to a select list of conditions. When genetic diseases occur early in life, their expression (usually determined by a single gene) is little influenced by environmental and lifestyle factors. At the older adult ages, however, lifestyle and environmental factors are more likely to be collaborating with genetic forces (usually represented by combinations of genes) in influencing the

cause and timing of illness and death. These are the usual conditions under which the various endogenous/intrinsic illnesses manifest themselves. Through all these developments, chance is always involved and the determination of the time, form, and site of the chronic conditions of later life can be described as an essentially stochastic process.

Environment and behavior can be viewed as dominant over genetics in the causation of many diseases on the ground that they alter gene activity, determining when and how long given genes are active, i.e., when they manufacture new proteins. This can be important because the genes may have multiple subtle abnormalities and, depending on environmental and behavioral influences, may express themselves normally or abnormally. External influences having notable effects on gene expression are stress, exercise, radiation, use of drugs, and diet. For example, neurogenesis, i.e., the creation of new nerve cells, can occur even in adult life as a result of appropriate environmental and behavioral influences.

The presentation of celiac disease illustrates how genetic and environmental forces interact to determine the course of some diseases, even clearly genetic diseases. Celiac disease is a chronic disorder of the gastrointestinal tract characterized by an inability to metabolize gluten and causing chronic inflammation of the intestinal mucosa. Although the disease has its major genetic risk factors (HLA-DQ2 and HLA-DQ8), it requires an environmental trigger, namely, specific peptides present in wheat, rye, and barley, to activate it. Most patients experience complete "remission" from their symptoms after these grains are excluded from their diet.

Earlier in this chapter I cited some estimates of the relative contributions of genetics and environment to the causation of mortality. The National Institute on Aging (Hodes 2005) estimates that two-thirds of the variance in mortality can be accounted for by nongenetic forces and only one-third by genetic forces. Carnes (2005) provides a similar estimate. Finch and Tanzi (1997) also estimate the relative contributions to be about one-third genetic and two-thirds environmental. Vaupel et al. (1998) offered a more nuanced explanation of the role of genetics and environment that reflects the complexity of the relation better:

> A frailty model applied to Danish twin data .... suggests that about 50% of the variation in human life-spans after age 30 can be attributed to survival attributes that are fixed for individuals by the time they are 30; a third to a half of this effect is due to genetic factors and half to two-thirds to nongenetic survival attributes (related to, for example, socioeconomic status or nutritional and disease history). The model suggests that the importance of survival attributes may increase with a person's life expectancy. For persons who at age 30 can expect to live into their 90s, more than 80% of the variation in life-span may be due to factors that are fixed by this age. These calculations were made by I.A. Iachine on the basis of a frailty model described by Yashin and Iachine (1997).

Rowe and Kahn (1998:64–65) maintain that, with advancing age, genes play an increasingly minor role in the promotion of single risk factors, such as hypertension, and even full-blown disease, such as heart disease. Their view is that, with advancing age, the lifestyle factor is of paramount importance in causing these conditions. They maintain further that, inasmuch as lifestyle determines the ultimate impact of genes, lifestyle changes can prevent the presentation of a potential genetic trait. This

view of the role of the lifestyle factor goes beyond any of those previously cited. It is arguably too expansive, since genetics, certain environmental influences, and chance are still primary players in this "dance of power." Even if such a broad role for lifestyle has limited demonstrable validity, however, the championing of it may have considerable practical merit in serving to induce people to be responsible for their own health and try to be proactive in maintaining it. While the dictum that people are entirely responsible for their own health is patently not true, promoting it can have highly positive consequences. Under a future regime of personal medicine, when the genomic profile of each individual can be determined quickly and cheaply and developments in nanomedicine/nanogenetics will provide expanded understanding of the function of individual genes, this view can be reviewed more critically.

Heritability Index

The measure used for the proportion of an observed disease or physical characteristic attributable to genetic influences is called the heritability index. In an informal sense, the heritability index refers to the degree to which a specific trait is passed on from parent to offspring. It is formally defined as the degree to which the total phenotypic variation within a population is due to genetic variation, that is, the ratio of the genetic variance to the phenotypic variance. The phenotypic variance (V_P) equals the sum of the genetic variance (V_G) and the environmental (i.e., nongenetic) variance (V_E):

$$h^2 = V_G \div V_P = V_G \div (V_G + V_E) \tag{6.15}$$

The heritability index is represented by h^2. An h^2 of 1 signifies that all phenotypic variance is due to genetic variance. The complement of the heritability index ($1-h^2$) is a measure of the environmental (i.e., nongenetic) effect. Note that in this construct of heritability no allowance appears for stochastic variance, which is assumed to express itself through the other factors. When the index is considered in terms of variations by age and cause of death, a relatively low heritability index would be expected at the younger ages because of the predominance of exogenous (extrinsic) causes and a higher index would be expected at the later ages where endogenous (intrinsic) diseases prevail. Recall, however, the powerful role of lifestyle and environment in modifying these genetic tendencies in later life.

Individual vs. Categorical Risks

The vulnerability to morbidity and mortality of a person varies according to the various demographic/social/economic/biological characteristics of the group to which a person belongs (i.e., categorical risks) as well as according to the unique genetic endowment and personal experience of the members of the group (i.e., individual risks). In groups, however there is heterogeneity, especially in health. Heterogeneity of risk is pervasive in human experience. Age, gender, and race may

be viewed as biological characteristics for which risks vary, that is, characteristics that add to heterogeneity of risk. In Chaps. 3 and 4 the variation of mortality with age was discussed in considerable detail. Children and the elderly are more likely to succumb to various mortality risks (e.g., accidents, starvation, infectious diseases) than persons in the intermediate ages. Through these age-associated effects, among other ways, biological factors contribute to mortality in varying degrees, even of extrinsic mortality. Adding to the heterogeneity of risk of incurring disease in later life are three aspects of family life – socioeconomic conditions of the family when the individual was a child or teenager, life-course influences, especially health experiences, and current socioeconomic conditions of the family, as described below.

Earlier classifications of deaths have tended to assign whole categories of causes either to the endogenous/intrinsic or exogenous/extrinsic classes. For example, all cancers, cardiovascular diseases, kidney diseases, liver diseases, diabetes, congenital malformations and chromosomal abnormalities, and similar categories of causes were classified as endogenous; and all infectious and parasitic diseases, acute lower respiratory infections, and deaths due to violence (i.e., accidents, homicides, and suicides) were classified as exogenous. Bourgeois-Pichat (1952) divided deaths into these two broad classes composed of these broad categories of causes in his original formulation.

Many "endogenous/intrinsic" diseases result, however, from a complex combination of biological and environmental/lifestyle factors or from environmental/lifestyle factors almost wholly. For example, some diseases in the broad categories labeled endogenous/ intrinsic are initiated by external factors such as pathogenic microbes or self-destructive behavior. Accordingly, a narrower and more refined definition of intrinsic deaths than the one described above was employed by Carnes et al. (2006) for their analysis of mortality partitions in biogerontological research. In their classification, several causes of death included in the intrinsic categories listed above were reassigned to extrinsic mortality, either because the principal etiological factor is an infectious agent (e.g., rheumatic heart disease, cervical cancer, infection of the kidney, chronic hepatitis, urinary tract infection), the abuse of a toxic substance (e.g., lung cancer, cirrhosis, drug dependence), or a food deficiency or excess (e.g., nutritional deficiency).

Summary Note

Trying to measure the exact contribution of genetic and nongenetic forces to human mortality is a speculative task, like trying to determine the relative importance of the length and width of a rectangle when the use of the rectangle is unspecified. Genetic and nongenetic factors are inextricably intertwined. Important conditions such as age, sex, and cause of death have to be considered to make this determination meaningful. Moreover, chance plays a significant role in the development and expression of morbidity and mortality – a fact clearly evident in the case of the extrinsic causes of death involving violence, but also, as indicated, in the case of the

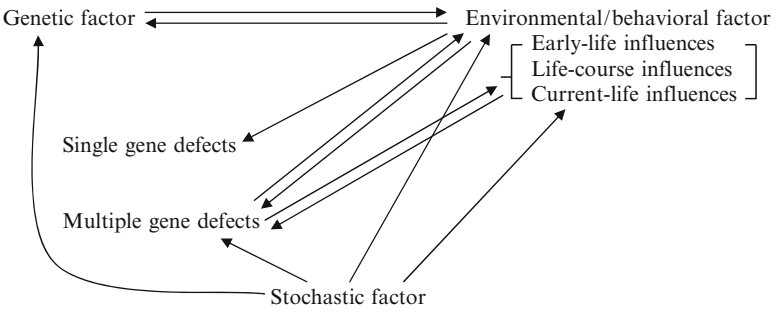


Exhibit 6.4 Schema relating various factors determining health and longevity in later life

endogenous causes of later life (See Exhibit 6.4). As result, the development and the nature, timing, and form of the expression of genetic and nongenetic factors are inevitably subject to stochastic influences.

Some important causes of death have a purely genetic basis that is present as a risk factor at conception (i.e., single-gene hereditary disorders), e.g., Huntington's chorea and sickle-cell anemia. Others involve genetic susceptibility and predisposition, and then triggering by the environment, e.g., Type I diabetes and celiac disease. Still others arise from somatic mutations that accumulate over the life course, e.g., congestive heart disease and prostate cancer, although there may be a genetic basis for these diseases also. Cumulative somatic mutations result from such factors as oxygen free-radical damage, radiation damage, protein cross-linking, reduced levels of heat shock proteins, and a decrease in hormone production (see Chap. 13 and U.S. NIA 2006). In all of this, chance may play the determinative role as to which genes are adversely affected, how they are affected, and the nature of the adverse effect on other genes and their protein production.

Some endogenous diseases have very large genetic contributions; for others the genetic contribution is small. There is always some genetic contribution if only because of the variation in the vulnerability of persons of different ages to different health conditions.

Early Life Influences on Health and Mortality in Later Life

Among the nongenetic factors that play an important role in the causation of chronic illness and death in later life are one's earlier-life experiences and environment, especially the socioeconomic status of one's family and one's health history in the earlier years. I discussed cohort effects on later-life morbidity and mortality to a limited extent in Chap. 5, and I pursue this theme further in this section. Four different interpretations can be given to the effect of early life experiences on later-life health and longevity of the members of a birth cohort. Much of the recent research has focused on the deleterious effects of ill health in childhood on

health in later life. Some of the research emphasizes the rise in socioeconomic status of the members of a cohort and the improvements in public health and medical knowledge as the members of the cohort grow older. Under these circumstances the members of the cohort tend to enjoy improved health opportunities and improved health conditions in later life as compared with their younger years. Some research has focused on the positive effects of having survived illness in childhood, whether by acquiring immunity to various infectious diseases or by becoming physically and emotionally hardened for survival to later years.[16] Finally, early sickness and death tend to weed out the weaker members of a cohort so that the stronger members constitute a relatively larger share of the survivors in later life – an expression of the concept of heterogeneity of frailty in cohorts.

The life-course perspective is associated with the concept of a chain of risk, which posits that adverse experience in early life modifies the life course, usually in an adverse direction. The opposing concept, called the acquired-immunity concept, posits that early exposure to adverse and stressful experiences reduces the risk of a similar adverse experience in later life. The evidence would appear to argue that the chain-of-risk concept is the more tenable general interpretation of later-life health events than the acquired-immunity concept although both are at play among the individuals of a cohort and serve to explain the health-history of members of the cohort in different degrees. The heterogeneity-of-frailty concept has a more general application in characterizing all cohorts, with evident consequences for the health composition of the cohort in late life.

There is accumulating evidence of the importance of early life experiences in affecting health status in later life, including the experience of having lived both with the risk factors for disease and with actual disease (O'Rand and Hamil-Luker, 2005; Costa and Lahey 2005; Blackwell et al. 2001; Preston et al. 1998; Ferraro et al. 2003; Elo and Preston 1992, 1996; Hayward and Gorman 2004). O'Rand and Hamil-Luker found that early disadvantage and childhood illness increase the risk for a heart attack in later life but that adult experiences may mediate the effects of early disadvantage. Costa and Lahey estimated that at least one-fifth of the increase between 1900 and 1999 in the probability of a 65-year-old surviving to age 85 may be attributable to improved early life conditions. They attribute the remainder of the gain to a host of other changes, mostly public health improvements, such as filtration and chlorination of water, construction of integrated sewage systems, widespread vaccination against childhood diseases, improvements in the milk and food supply, and reductions in environmental pollutants and poisons.

Some studies point to the month or season of birth as strongly associated with longevity (e.g., Gavrilov and Gavrilova 2003). Fogel (1993), Steckel (1995), and Komlos (1993) have linked developed height, as a measure of net nutrition and health in infancy and early childhood, with health in later life. Barker (1997) offers

[16]In this view, exposure to a moderate level of stress may provide protection against the original stressor and sometimes against other stressors in later life and thereby add to health in later life and longevity. This process is called hormesis.

medical evidence connecting nutrition during life *in utero* and in infancy with adult health. He maintains that short-term disturbances in nutrition, especially during fetal life, as well as chronic malnutrition during early life, have long-term negative effects on health that may result in any of several chronic conditions (e.g., hypertension, stroke, coronary artery disease, and adult-onset diabetes). This conclusion has been subject to considerable debate and criticism (e.g., Ben-Shlomo and Smith 1991; Lundberg 1993; Vågerö and Leon 1994; Järvelin et al. 1998; Leon et al. 1998; Preston et al. 1998; Vaupel et al. 1998; Bengtsson and Lindström 2000, 2003). These other researchers focus on conditions in early childhood as compared to conditions in fetal life. In the mix of early-life adverse conditions, stress and mental disorder may add to the likelihood of later-life adverse circumstances and the intergenerational transmission of social adversity and poorer physical and mental health (e.g., Wickrama et al. 2005).

The research on early life influences mostly distinguishes the mortality experiences of different members of the same cohort according to their own early health experiences. Other research shows that the early collective life experiences of a cohort are reflected in higher mortality of the cohort in later life as compared with the preceding and following cohorts. Finch and Crimmins (2004) single out inflammatory disease processes that severely affect basic somatic functions as such a precursor to increased death rates in the adult years. On the other hand, Kannisto et al. (1997) found no difference in the later mortality experience of cohorts born during famine and cohorts born before and after the famine.

We can extend the focus on early-life morbidity to include the accumulation of risk through the entire life course, as some authors have done (Kuh and Ben-Schlomo 1997; Ben-Shlomo and Kuh 2002). They examined the later-life effect not only of fetal and childhood conditions but health conditions over the life span. Environmental and lifestyle influences on health, experienced at different stages of the life course, can affect the development of chronic diseases in different ways through their differential effect on physiological processes at different life stages (Lynch and Smith 2005).

Many health conditions of later life first make their clinical appearance in childhood, adolescence, or youth. For example, atherosclerosis has been found in many young men who were battle casualties. Most serious mental diseases, e.g., schizophrenia, depression, and schizoaffective disorder, first appear in the teen ages or youth. Some disease-associated conditions, such as being obese, appear first in childhood. Ferraro et al. (2003) have demonstrated the importance of excessive weight in childhood as a risk factor for obesity over the life course, especially between the ages of 45 and 64. This finding does not preclude a role for genetic and other factors as well in the chain of risk linking excess weight in childhood and severe obesity in later life.

On the positive side, good health in one's younger years fosters good health in the later years. If one becomes informed about practices conducive to good health as a child, follows good health practices during those years, and secures the care of health providers as needed, one's chances for good health in later life are greatly

improved. A favorable socioeconomic status in childhood and youth can contribute to good health practices and experiences in childhood, which become a strong foundation for maintaining good health in later life. A favorable socioeconomic status as a child or youth is defined by having parents with greater education, higher income, or a white-collar occupation.

Analysts have long noted the importance of cohort changes in accounting for the improvements in generational health. They report a general tendency for the mortality of infants and children to decline much earlier than for adults (Preston and van de Walle 1978; Kermack et al. 1934). Wilmoth (1988) and Kannisto (1994) maintain, however, that the changes that occur over the life course, such as the rise in the socioeconomic status of members of a cohort and the improvements in public health and medical practice in the community as the members of the cohort grow older, are more important than childhood circumstances in affecting health outcomes in later years.

It is evident that the momentum of cohort succession is superimposed on the role of early life influences on health in later life and may modify them. Given the considerable class mobility in the United States, the generally improving health of the younger population is reflected in improved health in a cohort in later years. This change results in the cohort's having more healthful children who, in turn, become more healthful adults. Thus, upward mobility leads to better health for the families of the upwardly mobile, which they "pass on" to their children by example, education, and provision of better care. These changes have been supported by progress in social and biological medicine and in medical technology. Medical and other human interventions have contributed greatly to the improvement of the health of children and youth as well as adults. The children have benefited particularly from immunizations, water purification, cleaner air, and a greater supply of nutritious food, and the adults have benefited from coronary artery, cardiac, and other surgery, organ transplantation, new diagnostic methods, improved medications for chronic diseases, and similar developments.

In sum, the focus on early life influences can be extended to life-course health-associated experiences, to cover transitions from childhood and youth to adulthood and older age. This perspective broadens the basis of late-life mortality and morbidity beyond the events of early life to include the totality of experiences throughout the life course. We can identify three component periods for the environmental/behavioral factor in later-life health and longevity, namely, early-life influences, midlife transitions, and current adult living conditions. Complementing the environmental/behavioral factor is the genetic factor. As stated, these factors interactively influence one another to determine health and longevity in later life. In addition, these factors are heavily influenced by chance at every stage of the life course. In the case of extrinsic (exogenous) morbidity and mortality, e.g., accidents, the role of the genetic factor is vastly reduced and may be described as secondary, whereas in the case of intrinsic (endogenous) morbidity and mortality, the role of the genetic factor is primary, but not usually determinative. A sketch of these relationships is shown in Exhibit 6.4.

Reclassification of Diseases According to Social/Preventive Causes

The above discussion of the genetic, environmental, lifestyle, and stochastic factors contributing to disease and causes of death included in the WHO international classification provides the background for an alternative grouping of the causes, one that goes beyond the conventional classification and even the classification of the causes as exogenous and endogenous. This alternative classification divides the causes into those that can be considered preventable by the individual and society and those that cannot. It calls for a division of the cause categories commonly classified as endogenous into those caused by social factors such as inactivity, use of tobacco, poor diet, and infectious agents and those caused by purely genetic and biological factors. This reclassification recognizes, for example, the role of tobacco use in causing lung cancer and heart disease, or of a poor diet and inactivity in causing cardiovascular disease.

One such reclassification was carried out by McGinnis and Foege (1993), who reviewed the literature on the etiology of the conventional diseases as listed on death certificates and developed a new distribution of the diseases in the United States for 1990 in terms of social/nonsocial causes. Some would call these social causes the "real" causes of the deaths with which they are associated. According to the recast distribution of deaths derived by McGinnis and Foege, of the 2,148,000 deaths in the United States in 1990, they classified 1,060,000, or about half, as socially caused and hence, preventable. About 19%, or 400,000, were attributed to the use of tobacco, 14%, or 300,000, to poor diets and inactivity, and 5%, or 100,000, to the use of alcohol. Smaller numbers were assigned to microbial agents, toxic agents, firearms, unsafe sexual behavior, motor vehicles, and illicit use of drugs. Recall that the leading "real" cause, tobacco use, has been indicted in the causation of a very wide range of lethal diseases: cancer, especially cancer of the lungs, esophagus, oral cavity, bladder, pancreas, and kidney; the leading cardiovascular diseases; chronic obstructive pulmonary diseases, particularly asthma, bronchitis, and emphysema; prematurity and low birth weight of babies of mothers who smoke; and accidental deaths from cigarettes causing fires.

The obvious implication of this reclassification is that a large share of all deaths is preventable even though it may not be possible to eliminate all "social" deaths entirely. If the preventable deaths were wholly eliminated, life expectancy would be increased by possibly 8 years. A substantial addition to life expectation would be achieved if only the population practiced a more healthful life style. For example, eliminating the effect of smoking would add possibly 2–3 years to life expectancy.

Different analysts would be expected to assign some causes to different categories than McGinnis and Foege did. More than one social cause may be associated with the same health outcome. Multiple causes may contribute synergistically to a particular outcome. For example, poor diet, inactivity, and smoking lead to heart disease and stroke. The analyst has the difficult task of apportioning the conventional underlying causes among the social causes without overlap. Cultural differences in different countries and in different regions of the same country may suggest

assigning different shares of deaths to different social causes. Furthermore, an evolution of thinking in the interpretation of the causes of morbidity and mortality may dictate an allocation of the causes to different social factors.

Obesity is now emerging as a new villain, replacing smoking. The proponents of the view that "obesity is the new smoking scourge" predict that the decline in life expectancy may come to an end within a few decades in the United States when the obese children of today reach adulthood (Olshansky et al. 2005). In the face of a so-called epidemic of obesity in the United States in the 1970–2004 period, the Centers for Disease Control and Prevention (CDC) announced that an estimated 400,000 deaths were due to obesity in 2005. For this estimate, obesity was presumably charged with many of the deaths from cardiovascular disease, cancer, diabetes, and other causes that were previously attributed to tobacco, inactivity, and poor diet.

Critics of this estimate argue that the boundaries of overweight and obesity were set too low, that a moderate degree of overweight is favorable for elderly persons, that obesity is merely a visible marker for other less obvious factors that account for many lethal diseases, and that there is no evidence that the rise in obesity has caused a rise in heart disease or cerebrovascular disease (diseases which, as we may recall, declined sharply in the period 1970–2004). (See Campos 2004; Oliver 2005; Gibbs 2005.) However, the rise in obesity can be associated with the rise in deaths from diabetes between 1988 and 2004. The CDC modified its estimate of deaths due to obesity after it was criticized for overstating its case; its latest estimate of obesity-related deaths is 365,000. A recent study by Flegal et al. (2005), based on all three NHANES surveys, maintains that the evidence is lacking that there is any measurable mortality toll among overweight Americans as a group in relation to people who are of healthy weight.

Limited Effectiveness of Personal Behavior in Retarding Age-Related Health Changes

Many research studies point to the differences in longevity between those who are free from the main risk factors and those who are not, but evidence that risk-factor modification is fully effective in preventing the occurrence or halting the progression of the leading chronic diseases is not strong. Regular exercise, proper eating, avoidance of stress, better sleep practices, cessation of smoking, and moderate consumption of alcohol may aid in slowing the progression of some diseases for some people, but their effectiveness in retarding age-related health conditions is limited. That is because of the continuing and powerful effect of age-related physiological dysregulation of the cells of the body and the important roles of genetics and chance in the etiology of disease. For example, although hyperlipidemia, hypertension, hyperglycemia, and hyperhomocysteinemia have been identified as risk factors for peripheral artery disease and coronary artery disease, efforts to modify these risk factors will not necessarily prevent these disease outcomes or keep them from becoming progressively worse. Millions of Americans

who closely follow the guidelines for a healthy lifestyle still acquire many serious diseases and, having acquired them, cannot stop their inexorable and relentless progression. Risk-factor modification may be more effective in this regard in the earlier ages of the post-reproductive years than in the later ages of life.

Demographic Characteristics of the Disabled and Nursing Home Populations

This chapter concludes with some notes on the demographic characteristics of the disabled and the nursing-home populations.

Disabled Population

According to U.S. American Community Survey (ACS), the disabled population 5 years old and over in the United States numbered an estimated 39.8 million in 2005, or 14.9% of the population 5 years old and over. The survey questions distinguished several types of disability: self-care, sensory, physical, mental, employment, and "go-outside-home" disability (*i.e.*, inability to go alone to shop or visit a doctor's office). The survey total excludes the disabled population living in group quarters, which encompasses nursing homes, correctional institutions, college dormitories, military quarters, and group homes. In spite of the fact that, according to the 2000 census, all or nearly all residents of nursing homes have at least one disability, adjustment of the data for the omission of the nonhousehold population in the ACS of 2005 raises the original ACS percentage to only about 15.3%. This adjusted ACS figure is well below the 2000 census figure of 19.3%, which includes the nonhousehold population as well as the household population.

As expected, age-specific disability ratios rise with advancing age. For the household population in 2005, the disability ratio is 6% for ages 5–15 years, 12% for ages 16–64 years, and 41% for ages 65 years and over (Table 6.10; Fig. 6.9). (See also Freedman et al. 2004; Lutz and Scherbov 2003.) The adjusted ACS figure for ages 65 years and over for 2005 is 43%, as compared with the reported ACS figure of 41%. The adjustment of the ACS to include the nonhousehold population raises the figure for the percent disabled among the elderly population only modestly. Among the elderly, a limitation in physical functioning (i.e., walking, climbing stairs, reaching, lifting, carrying) is by far the most common type of limitation, affecting 31% of that population. The second most common limitation, affecting some 17% of the elderly population, is "going-outside-the-home" disability.

The disabled elderly make up a substantial share of the disabled population 5 years old and over (one-third), but the great majority of persons with disability limitations are younger than age 65 (two-thirds), and they are mostly persons of

Table 6.10 Disability status of the U.S. population, by age, sex, and type of disability: 2005

Type of disability and age	Total	Male	Female
Population 5 years and over	267,388	130,302	137,086
Without any disability	85.1	85.7	84.6
With one type of disability	6.9	7.1	6.7
With two or more types of disability	8.0	7.2	8.7
Population 5–15 years	44,586	22,811	21,776
With any disability	6.5	8.0	4.8
With a sensory disability	1.2	1.3	1.1
With a physical disability	1.2	1.3	1.1
With a mental disability	5.2	6.8	3.6
With a self-care disability	0.9	1.1	0.7
Population 16–64 years	188,041	92,647	95,394
With any disability	12.1	12.0	12.2
With a sensory disability	2.8	3.3	2.4
With a physical disability	7.2	7.7	6.7
With a mental disability	4.5	4.5	4.5
With a self-care disability	2.0	1.8	2.2
With a go-outside-home disability	3.0	2.6	3.4
With an employment disability	6.8	6.6	7.1
Population 65 years and over	34,761	14,844	19,916
With any disability	40.5	38.2	42.1
With a sensory disability	16.4	18.2	15.1
With a physical disability	30.8	27.4	33.3
With a mental disability	11.5	10.6	12.1
With a self-care disability	9.7	7.8	11.0
With a go-outside-home disability	16.6	12.0	20.0
Population 65–74 years	18,360	8,401	9,959
With any disability	30.3	30.0	30.5
With a self-care disability	5.8	5.0	6.4
Population 75 years and over	16,400	6,443	9,957
With any disability	51.9	49.0	53.8
With a self-care disability	14.0	11.0	15.6

Source: U.S. Census Bureau (2005) American Community Survey. www.census.gov
Note: See text for definition of types of disability
Percents of population in age group, except for absolute population numbers shown for each age group. Data are limited to the household population and exclude the population living in institutions and other group quarters. Data are based on a sample, and are subject to sampling error and to various types of nonsampling error. Numbers are given in thousands

working age.[17] The proportions disabled are higher for females than for males at every age above about age 40 and, accordingly, at the older ages disabled women greatly outnumber disabled men. Both self-reports and performance-based measures

[17]According to the 2000 census 9% of the nursing-home population (155,000) is in the age group 16 to 64, the "working ages." All of these persons may be assumed to have one or more

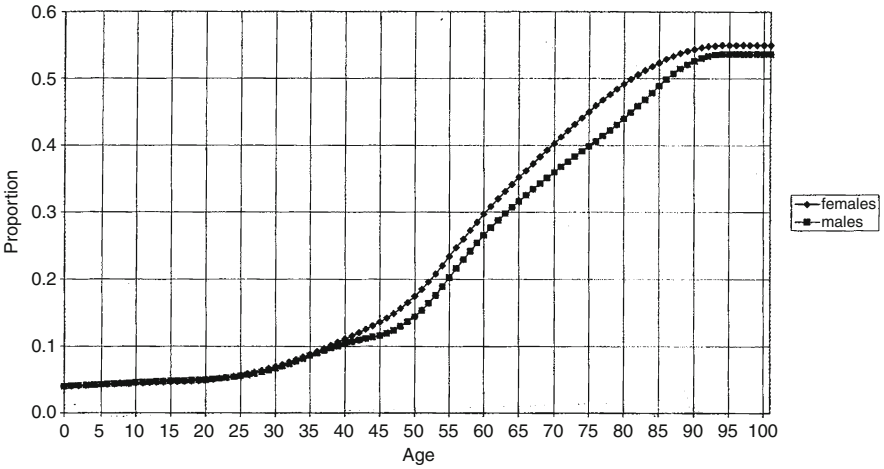


Fig. 6.9 Proportion disabled, graduated to single years of age, for males and females in the European Union-15: 1994–1996 (Source: Lutz and Scherbov 2003), Figure 2. Reprinted with permission. Primary source: European Community Household Panel (ECHP), 1994 and 1996; data provided by J.-M. Robine to Lutz and Scherbov)

indicate that blacks have significantly higher disability levels than whites, but socioeconomic status accounts for a substantial share of this difference (Mendes de Leon et al. 2005). Of the persons with disabilities, 91% live in housing units, 2% live in noninstitutional group quarters, and 6% live in institutional facilities.

There is strong evidence that age-specific disability ratios in the United States have been falling in recent decades. During the 1970s the increase in life expectancy was associated with an increase in years of disability of persons aged 65 and over, but during the 1980s and 1990s added years of life were associated with a substantial decrease in the proportion of disabled persons at these ages (Crimmins 2001, 2004; Manton and Gu 2001; Freedman et al. 2002; Manton et al. 2006). Further, the analysis of Manton et al., based on the National Long-Term Care Survey, shows that the decrease was greater during the 1990s than during the 1980s. The age-standardized figures for the proportion of chronically disabled persons decreased steadily, on a per annum basis, by 0.6% from 1982 to 1984 and by 2.2% from 1999 to 2004/2005. These declines were found to be significant for both persons with less severe disability and persons with more severe disability. Manton et al. estimate that, if the disability ratios of 1982 had not changed since that year, there would have been 9.2 million disabled persons 65 years and over in 1999 rather than the actual 7.0 million (Fig. 6.10). These declines in disability cannot be interpreted to mean that the health of the U.S. population is better now than a few decades

disabilities. About 22% of the noninstitutional group-quarters population (818,000) and over half of the institutional population excluding the nursing-home population (800,000) is disabled. The great majority of these persons are under 65 years of age.

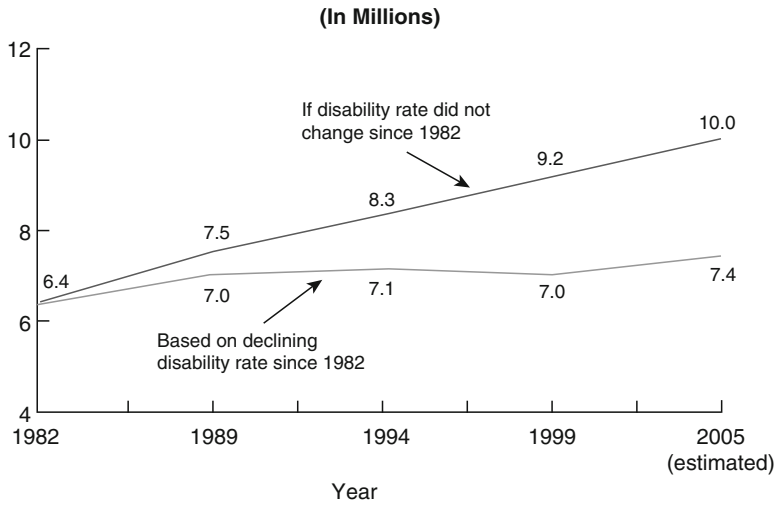


Fig. 6.10 Trend in actual numbers of disabled persons 65 years of age and over and in numbers assuming no change in disability ratios after 1982: United States,1982 to 2004/2005 (Source: Based on Manton et al. (2006), Copyright © (2006) National Academy of Sciences, U.S.A. Reprinted with permission. Primary source: U.S. National Long-Term Care Survey, 1982–2004/2005)

earlier because alternative definitions of health (e.g., chronic disease prevalence) give contrary results (Parker and Thorslund 2007).

Lutz and Scherbov (2003) attribute most of the decline in disability ratios in the United States to the rise in educational attainment in the United States. They note that disability ratios tend to be much lower for the more educated segments of the population and that educational levels have been rising for the elderly population. The rise in educational levels at each age have combined with the variations in disability according to educational level and with the declines in disability within each educational level to account for the large decrease in the overall disability ratio at each age among the elderly and in the general disability ratio of the elderly.

About 16% of the population of the European Union was disabled in 1994–1996, according to the European Community Household Panel (Lutz and Scherbov 2003), but data on trends in disability for the European Union and other regions of the world, such as are available for the United States, are lacking. Jacobzone et al. (1998) reported disability ratios for six European countries at a single date, but not for a series of dates. However, Egidi (2003) reported decreases in the disability ratios of the elderly in selected countries of Europe for the 1980s and 1990s.

These findings have important implications for the outlook of the growth of the disabled population in the United States and Europe. If we combine the expectation of a tremendous aging of the population of the United States and Europe in the next half century (caused in part by the advent at the older ages of the massive number of baby boomers) with the fact that disability ratios increase with advancing age, we should expect a rapid growth in the number of disabled people in the United States and Europe in the next few decades. This expectation, however, is based on

the assumption that disability ratios will remain at present levels. If, as has recently occurred in the United States, disability ratios fall, it is conceivable that the number of disabled persons will not rise and may even fall in the United States (Fig. 6.10).

Sensitivity calculations to evaluate the effect of falling disability ratios have also been carried out for Europe as for the United States. If Europe experiences a decline in disability ratios at each age, such as the United States did, it is possible that population aging in Europe will not result in an increase in the number of persons with disabilities. Lutz and Scherbov (2003) show that, if disability ratios are shifted only two ages up the age scale per decade, the number and proportion of disabled persons will be about the same in 2050 as in 2000. Further, if the disability ratios are shifted three ages up the age scale per decade, the number and proportion of disabled persons will fall by 2050.

Nursing-Home Population

In the United States and many European countries, there has been, and continues to be, a tremendous increase in the number and proportion of persons who live alone and a corresponding decrease in the proportion of persons living in married-couple households and multigenerational households. This is significant because single persons are much more likely to require formal health care than married persons and are, therefore, at greater risk than married persons of entering an institution. Yet, nursing-home residence has been declining in the last several years. According to the National Nursing Home Survey (NNHS), 1,492,000 persons resided in nursing homes in the United States in 2004, while 1,628,000 persons resided in them in 1999.[18] In the two prior decades, from 1977 to 1999, the number had been steadily increasing but its share of the total population remained steady.

The residents of nursing homes in the United States as a group are quite old and are getting older. The NNHS of 2004 reported that over three-quarters of the residents of U.S. nursing homes were 75 years of age or over and nearly half were 85 years or over:

	Percent of all ages		
Age group	2004	1999	1977
Under 65	11.7	9.7	13.0
65–74	11.7	12.0	16.2
75–84	31.4	31.8	36.0
85 and over	45.2	46.5	34.8
All ages	100.0	100.0	100.0

Source: U.S. National Center for Health Statistics/ Decker 2004; NCHS (2008)

[18] Alternative estimates are provided by the 2000 census and the National Long Term Care Survey. The census figure is 1.72 million, for example.

Yet, many persons under 65 years of age reside in nursing homes; about 12% of the nursing home population, numbering 175,000 persons, was under 65 years of age in 2004. The nursing-home population has been aging, as shown by the sharp increase in the share of residents over age 85 in the two decades prior to 1999. This would be expected, given the aging of the general population during the same period.

The nursing-home population is a largely female population, only one-quarter of the residents being male. The main reason for this imbalance of the sexes is that women have greater longevity than men and, hence, are more likely to survive to the oldest ages, usually as widows. Recall that the chronic illnesses of men tend to be "killers" while the chronic illnesses of women tend to be "disablers." The widowed status of women also means that they are less affluent and usually cannot afford private paid caregivers. Elderly persons tend to enter institutions for a combination of reasons, including mainly the unavailability of, or inability to afford, private caregivers, in addition to functional impairments and mental infirmity. Whites are overrepresented in nursing homes; blacks constitute 12% and Hispanics only 2% of the residents in nursing homes. Blacks and Hispanics show a great readiness to care for their elders at home and typically lack the means to support institutional care for them.

About one-third to one-half of persons 65 years old and over may expect to enter a nursing home some time in their lives. Two out of three persons will either never enter a nursing home or will stay in one for less than 3 months. Nursing home stays tend to be relatively short in spite of the public's conception of their duration. Although the mean length of stay of residents from the time of admission, according to the NNHS of 2004, is over 2 years, the distribution is skewed sharply to the right because more than one-quarter of the residents stay more than 3 years. The median length of stay is only a little over a year. The percent distribution of residents of nursing homes by length of stay since the time of admission according to the NNHS of 2004 and a few earlier years is as follows:

	Percent		
Length of stay	2004	1999	1977
Less than 3 months	29.8[a]	17.8	14.4
3 months to <1 year	14.3[b]	25.0	22.1
1 year to <3 years	30.3	30.1	32.8
3 years or more	25.6	27.1	30.7

Source: U.S. NCHS/Decker 2004; NCHS 2008.
[a]Less than 6 months
[b]6 months to less than 1 year

These figures indicate not only that a large share of the residents stay less than 1 year but that the residents were staying for shorter periods in 2004 than in 1999 and 1977. Other pertinent information regarding the trend in the length of stay is provided by the rate of discharge for stays of less than 3 months per 100 nursing-home beds. The rate doubled between 1977 and 1999, rising from 46 in 1977 to 92 in 1999, while the rate of discharge for the other lengths of stay remained steady in this period.

We tend to assume that the entire nursing-home population is disabled in one way or another. The NNHS of 1999 shows that this is almost true. About half can eat on their own but only 6 percent can bathe themselves and only 13% can dress themselves. Two decades earlier the situation was much more favorable, when 30% could dress themselves, 13% could bathe themselves, and two-thirds could eat without assistance. This trend may be accounted for by the aging of the elderly population and the rise of assisted living residences and other alternative ways of providing long-term care. These residential alternatives to nursing homes take care of persons less in need of medical and rehabilitative care.

References and Suggested Readings

Measuring Inequality in Health and Mortality

Barbi, E., & Caselli, G. (2003). Selection effects on regional differences in survivorship in Italy. *Genus, 49*(2), 37–62.

Gakidou, E. E., Murray, C. J. L., & Frenk, J. (2000). Defining and measuring health inequality: an approach based on the distribution of health expectancy. *Bulletin of the World Health Organization, 78*, 42–52.

Goesling, B., & Firebaugh, G. (2004). The trend in international health inequality. *Population and Development Review, 30*(1), 131–146.

Murray, C. J. L., Gakidou, E. E., & Frenk, J. (1999). Health inequalities and social group differences: what should we measure? *Bulletin of the World Health Organization, 77*, 537–543.

Regidor, E. (2004). Measures of health inequalities: Part 1. *Journal of Epidemiology and Community Health, 58*, 858–861.

Siegel, J. S. (1993). *A generation of change: A profile of America's older population*. New York: Russell Sage Foundation.

Siegel, J. S., & Swanson, D. A. (Eds.). 2004. *Methods and materials of demography*. (2nd ed.). San Diego, CA: Elsevier.

Trends in Health and Mortality

Caldwell, J. C., & Santow, G. (1989). *Selected readings in the cultural, social, and behavioral determinants of health* (Health Transition Series, No. 1). Canberra, Australia: Health Transition Centre, Australian National University.

Carnes, B. A., & Olshansky, S. J. (1993). Evolutionary perspectives on human senescence. *Population Development Review*, *19*, 793–806.

Carnes, B. A., Olshansky, S. J., & Grahn, D. (1996). Continuing the search for a law of mortality. *Population Development Review*, *22*, 231–264.

Casselli, G., & Lopez, A. D. (Eds.). (1996). *Health and mortality among elderly populations*. Oxford, UK: Clarendon Press.

Costa, D. L. (2002). Changing chronic disease rates and long-term declines in functional limitation among older men. *Demography*, *39*(1), 119–137.

Costa, D. L. (2005). Causes of improving health and longevity at older ages: A review of the explanations. *Genus*, *LXI*(1), 21–38.

Cutler, D., & Miller, G. (2005). The role of public health improvements in health advances: The twentieth century United States. *Demography, 41*(1), 1–11.

Elman, C., & Myers, G. C. (1997). Age and sex differentials in morbidity at the start of an epidemiological transition: Returns from the 1880 U.S. Census. *Social Science and Medicine, 45*, 943–956.

Elman, C., & Myers, G. C. (1999). Geographic morbidity differentials in the late nineteenth-century United States. *Demography, 36*, 429–443.

Ewbank, D. (2004). From Alzheimer's disease to a demography of chronic disease: The development of demographic synthesis for fitting multistate models. In L. J. Waite (Ed.), *Aging, health, and public policy: Demographic and economic perspectives* (pp. 63–83). A Supplement to Vol. 30, 2004, *Population and Development Review*.

Fried, L. P. (2000). Epidemiology of aging. *Epidemiological Review, 22*(1), 95–106.

Golini, A. (2002). Teaching demography of aging. *Genus, LVIII*(3–4), 135–164.

Gribble, J. N., & Preston, S. H. (1993). *The epidemiological transition: Policy and planning implications for developing countries*. Washington, DC: National Academy Press.

Herbert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D.A., & Evans, D.A. (2003). Alzheimer's disease in the U.S. population: Prevalence estimates using the 2000 census. *Archives of Neurology, 60*(8), 1119–1122.

Hotez, P. J. (2010, January). A plan to defeat neglected tropical diseases. *Scientific American, 302*(1), 90–96.

Johansson, S. (1991). The health transition: The cultural inflation of morbidity during the decline of mortality. *Health Transition Review, 1*(1), 39–68.

Kessler, R. C., Chiu, W. T., Demler, O., et al. (2005). Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry, 62*(6), 617–627.

Lancaster, H. O. (1990). *Expectations of life: A study in the demography, statistics, and history of world mortality*. New York: Springer.

Linder, F. E., & Grove, R. D. (1947). *Vital statistics rates in the United States: 1900–1940*. Washington, DC: U.S. National Office of Vital Statistics/U.S. Government Printing Office.

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., & Murray, C. J. L. (Eds.). (2006). *Global burden of disease and risk factors, 1990–2001*. New York: Oxford University Press and World Bank.

Malaspina, D. (2002, Summer 2002). *Narsad Research Newsletter, 14*(2).

National Research Council, Institute of Medicine. (2000). *Informing the future: critical issues in health*. Washington, DC: National Academy Press.

Olshansky, S. J., & Ault, B. (1986). The fourth stage of the epidemiologic transition: The age of delayed degenerative diseases. *Milbank Quarterly, 64*(3), 355–391.

Olshansky, S. J., Carnes, B. A., Rogers, R. G., & Smith, L. (1997). Infectious diseases: New and ancient threats to world health. *Population Bulletin, 52*(2), 1–52. Washington, DC: Population Reference Bureau.

Omran, A. R. (1971). The epidemiologic transition: A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly, 49*, 509–538.

Omran, A. R. (1977). Epidemiologic transition in the United States: The health factor in population change. *Population Bulletin*, *32*(2). Washington, DC: Population Reference Bureau.

Parker, M. G., & Thorslund, M. (2007). Health trends in the elderly population: Getting better and getting worse. *Gerontologist*, *47*(2), 150–158.

Population Reference Bureau. (2005). *World Data Sheet, 2005*. Washington, DC: Population Reference Bureau.

Reddy, K. S. (2004). Cardiovascular disease in non-Western countries. *The New England Journal of Medicine, 350*(24), 2438–2440.

Riley, J. C. (1990). The risk of being sick: Morbidity trends in four countries. *Population and Development Review, 16*(3), 403–431.

Riley, J. C. (1992). From a high mortality regime to a high morbidity regime: Is culture everything in sickness? *Health Transition Review, 2*(1), 71–78.

Riley, J. C. (2001). *Rising life expectancy: A global history*. Cambridge, UK: Cambridge University Press.
Riley, J. C. & Alter, G. (1996). The sick and the well: Adult health in Britain during the health transition. *Health Transition Review, 6*, 19–44.
Rogers, R., & Hackenburg, R. (1987). Extending epidemiologic transition theory: A new stage. *Social Biology, 34*, 234–243.
Sonnenschein, E., & Brody, J. (2005). Effect of population aging on proportionate mortality from heart disease and cancer, U.S. 2000–2050. *Journal of Gerontology: Social Sciences, 60B*(2), S110–S112.
United Nations, Department of Economic and Social Affairs, Population Division. (2007). *World population prospects: The 2006 revision, highlights*. New York: United Nations.
U.S. National Center for Health Statistics. (2007a). Deaths: Final data for 2004. By A. M. Miniño, M. P. Heron, S. L. Murphy, & K. D. Kochanek. *National Vital Statistics Reports, 55*(19).
U.S. National Center for Health Statistics. (2007b). Deaths: Leading Causes for 2003. By M. P. Heron & B. L. Smith. *National Vital Statistics Reports, 55*(10).
Vallin, J., & Meslé, F. (2005). Convergence and divergence: an analytical framework of national and sub-national trends in life expectancy. *Genus, LXI*(1), 57–82.
Verbrugge, L. (1984). Longer life but worsening health? Trends in health and mortality of middle-aged and older persons. *Milbank Memorial Fund Quarterly, 62*, 474–519.
Weinberg, A. N. (2006). Zoonoses: Where the wild things are. Lecture, March 8, 2006. *Great Teachers Lecture Series*. National Clinical Center, National Institutes of Health.
World Bank. (2004). *Millennium development goals for health in Europe and Central Asia. Relevance and policy implications*. Washington, DC: World Bank.
World Health Organization. (2002). *World health report 2002: Reducing risk, promoting healthy life*. Geneva, Switzerland: World Health Organization. www.who.int.
World Health Organization. (2003). *Global burden of disease in 2002: Data source, methods, and results*. Geneva, Switzerland: World Health Organization. www.who.int.
World Health Organization. (2004). *The European health report*, 2002. www.who.dk/europeanhealthreport.
World Health Organization and the World Bank. (1996). *The global burden of disease: summary*. C. J. L. Murray & A. D. Lopez (Eds.). Cambridge, MA: Harvard School of Public Health.
Yach, D., Leeder, S. R., Bell, J., & Kistnasamy, B. (2005, January 21). Global chronic diseases. *Science, 307*(5708), 317.
Yanagishita, M., & Guralnik, J. M. (1988). Changing mortality patterns which led life expectancy in Japan to surpass Sweden: 1972–1982. *Demography, 25*, 611–624.

Etiological and Risk Factors

AARP Foundation and ILC-USA. (2003). Getting your Zzzzzz's: How sleep affects health and aging.
Aronow, W. S. (2003). "Commentary" on J. S. Goodwin. *Journal of Gerontology: Medical Sciences, 58A(7)*, 659–661.
Boshuizen, H. C., Izaks, G. J., VanBuure, S., & Ligthart, G. J. (1998). Blood pressure and mortality in elderly people aged 85 and older: Community based study. *British Medical Journal, 316*, 1780–1784.
Campos, P.F. (2004). *The obesity myth: Why america's obsession with weight is hazardous to your health*. New York: Gotham Books.
Cappuccio, F. P. (2007). *Address before the British sleep society, Cambridge*, 2007. Coventry, UK: Warwick Medical School.
Carnes, B. A. (2005, November 20). *Do genes govern the aging process*? Paper presented at the annual meeting of the Gerontological Society of America, Orlando, FL.

Carnes, B. A., Holden, L. R., Olshansky, S. J., Witten, T. M., & Siegel, J. S. (2006). Mortality partitions and their relevance to research on senescence. *Biogerontology, 7*, 183–198.

Carroll, M., Lacher, D. A., Sorlie, P. D., et al. (2005). Trends in serum lipids and lipoproteins of adults, 1960–2002. *Journal of American Medical Association, 294*(14), 1773–1781.

Case, A., & Deaton, A. (2003). *Broken down by work and sex*: How our health declines (National Bureau of Economic Research Working Paper No. 9821).

DiMaria-Ghalili, R. A., & Guenter, P. A. (2008). The mini nutritional assessment. *American Journal of Nursing, 108*(2), 50–59.

Epel, E., Blackburn, E., Cauthon, R. M., et al. (2004). Accelerated telemere shortening in response to chronic stress. *Proceedings National Academy of Sciences, 101*, 17312–17315.

Finch, C. E., & Kirkwood, T. B. L. (2000). *Chance, development, and aging*. New York: Oxford University Press.

Finch, C., & Tanzi, R. E. (1997). Genetics of aging. *Science, 278*, 407–411.

Flegal, K. M., Carroll, M. D., Kuczmarski, R. J., & Johnson, C. L. (1998). Overweight and obesity in the United States: Prevalence and trends, 1960–1994. *International Journal of Obesity and Related Metabolic Disorders, 22*, 39–47.

Flegal, K. M., Graubard. B. I., Williamson, D. F., & Gail, M. H. (2005). Excess deaths associated with underweight, overweight, and obesity. *Obstetrical and Gynecological Survey, 60*(9), 593–595.

Gibbs, W. W. (2005). Obesity: An overblown epidemic? *Scientific American*, *292*(6), 70–77.

Goldman, N. (2002). Incorporating health into demographic training. *Genus*, *LVIII*(3–4), 91–107.

Goodwin, J. S. (2003). Embracing complexity: A consideration of hypertension in the very old. *Journal of Gerontology: Medical Sciences, 58A*(7), 653–658. See also commentaries on this paper by W. S. Aronow, S. Denson, R. R. Hajjar, T. B. Harris, D. T. Lowenthal, J. -P. Michel et al., A. B. Newman, and D. T. Thomas.

Hairston, I., Little, M. T. M., Scanlon, M. D., et al. (2005). Sleep restriction suppresses neurogenesis induced by hippocampus-dependent learning. *Journal of Neurophysiology, 94*(6), 4224–4233.

Hajjar, R. R. (2003). "Commentary" on J. S. Goodwin. *Journal of Gerontology: Medical Sciences, 58A(7)*, 661–662.

Hakala, S. M., Tilvis, R. S., & Strandberg, T. E. (1997). Blood pressure and mortality in an older population. *European Heart Journal, 18*, 1019–1023.

Hayflick, L. (2001). Giving chance a chance or, if there are no dice, there is no game. Book Review of Finch and Kirkwood (2000). *Gerontologist, 41*(2), 272–275.

Hjelmborg, J. vB., Iachine, I., Skytthe,A., Vaupel, J. W., et al. (2006). Genetic influence on human lifespan and longevity. *Human Genetics, 119*(3), 312–321.

Hodes, R. J. (2005, November). Growing older: Challenges and opportunities in aging. Public address at the National Institutes of Health, Bethesda, MD, by Director, National Institute on Aging.

Jakicic, J. M., Marcus, B. H., Gallagher, K. I., Napolitano, M., & Lang, W. (2003). Effect of exercise dose and intensity on weight loss in overweight sedentary women: A randomized trial. *Journal of American Medical Association, 290*(10), 1323–1330.

Kuczmarski, R. J., Flegal, K. M., Campbell, S., & Johnson, C. L. (1994). Increasing prevalence of overweight among US adults: The National Health and Nutrition Examination Surveys, 1960–1991. *Journal of American Medical Association, 272*, 205–211.

Lew, E. A., & Garfinkel, L. (1990). Mortality at ages 75 and older in the Cancer Prevention Study (CPSI). *Ca-A Cancer Journal for Clinicians, 40*, 210–224.

Ling, S. M., Simonsick, E. M., & Ferrucci, L. (2007). Editorial. A painful interface between normal aging and disease. *Journal of Gerontology: Medical Sciences*, *62A*(6), 613–615.

Locher, J. L., Roth, D. L., Ritchie, C. S., et al. (2007). Body mass index, weight loss, and mortality in community-dwelling older adults. *Journal of Gerontology: Medical Sciences, 62A*(12), 1389–1392.

Manton, K. G., & Stallard, E. (1994). Interaction of disability dynamics and mortality. In L. G. Martin & S. H. Preston (Eds.), *Demography of aging* (pp. 217–278). Washington, DC: National Academy Press.

Malaspina, D. (2002, Summer). *Narsad Research Newsletter 14*(2).

Mattila, K., Haavisto, M., Rajala, S., & Heikinheimo, R. (1988). Blood pressure and five-year survival in the very old. *British Medical Journal, 296*, 887–889.

McGinnis, J. M., & Foege, W. H. (1993). Actual causes of death in the United States. *Journal of American Medical Association*, *270*(18), 2207–2212.

Miles, L. E., & Dement, W. C. (1980). Sleep and aging. *Sleep, 3*(2), 1–220.

Mokdad, A., Serdula, M., Dietz, W., et al. (1999). The spread of the obesity epidemic in the United States, 1991–1998. *Journal of American Medical Association, 282*, 1519–1522.

National Research Council, Institute of Medicine. (2000). *Informing the future: critical issues in health*. Washington, DC: National Academy Press.

Newman, A. B. (2003). "Commentary" on J. S. Goodwin, "Embracing complexity: A consideration of hypertension in the very old." pp. 666–667. *Journal of Gerontology: Medical Sciences*. *58A*(7), 653–658.

Nicholson, A., Bobak, M., Murphy, M., & Marmot, M. (2004, March 31–April 3). Alcohol consumption is associated with increased all-cause mortality in Russian men and women: A study based on the mortality of relatives. Paper presented at the annual meeting of the Population Association of America, Boston, 2004.

Ogden, C. L., Carroll, M. D., Curtin, L. R., McDowell, M. A., et al. (2006). Prevalence of overweight and obesity in the United States, 1999–2004. *Journal of American Medical Association*, *295*(13), 1549–1555.

Oliver, J. E. (2005). *Obesity: The making of an American epidemic*. Oxford, UK: Oxford University Press.

Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., et al. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine, 352*(11), 1138–1145.

Reed, T., Quiroga, J., Selby, J. V., et al. (1991). Concordance of ischemic heart disease in the NHLBI Twin Study after 14–18 years of follow-up. *Journal of Clinical Epidemiology, 44*, 797–805.

Rogers, R. G., Hummer, R. A., Kreeger, P. M., & Pampel, F. C. (2005). Mortality attributable to cigarette smoking in the United States. *Population and Development Review, 31*(2), 259–292.

Rowe, J. W., & Kahn, R. L. (1998). *Successful aging*. New York: Pantheon Books.

Seeman, T. E., & Crimmins, E. (2001). Social environment effects on health and aging: Integrating epidemiologic and demographic approaches and perspectives. *Annals of the New York Academy of Medicine, 954*(1), 88–117.

Seeman, T., & McEwen, B. S. (1996). Impact of social environment characteristics on neuroendocrine regulation. *Psychosomatic Medicine, 58*, 458–471.

Seeman, T. E., Singer, B. H., Ryff, C. D., Love, G. D., & Levy-Storms, L. (2002). Social relationships, gender, and allostatic load across two age cohorts. *Psychosomatic Medicine, 64*, 395–406.

Sellers, T. A., Bailey-Wilson, J. E., Elston, R. C., et al. (1990). Evidence for Mendelian inheritance in the pathogenesis of lung cancer. *Journal of National Cancer Institute, 82*, 1272–1279.

Strauss, R. S., & Pollack, H. A. (2001). Epidemic increase in childhood overweight, 1986–1998. *Journal of American Medical Association, 286*, 2845–2848

Troiano R. & Flegal, K. (1998). Overweight children and adolescents: Description, epidemiology, and demographics. *Pediatrics, 101*, 497–504.

U.S. Agency on Healthcare Research and Quality (AHRQ). (2005). *2005 National healthcare disparities report* (AHRQ Publication No. 06-0017). Rockville, MD: AHRQ.

U.S. National Center for Health Statistics. (2004). Mean body weight, height, and body mass index, United States 1960–2002. By C. L. Ogden, C. D. Fryar, M. D. Carroll, & K. M. Flegal. *Advance Data from Vital and Health Statistics* (No. 347). Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2005). Seroprevalence of six infectious diseases among adults in the United States by race/ethnicity: Data from the third National Health and Nutrition Examination Survey, 1988–94. By D. Kruszon-Moran & M. McQuillan. *Advance Data from Vital and Health Statistics* (No. 352). Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2006a). Summary health statistics for U.S. adults: National Health Interview Survey, 2004. By M. Lethbridge-Çejku, D. Rose, & J. Vickerie. *Vital and Health Statistics,* Series 10 (228). Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2006b). Health characteristics of adults 55 years of age and over: United States, 2000–2003. By C. A. Shoenborn, J. L. Vickerie, & E. Power-Griner. *Vital and Health Statistics: Advance Data* (No. 370). Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2006c). Health Behavior of Adults: United States, 2002–04. By P. F. Adams & C. A. Schoenborn. *Vital and Health Statistics*. Series 10 (230). Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2006d). *Health, United States: With chartbook on trends in the health of Americans, 2006*. Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2006e). Births: Final data for 2004. By J. A. Martin, B. E. Hamilton, P. D. Sutton, S. J. Ventura, et al. *National Vital Statistics Reports, 55*(1).

U.S. National Center for Health Statistics. (2009). One-third of U.S. adults embraced most heart healthy behaviors in 1999–2002. By J. D. Wright, R. Hirsch, & C. -Y. Wang. *NCHS Data Brief* (No. 17). Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Institute on Aging, National Institutes of Health. (1993). *In search of the secrets of aging*. Bethesda, MD: National Institute on Aging.

U.S. National Institute on Aging, National Institutes of Health. (2006). *Aging under the microscope: A biological quest* (NIH Publication No. 02-2756). Bethesda, MD: National Institute on Aging.

Vaupel, J. W., Carey, J. R., Christensen, K., et al. (1998). Biodemographic trajectories of longevity. *Science*, *280*, 855–860.

Yashin, A. I., & Iachine, I. A. (1997). How frailty models can be used in evaluating longevity limits. *Demography, 34*(1), 31–48.

Life Course Studies and Early Life Influences

Barker, D. J. P. (1997). Maternal nutrition, fetal nutrition and diseases in later life. *Nutrition*, *13*(9), 807–813.

Bengtsson, T., & Lindström, M. (2000). Childhood misery and disease in later life: The effects on mortality in old age by hazards experienced in early life, southern Sweden, 1760–1894. *Population Studies, 54*, 263–277.

Bengtsson, T., & Lindström, M. (2003). Airborne infectious diseases during infancy and mortality in later life in southern Sweden, 1766–1894. *International Journal of Epidemiology, 32*, 286–294.

Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual model, empirical challenges, and interdisciplinary perspectives. *International Journal of Epidemiology, 31*, 285–293.

Ben-Shlomo, Y., & Smith, G. D. (1991). Deprivation in infancy or in adult life: Which is more important for mortality risk? *Lancet, 337*, 530.

Blackwell, D. L., Hayward, M. D., & Crimmins, E. M. (2001). Does childhood health affect chronic morbidity in later life? *Social Science and Medicine, 52*, 1269–1284.

Costa, D. L., & Lahey, J. (2005). Becoming oldest-old: Evidence from historical U.S. data. *Genus, LXI*(1), 125–161.

Elo, I. T., & Preston, S. H. (1992). Effects of early-life conditions on adult mortality: A review. *Population Index, 58*(2), 186–212.

Elo, I. T., & Preston, S. H. (1996). Educational differentials in mortality: United States, 1970–85. *Social Science Medicine, 42*, 42–57.

Ferraro, K. F., Thorpe, Jr., R. J., & Wilkinson, J. A. (2003). The life course of severe obesity: Does childhood overweight matter. *Journal of Gerontology: Social Sciences, 58B*(2), S110–S119.

Finch, C. E., & Crimmins, E. M. 2004. Inflammatory exposure and historical changes in human life-spans. *Science, 305*, 1726–1730.

Fogel, R. W. (1993). New sources and new techniques for the study of secular trends in nutritional status, health, mortality, and the process of aging. *Historical Methods, 26*(1), 5–43.

Gavrilov, L. A., & Gavrilova, N. S. (2003). Early-life factors modulating lifespan. In S. I. S. Rattan (Ed.), *Modulating aging and longevity* (pp. 27–50). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Hayward, M. D., & Gorman, B. K. (2004). The long arm of childhood: The influence of early-life social conditions on men's mortality. *Demography, 41*(1), 87–107.

Järvelin, M. -R., Lauren, L., Chee, D., & Vartiainen, E. (1998). Intrauterine and life course factors in the aetiology of adult cardiovascular disease. *International Journal of Circumpolar Health, 57*(2–3), 114–122.

Kannisto, V. (1994). *Development of oldest-old mortality, 1950–1990*. Odense, Denmark: Odense University Press.

Kannisto, V., Christensen, K., & Vaupel, J. W. (1997). No increased mortality later in life for cohorts born during famine. *American Journal of Epidemiology, 145*(11), 987–994.

Kermack, W. G., McKendrick, A. G., & McKinlay, P. L. (1934). Death rates in Great Britain and Sweden: Some regularities and their significance. *Lancet, 31*, 698–703.

Komlos, J. (1993). The secular trend in the biological standard of living in the United Kingdom. *Economic History Review, 46*, 115–144.

Kuh, D., & Ben-Schlomo, Y. (Eds.). (1997). *A life course approach to chronic disease epidemiology*. Oxford, UK: Oxford University Press.

Leon, D. A., Lithell, H. O., & Våger?, D. (1998). Reduced fetal growth rate and increased risk of death from ischaemic heart disease: cohort study of 15,000 men and women born 1915–1929. *British Medical Journal, 317*, 241–245.

Lundberg, O. (1993). The impact of childhood living conditions on illness and mortality in adulthood. *Social Science and Medicine, 36*, 458–462.

Lynch, J., & Smith, G. D. (2005). A life course approach to chronic disease epidemiology. *Annual Review of Public Health, 26*, 1–35.

Manton, K. G. & Myers, G. C. (1987). Recent trends in multiple-caused mortality 1968 to 1982: Age and cohort comparisons. *Population Research and Policy Review, 6*, 161–176.

Marmot, M. G., & Shipley, M. J. (1996). Do socioeconomic differences in mortality persist after retirement? 25-year follow-up of civil servants from the first Whitehall study. *British Medical Journal, 313*, 1177–80.

O'Rand, A. M., & Hamil-Luker, J. (2005). Processes of cumulative adversity: Childhood disadvantage and increased risk of heart attack across the life course. *Journal of Gerontology:* Series B, *60B*(Special Issue II), 117–124.

Power, C., & Peckham, C. (1990). Childhood morbidity and adulthood ill health. *Journal of Epidemiology and Community Health, 44*, 69–74.

Preston, S. H., & van de Walle, E. (1978) Urban French mortality in the nineteenth century. *Population Studies, 32*(2), 275–297.

Preston, S. H., Hill, M. E., & Drevenstedt, G. L. (1998). Childhood conditions that predict survival to advanced ages among African-Americans. *Social Science and Medicine, 47*, 1231–1246.

Steckel, R. (1995). Stature and standard of living. *Journal of Economic Literature, 33*, 1903–1940.

Tsenkova, V. K., Love,G. D., Singer, B. H., & Ryff, C. D. (2008). Coping and positive affect predict longitudinal change in glycosylated hemoglobin. *Health Psychology, 27*(2)(Suppl.), S163–171.

Vågerö, D., & Leon, D. (1994). Ischaemic heart disease and low birth weight: A test of the fetal-origins hypothesis from the Swedish Twin Registry. *Lancet, 29*, 260–263.

Vaupel, J. W., Carey, J. R., Christensen, K., et al. (1998, May 8). Biodemographic trajectories of longevity. *Science, 280*, 855–860.
Wickrama, K. A. S., Conger, R. D., & Abraham, W. T. (2005). Early adversity and later health: The intergenerational transmission of adversity through mental disorder and physical illness. *Journal of Gerontology*, Series B, *60B*(Special Issue II), 125–129.
Wilmoth, J. R. (1988). *On the statistical analysis of large arrays of demographic rates*. Ph.D. dissertation, Department of Statistics and Office of Population Research, Princeton University, Princeton, NJ.

Disability and Frailty

Bortz, W. (2002). A conceptual framework of frailty: A review. *Journal of Gerontology: Medical Sciences, 57A*, M283–M288.
Crimmins, E. M. (2001). Are Americans healthier today than they were 20 years ago?, *Population Today,* Population Reference Bureau, *29*, 5,8.
Crimmins, E. M. (2004). Trends in the health of the elderly. *Annual Review of Public Health, 25*, 75–98.
Egidi, V. (2003). Health status of older people. *Genus, LIX*(1), 169–200.
Ferrucci, L., Windham, B. G., & Fried, L. (2002). Frailty in older persons. *Genus, LXI*(1), 39–56.
Freedman, V. A., Martin, L. G., & Schoeni, R. F. (2002). Recent trends in disability and functioning among older adults in the United States: A systematic review. *Journal of American Medical Association, 288*(24), 3137–3146.
Freedman, V. A., Martin, L. G., & Schoeni, R. F. (2004). Disability in America. *Population Bulletin, 59*(3). Washington, DC: Population Reference Bureau.
Fried, L. P, Tangen, C. M., Walston, J., et al. (2001). Frailty in older adults; evidence for a phenotype. *Journal of Gerontology: Medical Sciences, 56A*(3), M146–M156.
Jacobzone, S., Cambois, E., Champlain, E., & Robine, J. M. (1998). The health of older persons in OECD countries: *Is it improving fast enough to compensate for population aging?* (OECD Occasional Paper No. 37). Paris: Organization for Economic Co-operation and Development.
Lutz, W., & Scherbov, S. (2003). *Will population ageing necessarily lead to an increase in the number of persons with disabilities? Alternative scenarios for the European Union* (European Demographic Research Papers, No. 3). Vienna, Austria: Vienna Institute of Demography.
Manton, K., & Gu, X. (2001). Changes in the prevalence of chronic disability in the United States black and nonwhite population above age 65 from 1982 to 1999. *Proceedings of the National Academy of Sciences, 98*(11), 6354–6459.
Manton, K., Gu, X., & Lamb, V. (2006). Change in chronic disability from 1982 to 2004/05 as measured by long-term changes in function and health in the U.S. elderly population. *Proceedings of the National Academy of Sciences, 103*(48), 18374–18379.
Mendes de Leon, C. F., Barnes, L. F., Bienias, J. L., Skarupski, K. A., & Evans, D. A. 2005. Racial disparities in disability: Recent evidence from self-reported and performance-based disability measures in a population-based study of older adults. *Journal of Gerontology Series B, 60*(5), S263–S271.
U.S. National Center for Health Statistics. (2004). Nursing homes, 1977–99: What has changed, what has not? Facts from the National Nursing Home Survey. By F. Decker. Hyattsville, MD: U.S. National Center for Health Statistics.
U.S. National Center for Health Statistics. (2008). *National Nursing Home Survey and National Nursing Assistant Survey*. www.cdc.gov/nchs.
Wolf, D. A., Mendes de Leon, C. F., & Glass, T. A. (2007). Trends in rate of onset of and recovery from disability at older ages: 1982–1994. *Journal of Gerontology: Social Sciences, 62B*(1), S3–S10.
World Health Organization. (2004). *The world health report 2004: Changing history* (Annex Table 3). Geneva, Switzerland: World Health Organization.

Chapter 7
Demographic and Socioeconomic Group Differences in Morbidity and Mortality

This chapter is concerned with differences in the morbidity and mortality of major demographic, social, and economic groups, such as gender, race/ethnicity, and socioeconomic groups. I consider whether and to what extent health varies among males and females, whites, blacks, and Hispanics, marital status groups, income, education, and occupation groups, and groups distinguished by religious participation, geographic area within countries, and urban-rural residence. Group differences in the levels of morbidity and mortality, and in access to and in the quality of health care, have received considerable attention in recent years in the United States. It is the focus of recent reports of the National Research Council's Institute of Medicine (IOM), the National Research Council's Board of Health Sciences Policy (HSP), reports of the U.S. Agency for Healthcare Research and Quality (AHRQ), national monitoring efforts of the Centers for Disease Control and Prevention (CDC), university conferences,[1] and public policy debates. The reports of the U.S. Agency for Healthcare Research and Quality (U.S. AHRQ 2005a, b) include a broad set of performance measures that will be used to monitor the nation's progress in improving the quality of its health care system and in its access to health services, especially on the part of lagging groups such as some race/ethnic groups and lower income groups. Less attention has been given to some other socioeconomic-status groups (e.g., education; occupation), residence groups (e.g., urban vs. rural populations), types of households (e.g., family households vs. persons living alone), and marital-status groups (e.g., never-married vs. married persons).

[1] A university conference (Penn State University, June 2004) is the basis of the volume, *Health Inequalities Across the Life Course,* edited by S. H. Zarit, L.I. Pearlin, and J. Hendricks. It was published in the *Journals of Gerontology, Series B, Psychological Sciences and Social Sciences,* Special Issue II, Vol. 60B, October 2005. This volume gives a much more detailed picture of health inequalities among some groups than is presented here.

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_7, © Springer Science+Business Media B.V. 2012

Concepts of Group Differences

Before discussing the group differences themselves, we need to consider the concept of group differences and the methods of measuring them. The measurement of group differences is more complex than initially appears to be the case. Nominal group differences will vary depending on the concept of group differences applied (e.g., inequality, inequity), the population characteristic for which group differences are measured (e.g., education, income), the metric employed for measuring the differences (e.g., life expectancy, morbidity prevalence ratio), and the level of disaggregation of the population groups being analyzed or of the metric being applied (e.g., age, cause of death). The interpretation of the group differences will vary depending on these choices as well as the complexities introduced by such issues as the interaction of the independent variables (e.g., education and income, race and socioeconomic status) and the directness or indirectness of the causal path of any factor.

The terms differentials, disparities, inequalities, inequities and differences all appear in the writings on this subject. Loosely speaking, all five terms have the same meaning, but for technical use I think it is necessary to draw some distinctions among them. I prefer to limit the use of the term differentials to differences that have been subjected to statistical analysis and even significance testing.

Health disparity may have several meanings. In the U.S. program *Healthy People* 2010 (U.S. NCHS 2000), disparity encompasses the total difference between particular groups in the measures of morbidity and mortality. Disparity may also be viewed as the quantity that separates a group from a reference point on a particular measure of health expressed as a summary quantitative measure. Alternatively, disparity may be defined more narrowly as the difference that remains after taking into account patient needs and preferences and the availability of health care. This is the way the Institute of Medicine defined the term (National Research Council, IOM 2003). Some consider the term only in association with adverse health outcomes for which they hold the health providers responsible, while others associate the term with adverse health outcomes for which the individual is responsible. I think that Carter-Pokras and Baquet (2002) are on the right track in associating the term with injustice and unfairness, but I do not think that the latter are measured well by whether the difference is avoidable or immutable, as is suggested by Adler (2006). Adler defines disparity as,

> the extent to which individuals or segments of the population fail to achieve their highest potential state of health at a given age, given currently available medical treatments.

The practice of smoking or a suicide is usually considered avoidable but the individual has made his/her own choice, and hence the resulting disparity in health level is not unfair or unjust. The state of medical treatment, especially the degree to which genetic conditions are treatable, is always fluid, so that the "content" of disparity is steadily changing. Given that there are multiple definitions of the term

disparity and opinions differ on its appropriate usage, I suggest avoiding the term disparity in favor of some more neutral term in referring to the difference in health between groups.

The total difference in health status or mortality levels between two groups may be labeled health inequality. A formal definition of health inequality is the systematic difference in health between different demographic or socioeconomic groups within a society as reflected in recognized health indicators. The part of the total difference that may be viewed as indicating unfairness or injustice may be called inequity. Inequity may result from a difference in the quality of health care provided or in access to the health care available. It may also be viewed as including effectiveness (i.e., provision of care to all patients that could benefit), safety (i.e., care that avoids injuring patients), and timeliness (i.e., care that minimizes delay) (see U.S. AHRQ 2005a). In effect, inequality refers to total outcome while equity refers to unequal treatment affecting that outcome. The former includes the latter inasmuch as differences in outcomes may result in part from inequitable treatment.

It should be a matter of serious public concern when different groups in a population receive health care of variable quality. For example, while most of the difference in health status between young adults and aged adults is a result of age-related illness, some is a result of inequitable treatment. Equity in the provision of health care calls on health care systems to provide care that does not vary in quality as a result of such characteristics of the person as age, gender, race, ethnicity, socioeconomic status, marital status, and religion. The term discrimination is sometimes brought into such discussions to refer to a deliberate effort to deny some cognizable, or identifiable, group, because of its identity, the same quality of health care or degree of access to health care services, as other groups. The use of the term is intended to place responsibility for inequities in health care on some other group. Discrimination may be viewed as a special type of inequity.

Calling for *equality* in the provision of health care is quite different from calling for *equity* in the provision of health care. Groups vary in their risk of experiencing adverse health outcomes even if equity in the provision of health care is achieved. Groups such as age, sex, and race groups vary in the distribution of health risks and outcomes. Moreover, there is considerable heterogeneity of risks and outcomes within each of these groups. Heterogeneity of health status within groups is a natural characteristic of populations; homogeneity of health status of individuals within groups is largely a conceptual artifact assumed for statistical analysis, practical convenience, and planning purposes.

To achieve equality in health status between all groups is arguably an impossible goal. It is unreasonable to expect the health system to correct all the ills of a society, including those associated with variations in education, economic status, marital composition and household arrangements, and housing conditions. These are all associated with health status, however, as indicated below, and populations differ with respect to the distribution of these characteristics. Even countries with systems of universal health care experience socioeconomic differences in the health status of their populations. Analysts should therefore expect differences in health among

these groups. The goal of social measurement and its policy applications should be to quantify the extent of the group differences, identify the causes of the differences in terms of equity and other factors, and then suggest how much inequality is tolerable for the condition being measured. The first step in reducing inequality of health outcomes is to maximize equity in the availability of and access to health care. Maximizing equity calls for ethical guidelines, value-judgments, and setting of appropriate standards; on these even medical ethicists differ among themselves.

Some Measures of Group Differences

As suggested in Chap. 6, numerous measures of the health inequality among demographic and socioeconomic groups have been proposed. They vary in their design, their complexity, and the information they convey. More specifically, they vary with regard to the reference population from which the differences are measured, as to whether they are absolute or relative measures, and as to whether they are weighted on the basis of population size or take explicit account of population distributions. They vary as to whether they compare only two classes of the socioeconomic variable or take account of all classes in combination. Some are simple summary measures of the difference between the health status or frequency of a health event for different demographic and socioeconomic groups, and others are more complex summary measures designed to represent the degree of association between health status and various demographic and socioeconomic factors. The absolute or relative difference between the diabetes prevalence ratio for the two extreme educational levels illustrates the first type of measure, and the correlation between the diabetes prevalence ratio and the level of educational attainment illustrates the second type of measure. Since the choice of a summary measure of inequality may affect the interpretation of changes or differences in health status, it is advisable to apply more than one measure to derive the most valid and robust information regarding inequality.

Simple Comparisons

Absolute and Percent Difference Between all Classes and Highest Socioeconomic Class

One simple measure of inequality in the health levels of various socioeconomic groups is the absolute difference between the ratios or rates (e.g., diabetes prevalence ratio; mortality rate) for the highest socioeconomic class and the same measure for all classes combined. This absolute difference can be converted into a percent difference by dividing it by the rate or ratio for the highest socioeconomic class. The percent difference represents the degree to which the mortality rate or morbidity

ratio or rate for the "average" person differs from the corresponding measure for the highest socioeconomic class. Regidor (2004a) gives the following hypothetical examples, with mortality data for educational levels in 1990 and 2000:

	Mortality rate (per 100,000 persons)		
Educational level	1990	2000	Calculation of percent difference:
Total	134	72	For 1990: For 2000:
Third level	80	40	$\frac{134-80}{80} = 67\%$ $\frac{72-40}{40} = 80\%$
Secondary, upper level	110	58	
Secondary, lower level	128	80	
First level	160	91	

Depending on the year, mortality was 67% or 80% higher in the general population than in the highest educational class.

Absolute and Percent Difference Between Extreme Categories

The percentage difference between the highest and lowest categories in a socioeconomic hierarchy is a more common type of comparison. In this case, we take the percent difference between the rate or ratio (e.g., mortality rate; diabetes prevalence ratio) for the highest and the lowest socioeconomic classes, using the ratio or rate for the highest class as divisor. This measure shows the percent decrease that would have to occur in the death or disease rate in the lowest socioeconomic class to be the same as in the highest socioeconomic class.

This type of calculation has some important limitations that apply whether the difference is taken with the rate for the lowest socioeconomic class or the overall population. If it is interpreted as a measure of the relation of two variables, say mortality and educational level, it disregards the mortality rates for the intermediate educational levels. Next, this type of measure raises the question of the choice of base, whether to select the smallest or largest rate. The results may be very different depending on the choice of base. For example, using the data for 2000 above, we have either $91/40 = 2.28$ (a disadvantage of 128%) or $40/91 = .44$ (a disadvantage of 56%). Third, the result can be unstable if the mortality rate in the base, such as that for the highest educational level, is very low.

More Complex Comparisons

Dissimilarity, Gini, Theil, and Other Indexes

The several measures of inequality between countries described in Chap. 6, namely, the Gini coefficient (and its associated measures, the index of dissimilarity, Lorenz

curve), the Theil Index, the mean logarithmic deviation, and the squared coefficient of variation, are types of more complex measures that may be used to measure socioeconomic differences. They take all the health values in the socioeconomic distribution into account. They can be applied generally in the measurement of socioeconomic-group differences in health and mortality represented by such indicators as a disease prevalence ratio, life expectancy, a disability incidence rate, or any other interval-level health variable. The calculation of the Gini index, the Theil index, the mean logarithmic deviation, and the squared coefficient of variation is illustrated with data on the diabetes prevalence ratio for educational attainment groups in the United States in 2006 (Table 7.1). Although the data employed in the calculations are sparse, it is evident that these measures may yield different indications of the degree of inequality between the two variables being compared. In this case, all four measures show degrees of inequality between the segment of the population with diabetes and the segment free of diabetes, with respect educational attainment, of 90% or more.

The dissimilarity index can be applied in more than one way. It may be used to compare the distributions of two broad classes of a health indicator according to the classes of a socioeconomic variable (e.g., educational attainment),

Educational level	Good or better health	Poor or fair health	Abs. difference
Total	1.00	1.00	–
Less than 8th grade	.1	.4	.3
8th to 11th grades	.2	.3	.1
Completed high school to less than 3 years of college	.3	.2	.1
College graduate and higher	.4	.1	.3
Index of dissimilarity	1/2 sum of absolute differences = .4		

or to compare the geographic distributions of health status and socioeconomic status:

Census tract	High school graduates	Fair or poor health	Abs. diff.
Total	1.00	1.00	–
A	.22	.30	.08
B	.15	.20	.05
C	.18	.18	–
D	.20	.22	.02
E	.25	.10	.15
Index of dissimilarity	1/2 sum of absolute differences = .15		

The dissimilarity Index suffers from various limitations for intergroup comparisons. For example, the result is affected by the number of classes into which the variables are disaggregated. In addition, the measures may give the same values for populations whose health situations are radically different because they indicate merely the overall degree of dissimilarity. These measures can only be used if the socioeconomic and health variables can be ranked hierarchically.

Table 7.1 Illustrative calculations of several indexes of inequality between the diabetic population and the nondiabetic population with respect to educational attainment: United States, 2006

Educational attainment groups	Population (000's)			Prop. distribution = p_i[a]			Health measure (x_i)	
	All persons[b]	Diabetic	Non-diabetic	All persons	Diabetic	Non-diabetic	Diabetes ratio[c]	Non-diabetes ratio
Total	189,358	16,685	172,674	1.000	1.000	1.000	.0881	.9119
Less than high school diploma	3,750	4,191	27,559	.168	.251	.160	.1320	.8680
High school diploma or GED	54,586	5,475	49.111	.288	.328	.285	.1003	.8997
Some college	51,159	4,454	46,705	.270	.267	.270	.0871	.9129
Bachelor's degree or higher	51,863	2,565	49,299	.274	.154	.286	.0495	.9506

	$r_i = x_i / \sum x_i p_i$		Gini coefficient				Theil index			
	Diabetes ratio	Non- diabetes ratio	Diabetes ratio[d]		Non- diabetes ratio		Diabetes ratio[e]		Non-diabetes ratio	
Total	1.0000	1.000	—	.1723	—	.0167	—	.0525	—	.0005
Less than high school diploma	1.4983	.9519	.832	.2094	−.832	−.1331	.4043	.1018	−.0493	−.0079
High school diploma or GED	1.1385	.9866	.876	.1233	−.376	−.1068	.1297	−.0425	−.0135	−.0038
Some college	.9886	1.0011	−.182	−.0486	.182	.0492	−.0115	−.0031	.0011	.0003
Bachelor's degree or higher	.5619	1.0424	−.726	−.1118	.726	.2074	−.5764	−.0887	.0415	.0119

	Mean logarithmic deviation				Squared coefficient of variation			
	Diabetes ratio[f]		Non-diabetes ratio		Diabetes ratio[g]		Non-diabetes ratio	
Total	—	.0057	—	.0005	—	.0998	—	.0009
Less than high school diploma	−.4043	−.0679	.0493	.0083	.2483	.0417	.0023	.0004

(continued)

Table 7.1 (continued)

	Mean logarithmic deviation				Squared coefficient of variation			
	Diabetes ratio[f]		Non-diabetes ratio		Diabetes ratio[g]		Non-diabetes ratio	
High school diploma or GED	−.1297	−.0374	.0135	.0039	.0192	.0055	.0002	0000
Some college	.0115	.0031	−.0011	−.0003	.0001	.0000	0000	0000
Bachelor's degree or higher	.5764	.1579	−.0416	−.0114	.1919	.0526	.001	.0005

See the text for the formulas for the various measures
[a]Population 18 years and over
[b]Population weights
[c]Proportion of the population that is diabetic
[d]The first colum is $(q_i . - Q_i)$. The second is the final measure. q_i is the proportion of the total population in which the value of the diabetes ratio is lower than in category i and Q_i is the proportion of the total population in which the value of the diabetes ratio is higher than in category i
[e]The first column is lnr_i. The second is the final measure
[f]The first column is $ln(1/r_i)$. The second column is the final measure
[g]The first column is $(r_i - 1)^2$. The second column is the final measure

Regression Methods

To measure the extent to which a health variable is associated with socioeconomic status, some type of regression approach, linking the socioeconomic status (SES) and the health variable, may be applied. Regression analysis provides several indicators of the relation. Most important among them is the amount of change in health status or in a health event that is associated with a specified change in the level of a socioeconomic variable. Regression methods take account of all classes of the health variable. In using regression analysis to compare health differences between the categories of a SES variable (e.g., education), the health levels associated with the mean values for the lowest, highest, or other category of the SES variable may be determined from the regression equation and compared with one another.

This following section gives a brief description of regression methods.

Ordinary regression. The first form of regression to be considered is one where health status is assumed to vary linearly with the levels of a socioeconomic variable. Consider, for example, the relation between the sum of systolic and diastolic blood pressures and annual household income (for male householders 45–64 years of age). The general equation is,

$$\mathrm{Y} = \mathrm{a} - \mathrm{bX} + \mathrm{e} \tag{7.1}$$

where Y represents the sum of systolic and diastolic blood pressures, X represents annual household income in units of $10,000, and e represents an error term. The regression coefficient of the equation, b, and the Y-intercept, a, must be determined from observed data for individuals and households. The equation shows that blood pressure declines by b points for each $10,000 increase in annual household income. A regression equation (with hypothetical parameters) fitting these conditions would be,

$$\mathrm{Y} = 240 - 9\mathrm{x} + \mathrm{e}$$

where the blood pressure total is shown to decline by nine points for every increase of $10,000 in annual household income.[2]

The square of the correlation coefficient between the two variables, representing the degree of association between them, or the share of the total variation in Y (blood pressure) explained by the fitted line, may be viewed as one summary measure of the inequality in the health status of the population according to income. The correlation coefficient ranges from $-1\text{to}+1$, representing, respectively, a perfect negative association between the variables and a perfect positive association between them. The higher the absolute value of the correlation coefficient the greater the inequality in blood pressure according to household income. The measure's validity depends

[2]This linear relationship can be valid only over a limited range of the health variable since blood pressure can vary only over a very limited range before it is incompatible with survival. This suggests modifying the form of the equation, possibly by assuming that the slope of the curve declines at a declining rate with increasing household income or employing the reciprocal of household income as the independent variable.

on the assumption of linearity in the relation between the two variables and the specification of this one variable, household income, as accounting for the relation.

The analyst may wish to assume that more than one variable accounts for the variation in the independent variable (Y), say household wealth in addition to household income. The formula representing the relation between the variables may be extended to allow for several independent variables. Multivariate linear regression, as the new equation is called, is a statistical model of the form,

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots \beta_k x_k + \ldots. + e \tag{7.2}$$

where Y is a random variable whose value we want to "predict" in terms of the given values of the independent variables $x_1, x_2, \ldots x_k$. $\beta_1, \beta_2, \ldots \beta_k$ are regression coefficients associated with each independent variable and numerical constants that must be determined from actual data; β_0 is the Y-intercept, the value of Y when the regression coefficients are zero; and *e* is an error term for each observation. The regression coefficients are usually determined by the method of least squares.

Multivariate linear regression is a widely used statistical technique for describing the relationship between a dependent variable (e.g., health status or event) and several independent variables (e.g., demographic, socioeconomic, physical, and psychological variables), for assessing the strength of the relation between them, and for developing hypotheses as to whether there is a causal relationship between them.

Logistic regression. One limitation of ordinary univariate or multivariate linear regression is that, if the dependent variable is a dichotomous variable expressed as a percentage (e.g., percent self-reported as unhealthy, percent disabled, percent obese), the solution of the regression equation can yield estimated values of the dependent variable above 1.00 or below 0.0, that is, values outside an acceptable range for a percentage. To prevent such a result, a variant of the multiple linear regression equation, the logistic regression model, is usually employed. To apply this form of regression, the dependent variable must be transformed to logits, or the natural logarithms of odds ratios, a continuous variable that can take on any value in the range from $-\infty$ to $+\infty$. The logit of the proportion P_i is,

$$\text{logit}\, P_i = \ln[P_i \div (1 - P_i)] \tag{7.3}$$

where P_i is the probability that the *ith* person is in the category of interest (e.g., disabled) and $(1 - P_i)$ is the probability that he or she is in the opposite category (e.g., not disabled). If the linear part of the multivariate regression equation, i.e., the part with the independent variables and their coefficients, is exponentiated, the regression equation takes the form,

$$P_i \div (1 - P_i) = e^{b0+b1xil+b2xi2+\cdots\cdots+bkxik+\cdots\cdots+e} \tag{7.4a}$$

The operating algorithm is,

$$\ln[P_i \div (1 - P_i)] = b_0 + b_1x_1 + b_2x_2 + \cdots b_kx_k + \cdots + e \tag{7.4b}$$

given a particular vector, or array, of data on k predictor variables. A unit change in any independent variable x_i results in a change of magnitude b_i in the logit of P_i, holding constant the other independent variables.

In survival analysis a somewhat similar form is used, but one or more of the terms of the regression equation contains a time-varying element. In this case the natural logarithm of the hazard ratio for the health variable is expressed as a linear function of the explanatory variables. Here is one such regression equation:

$$\ln h(t) = b_0(t) + b_1x_1 + b_2x_2(t) + e \tag{7.5}$$

where $b_0(t)$ is any function of time, the variable x_1 is constant with respect to time, and the variable x_2 changes over time. This is the so-called proportional hazards model.

Poisson regression. In the cases above, the dependent variable is either numerical and normally distributed, or binary. One also encounters cases where the dependent variable is numerical but is a count of very uncommon events, such as the number of new cases of emphysema occurring in a particular year. Note that the counts are all positive. In this case the Poisson distribution is used to express the relation of the variables. The natural logarithm of the dependent variable (e.g., number of new cases of the disease) varies as a linear function of the explanatory variables:

$$\ln Y = b_0 + b_1x_1 + b_2x_2 + \cdots\cdots + b_kx_k + e \tag{7.6a}$$

$$\text{or } Y = (e^{b0})(e^{b1x1})(e^{b2x2})\cdots\cdots(e^{bkxk})(e^{e}) \tag{7.6b}$$

Here it is assumed that, like the Poisson distribution, at each level of the variables, the number of new cases has a variance equal to the mean.

For further discussion of these types of multivariate regression, and for computer programs for applying them, refer to any of the standard statistics texts on regression.

Slope index of inequality and relative slope index of inequality. In connection with a review of measures of health inequality for use in NCHS Health 2000 Goals, Pamuk et al. (1993) proposed two alternatives to existing measures, namely, the slope index of inequality (SII) and the relative slope index of inequality (RII). They defined the slope index of inequality as the slope of a least-squares regression line fitted to data on the relation of the socioeconomic variable and the health variable. The relative slope index of inequality is the ratio of the slope index of inequality to the rate for the health variable over all classes. If the coefficient b represents the slope in the

regression equation, $Y = a + bX$, then the relative slope index of inequality is represented by $b \div R$, where R is the overall rate of the health variable.

Alternative definitions of the SII and the RII are also employed, particularly by European analysts of health inequality (Mackenbach et al. 2008). These are also regression-based measures that take into account the whole socioeconomic distribution. The form of the regression is Poisson. As before, the regression relates a measure of mortality or morbidity to a measure of educational attainment, income status, or occupation. Each class figure is calculated as the mean proportion of the population having the specified level of education, income, or occupation. The relative slope index of inequality is the ratio between the estimated mortality or morbidity rate or ratio for persons of the lowest socioeconomic class to that of persons at the highest socioeconomic class. The slope index of mortality-inequality measures absolute differences in rates between the lowest socioeconomic class and the highest socioeconomic class, as derived from the Poisson regression. It is calculated from the relative slope index of mortality-inequality and the overall mortality rate as follows (Mackenbach et al. 2008):

$$\mathrm{SII} = 2 * \mathrm{m} * (\mathrm{RII} - 1) \div (\mathrm{RII} + 1) \tag{7.7}$$

where the mortality rate (m) is expressed per 100,000. Here is an evaluation of the formula with illustrative numbers,

$$640 = 2 * 1600 * (1.50 - 1) \div (1.50 + 1)$$

where the mortality rate is 1600 and the relative index of inequality is 1.50)

Population Attributable to Risk

The population attributable to risk (PAR) is an index of disease frequency often used in epidemiology. It is a measure of the total number or proportion of cases of death or disease that would be eliminated if persons in each socioeconomic class had the death or disease frequency of the highest socioeconomic class. With the proportion, we can state the percent decrease in the mortality or morbidity rate that would result from equalizing the rate for all classes at the level of the highest socioeconomic class. The PAR takes account of population size, so that if the groups with the highest mortality or morbidity rates are large in size, the reduction in the rate would tend to be large. PAR is derived by calculating a PAR value for each socioeconomic class and summing the results across all classes:

$$\mathrm{PAR}\,\% = \Sigma \frac{\mathrm{P_c}\,(\mathrm{RR_c} - 1)}{1 + \mathrm{P_c}\,(\mathrm{RR_c} - 1)} \tag{7.8}$$

where P_c is the proportion of the total population in the socioeconomic class and RR_c is the relative risk for the class compared to the risk for the highest socioeconomic class. To illustrate the evaluation of this measure, assume that there are three classes in a particular socioeconomic group, with population proportions .10, .50, and .40, and relative morbidity ratios relating the lowest socioeconomic class rate to the highest socioeconomic class, 1.0, 1.4, and 1.5:

$$\text{PAR\%} = 33.3\% = \left[\frac{.10(1.0-1.0)}{1+.10(0)} + \frac{.50(1.4-1)}{1+.50(1.4-1)} + \frac{.40(1.5-1)}{1+.40(1.5-1)}\right] * 100$$

Demographic and Socioeconomic Explanatory Variables

In measuring associations, it is useful to distinguish compositional/structural factors from causative/etiological factors among the explanatory variables, but this is no straightforward task. Some factors can be assigned to either class depending on the nature of the study. Structural factors in one analysis may become etiological factors in another analysis. Age, sex, race and ethnic origin, geographic area, and urban-rural status are usually included among the compositional/structural factors. Factors like religion, marital status, labor force status, socioeconomic status, and housing conditions are commonly viewed as "causative." Socioeconomic status (SES) is defined variously in terms of education, income, wealth, occupation, or some combination of them. Socioeconomic status and the other "causative" characteristics have been found to be strongly associated with the level of health and health care. This association is discussed in more detail below.

The corresponding "change" factors, such as migration, marriage, divorce, widowing, graduating high school, becoming poor, becoming unemployed, and becoming a homeowner, are typically viewed as causative factors as well. These factors can be further disaggregated in various ways. For example, marriages can be disaggregated in terms of the order of the marriage for each spouse or the ages of the spouses.

In further exposition of these distinctions, Exhibit 7.1 sets forth a list of factors that often appear in demographic and socioeconomic analyses of health differences. In addition to these factors, numerous other factors can be enlisted to explain specific health changes. Among these are the passage of a law or change in public policy (e.g., passage of Medicare legislation, increase in resources for health services, establishment of neighborhood mental health or family planning clinics, change in abortion policy or law), environmental change (e.g., hurricane, forest fire), change in lifestyle practices (e.g., less smoking and use of alcohol, more physical activity), and a change in the economy (e.g., rise in unemployment, rise in cost of living).

Structural	Status/causative	Change/causative	Disaggregating
Age	Marital status	Migration	Age of spouses
Sex	Education	Becoming a parent	Duration of marriage
Race	Labor force status	Death	Spacing of children
Ethnicity	Occupation	Marriage	Order of marriage
Urban-rural residence	Religion	Divorce	Age of children
Geographic location	Income	Widowing	Number of children
	Wealth	Entry into labor force	
	Household structure	Exit from labor force	
		Becoming unemployed	
		Enrolling in school	
		Graduating from school	
		Change in household structure	

Exhibit 7.1 Some explanatory demographic and socioeconomic factors employed in health analysis

Selection vs. Causation

Even after we have separated the compositional factors in health from the etiological factors, an important analytic question remains – the omnipresent selection-vs.-causation issue. One of the most difficult and persistent issues to be resolved in social science research on health outcomes is whether an apparent etiological factor is truly causative of a given health outcome or whether the factor is simply selective of those who have good or bad health. A factor can be both selective and causative, although in a particular association it tends to be primarily one or the other. For example, are persons who marry more healthy than those who do not, or does married life contribute to better health? A reasonable hypothesis is that both propositions are true in part. It is the task of the researcher to quantify the partial contributions of marital status to health. In any case a demonstration of association should not be treated as a demonstration of causation and the research protocol must therefore be designed to distinguish these relations.

Longitudinal studies are far superior to cross-sectional studies in dealing with the causation/selection issue because they can track the status of the respondents before and after the events of interest (e.g., marriage and health changes; graduating high school and health changes).

By documenting the time sequence of events, a carefully designed longitudinal study can eliminate the role of the later event as a cause of an earlier one. While such a design cannot prove cause and effect, it can increase the likelihood of a correct interpretation. It also reduces recall bias. On the other hand, longitudinal analysis is much more costly than cross-sectional analysis and is complicated by the fact that the various events can be repeated, statuses of individuals can change over the life course, and individuals can drop out of the sample, with resulting changes in the composition of the respondent population.

Further Complexities of Causation

The analysis of causation has further complications. A given extrinsic factor may be a cause of a given disease, but some people who have the disease may not have been exposed to this risk factor, and some people who do not have the disease have been heavily exposed to the risk factor, as in the case of smoking and lung cancer. A particular factor may operate as a positive influence on health at first but may become a negative influence after long-term exposure or overindulgence. Exercise may be such a factor.

The chain of causation may be extremely complex and factors sometimes operate directly and at other times indirectly as causal agents. One disease may be a risk factor for another disease, such as diabetes for heart disease, which in turn becomes a risk factor for a third disease, such as heart disease for kidney disease. Some risk factors are associated with more than one disease, such as obesity for both heart disease and diabetes, and some diseases may require the presence of more than one risk factor. Some agents may have a direct effect in causing a disease, as a particular virus' role in a particular infectious disease. On the other hand, they may merely add to the susceptibility of the person to a disease by compromising his/her immune system. Some established causal relations may have to be altered when the recommendations of earlier research studies are found by later research studies to be too general, as, for example, the shift in the recommendations regarding use of hormone replacement therapy for postmenopausal women. An NIH panel recommended in 2005 that menopause should not be treated as a medical condition, although hormone therapy is effective in "treating" some unpleasant "symptoms" of menopause.

In recent years social scientists have been questioning the methods used in reaching conclusions of cause and effect. These conclusions are often drawn on the basis of regression analysis and structural equation models. According to Moffitt (2005), typical practice currently ignores some key problems associated with cause-effect analysis or handles them inadequately. He refers to such problems as: The internal validity (i.e., exogeneity) of the independent variables; the external validity (i.e., generalizability to wider populations) of the associations; incorrectly specified mechanisms because of error or unresolved complexity, such as error in the choice of the causal factor(s), the large number of causal mechanisms for a single outcome, and the large number of effects of the single posited causal mechanism; the existence of reverse causation; extrapolation of the results (i.e., applications to data points outside the sample actually used in the analysis); and ecological correlation (i.e., applying the results derived from area units to individuals). Moffitt (2005) writes, "The question of whether there exist single, generalizable effects that are independent of the particular causes that induce them is a deep question, not easily resolved." To manage the problem and to achieve creditable cause-effect conclusions, he suggests weighing evidence from different studies with different strengths and weaknesses inasmuch as single studies cannot provide the best estimates.

Demographic and Socioeconomic Characteristics

Gender Differences

Gender is another biological factor like age that receives frequent attention as a risk factor. The risks for death and disease differ for the sexes but the direction of the differences between the sexes is largely in opposite directions. In general, men have a higher risk of death but a lower risk of experiencing a chronic degenerative nonlethal disease than women.

Mortality

The death rates of males are higher than those of females at every age of the life cycle in nearly all countries of the world. The current exceptions may be a few South Central Asian countries (e.g., Afghanistan, Bangladesh, and Nepal) at some ages. A quarter century ago the list of exceptions was longer and included several other countries in that region of the world, for example, India, Pakistan, Iran, and Iraq. The female advantage in life expectancy in 2005 ranged up to more than 10 years in some countries of Eastern Europe (e.g., Belarus, Russia, Ukraine) and Northern Europe (e.g., Estonia, Latvia, Lithuania). (See Population Reference Bureau 2005). The countries of Southeast Asia and South Central Asia show the smallest differences. In Japan, the country with the highest life expectancy in 2005 at 82 years, the difference between the sexes was 7 years. This is well above the world average of 4 years and similar to the difference for the More Developed Countries (MDC) as a whole (7 years). Male and female life expectancies for the MDC and LDC as a whole and world regions are shown in Table 7.2.

In the United States, as in the other industrialized countries, the survival curve for females is consistently above that for males over the age span and the level of life expectancy for females is much higher than for males (Figs. 7.1 and 7.2). Underlying this overall difference is an excess of male mortality over female mortality at every age. In 2005 the male-female gap in life expectancy at birth in the United States was 5.2 years (Table 7.3). The difference at age 65 was 2.8 years, so that a large part of the male-female difference is accounted for by differences in mortality at ages over 65. The gender difference in life expectancy at birth grew gradually from its first recorded level of 2.8 years in 1900–2002 (ODRS) to a peak of 7.8 years in 1975 and 1979, but since that time the advantage for females has been gradually diminishing (Fig. 7.3). A difference exceeding 5 years has prevailed in all the years since 1946, however. Clearly males and females have not shared equally in the reduction of mortality in the last century. During most of the century life expectancy at birth was rising faster for females than for males, but after about 1979 males moved ahead more rapidly. The patterns of male-female changes in life expectancy below age 65 and above age 65 were similar to those of life expectancy at birth (Table 7.3).

Table 7.2 Estimates of male and female life expectancy at birth, for world regions: 2005

Region	Years		
	Male	Female	Difference
World	65	69	4
More developed countries	73	80	7
Less developed countries	63	67	4
Less developed countries, excl. China	61	64	3
Africa	51	53	2
Northern Africa	66	70	4
Sub-Saharan Africa	47	49	2
Northern America	75	80	5
Latin America /Caribbean	69	75	6
Asia	66	69	3
Asia excl. China	64	67	3
China	70	74	4
Europe	71	79	8
Eastern Europe	63	74	11
Oceania	73	77	4

Source: Population Reference Bureau (2005). Reprinted with permission of the Population Reference Bureau

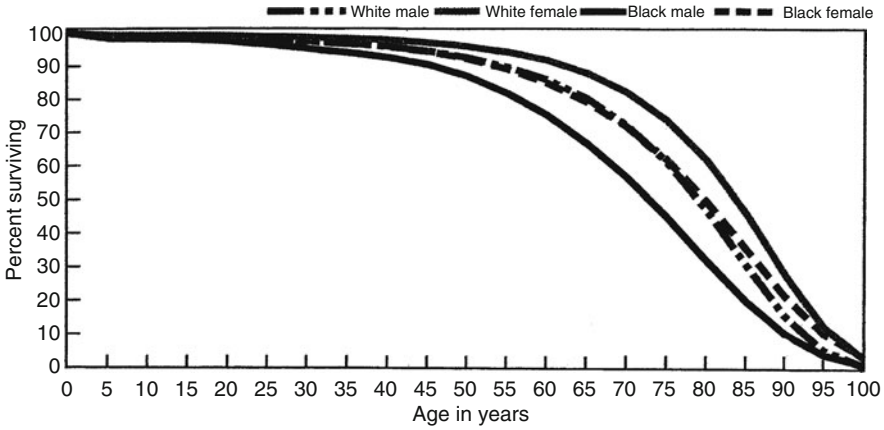


Fig. 7.1 Life table survival curves for sex-race groups: United States, 2004 (Source: U.S. National Center for Health Statistics (2007))

Relative role of genetic and nongenetic factors. There is considerable interest in and debate about the relative role of genetic and nongenetic factors in the difference in life expectancy between the sexes. There are several reasons for this interest. One is that, if the sex difference in longevity has a strong biological basis, it may be expected to persist; nongenetic influences are more amenable to change. The basis of the difference affects decisions regarding the mortality assumptions for population projections and regarding the premiums and benefits for insurance and retirement

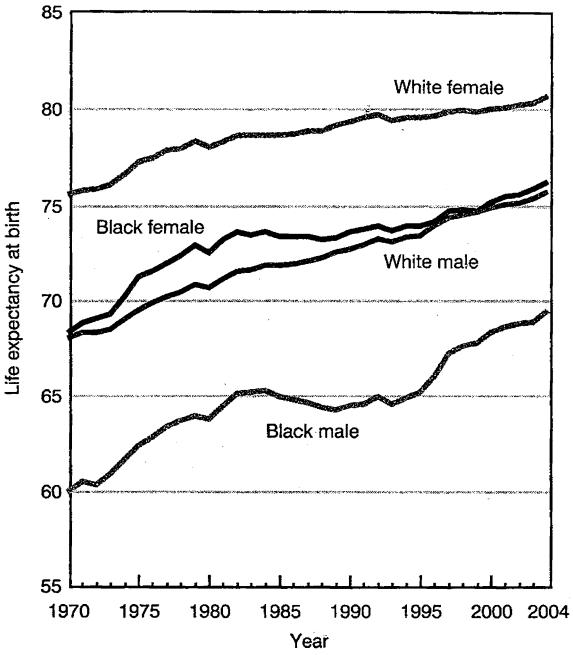


Fig. 7.2 Life expectancy at birth by sex and race: 1970–2004 (Source: U.S. National Center for Health Statistics (2007))

programs. It affects the design of public health programs and clinical trials, and treatment protocols in public health and clinical medicine. In spite of the importance of the issue and interest in it, the evidence regarding the relative roles of genetic and nongenetic factors in the female longevity advantage is inconclusive and can be used, selectively, to support both sides of the argument.

A strong case can be made for the view that the longevity advantage of females over males has a substantial genetic component (Waldron 1976, 1983; Holliday 1997:67–68). Holliday hypothesizes that there are sex-linked genes favoring female longevity over male longevity; these gene differences are a direct consequence of the genetic mechanism of sex determination. The female advantage applies to many of the leading causes of death, particularly the endogenous/intrinsic ones. A biological element in mortality can be inferred from the excess of male late-fetal losses and male infant deaths. The practices of female infanticide and of rearing male infants with greater care than female infants in some Asian countries would, however, contribute to an excess of female infant mortality and an understatement of any excess of male infant mortality, possibly even concealing it.

Several physiological responses appear to favor women. The hormonal conditions during the female reproductive period are clearly protective for survival. This is suggested by the 10-year lag in the prevalence of heart disease for women as compared with men. Heart disease does not become an important health condition for women until after the menopause. Premenopausal and postmenopausal women

Table 7.3 Sex and race differences in expectation of life in years and survival rates for selected age segments, for the Death Registration States and the United States: 1900–2002 to 2005 (Death registration states, 1900–1902 and 1919–1921; United States, 1939–1941 to 2005)

	Life expectation						Survival rate			
	Sex			Race			Sex		Race	
	Male	Female	Diff	Black	White	Diff	Male	Female	Black	White
Year	At birth						Birth to age 65[a]			
1900–1902	47.9	50.7	2.8	33.8	49.6	15.8	.387	.432	.205	.415
1919–1921	55.5	57.4	1.9	47.0	57.4	10.4	.492	.521	.327	.524
1939–1941	61.6	65.9	4.3	53.8[b]	64.9	11.1	.558	.655	.378[b]	.632
1959–1961	66.8	73.2	6.4	63.9	70.7	6.8	.642	.785	.560	.731
1979–1981	70.1	77.6	7.5	68.5	74.5	6.0	.706	.835	.644	.786
2005	75.2	80.4	5.2	73.2	78.3	5.1	.792[d]	.870[d]	.733[d]	.841[d]
	At age 65						Age 65 to age 85[c]			
1900–1902	11.5	12.2	0.7	10.9	11.9	1.0	.134	.163	.136	.149
1919–1921	12.2	12.7	0.5	12.2	12.5	0.3	.161	.183	.177	.173
1939–1941	12.1	13.6	1.5	13.0[b]	12.8	−0.2	.156	.213	.220[b]	.182
1959–1961	13.0	15.8	2.8	14.0	14.4	0.4	.200	.322	.274	.260
1979–1981	14.2	18.4	4.2	15.4	16.6	1.2	.255	.452	.310	.366
2005	17.2	20.0	2.8	17.2	18.8	1.6	.387[d]	.522d	.386[d]	.464[d]

Source: Based on U.S. National Center for Health Statistics (2008) U. S. decennial life tables for 1999–2001, United States life tables. *National Vital Statistics Reports*, *57*(1); U.S. National Center for Health Statistics (2007)

[a]Survival from birth to exact age 65

[b]For the entire nonwhite population

[c]Survival from exact age 65 to exact age 85

[d]For 2004

differ in their neuroendocrine responses to behavioral stressors. In general, however, women under stress secrete less of the stress hormones; that is, their neuroendocrine regulation is more favorable to their health. Hence, the destructive bodily changes that result from stress are less pronounced for women (Ramey 1982; Frankenhaueser et al. 1978; Moen et al. 1989). Females have a more favorable response than men with respect to cholesterol metabolism, the clotting factor, the elasticity and proneness to injury of the vascular lining, and the immune function, in addition to the more favorable hormonal balance and response to stress.

Oxidative stress and inflammatory processes are central elements in certain theories of aging. The female hormone, estrogen, is a powerful antioxidant in most tissues (excluding breast tissue) and an anti-inflammatory agent. According to Aviv et al. (2005), women's augmented antioxidative and anti-inflammatory capacity bespeaks a greater somatic fitness for them than for men. This superior fitness is expressed in greater leukocyte telomere length for women than men.[3] Aviv et al. also

[3]Telomeres are small caps at the ends of each chromosome and leukocyte telomeres are those in the white blood cells. Each time a cell replicates, the telomeres get slightly shorter. Chromosomes are the ordered collections of genes located in each cell; there are 46 of them in each cell.

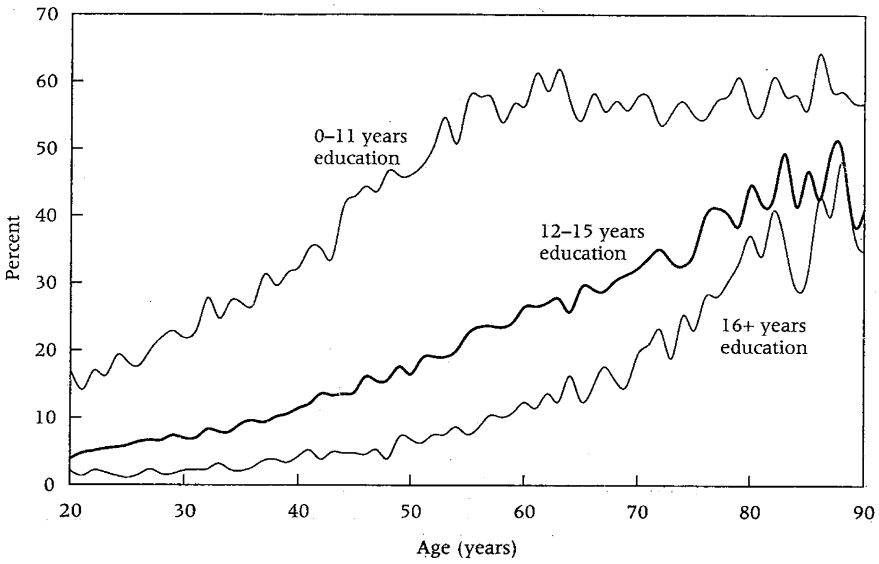


Fig. 7.3 Percent of population 20–90 years old reporting fair or poor health according to education: United States, 1991–1996 (Source: U.S. National Institute on Aging (1999); From Smith (2004), Figure 3, p. 127. Copyright © John Wiley & Sons. Reprinted with permission. Primary source: National Health Interview Surveys, 1991–1996)

maintain that the longer life span of women than males may originate from innate differences in somatic cell profiles between the two sexes and that, with advancing age the cells with superior somatic fitness within each woman, and presumably longer telomeres, are more likely to survive.

Generally, among mammals in captivity the males of the species have a shorter life span than the females of the species. The disadvantage of males has well established evolutionary (i.e., biological) origins, even though, over extended periods of geologic time, environmental factors are heavily involved in evolutionary changes. Higher male than female mortality during the prime reproductive ages is typical of sexually reproducing animals of many vertebrate and invertebrate species (Carey and Judge 2001). Under contemporary conditions, male reproductive competition selects men for riskier behavior and accounts for most deaths due to accidents and homicides during early adulthood (Carey 2002). Carey maintains that, as a result of differences in both biology and behavior, the mortality response of males and females will differ even in similar environments.

A study of the mortality of men and women in Catholic teaching orders, based on the premise that the living conditions of the sexes are essentially equal in a monastic milieu, tends to support the biological hypothesis (Madigan 1957). The assumption of this study with regard to similarity of living conditions for the sexes has been challenged, however.

An alternative hypothesis maintains that the tendency for women to live longer than men results largely from differences in the behavior and lifestyles of men and

women (Waldron 1976; Waldron and Johnston 1976). These factors can account for an important part of the male-female difference in longevity either directly or indirectly by their interaction with genetic or biological factors. Generally, males are engaged in the more stressful, physically demanding, and hazardous occupations. On the other hand, in Russia, where there is less differentiation in the occupational roles of men and women than in the United States, there is a far greater gap in life expectancy at birth in favor of females than in the United States; the figures in 2005 are 13 years vs. 5 years (Population Reference Bureau 2005).

Women are more likely to secure an earlier diagnosis and appropriate treatment for a health condition, including particularly a serious illness. Women are less likely to be self-destructive (suicide, alcoholism), to engage in acts of violence against others (homicide), or to be excessive risk takers (auto accidents). Cigarette smoking has been a more common habit of men than women and has been indicted as a leading contributor to the sex difference in longevity (Retherford 1975). On the other hand, smoking affects women's health differently from men's health. Women tolerate exposure to cigarette smoke more poorly than men and accordingly suffer the adverse health consequences of smoke ingestion more quickly and more severely than men. The rise in smoking among women and its decline among men may explain the recent divergence of trends in lung cancer for the sexes.

On the basis of a study of cloistered populations, Luy (2003) concluded that a combination of biological and non-biological factors is responsible for a small excess of male mortality in these populations. He maintains that a large part of the male excess in mortality in the general population is accounted for by the harmful lifestyle of blue-collar men.

Behavioral factors are certain to be the major cause of any considerable changes in the sex difference in mortality, such as occurred in the last century in the United States. Biological factors cannot explain such rapid developments. Rapid changes in the male-female mortality gap are largely influenced by changes in the male-female patterns of responding to environmental (e.g., work and home), lifestyle (e.g., diet, exercise, smoking, sleep), and behavioral (e.g., driving and drinking) risk factors and in sex differences in access to and use of health resources. The divergence of the mortality of the sexes in the past century may have been due in substantial part to the fact that a much greater share of men than women smoked, while an important factor in the partial convergence of the mortality of the sexes in the last few decades may have been the increase in the share of women who smoke and the decrease in the share of men who do. Alcohol consumption may have played a similar, if lesser, role.

The influence of holding more hazardous jobs may be decreasing for men, as women are employed in larger numbers at such jobs and blue-collar jobs for men are diminishing. Both men and women are benefiting with lower death rates from the improved access to medical care, but men may be benefiting more because of their initially higher death rates (Waldron 1993). HIV/AIDS has had a different impact on the two sexes historically. Initially men were almost exclusively affected by AIDS, but this has been changing and now deaths to women make up one-fourth of the total.

Incidence of deaths from heart disease and cancer. A comparison of the age difference in the incidence of heart disease and cancer mortality among men and women suggests the different role of biological forces for the two sexes in affecting mortality. In the United States heart disease and cancer figure among the 10 leading causes of death for each sex at every age, and from age 45 on they occupy the first, second, or third ranks among the causes of death (U.S. NCHS/Heron and Tejada-Vera 2009). For ages 35–74, cancer holds rank one and heart disease holds rank two or three among women and then their ranks reverse themselves. For most ages from ages 45 on, heart disease is the number one cause for men, outranking cancer.

Over all ages combined, heart disease deaths of women outnumber cancer deaths of women by nearly 25%; and cardiovascular deaths as a group outnumber gynecological cancers by about 7–1. There is clearly a disproportionate impact of heart disease and cancer on the two sexes even though the overall death rates from heart disease and cancer are roughly equal for the sexes. Men are afflicted by heart disease much earlier in life than women; women must wait until they are elderly for it to become number one in rank. Women's reproductive hormones protect them against heart disease while those of men do not. After the female menopause the difference between the sexes in the role of these two causes of death diminishes steadily until heart disease becomes the dominant cause of death for both sexes at the advanced ages.

Projections of the male-female difference in life expectancy. Whether genetic or non-genetic factors are dominant in explaining the male-female difference in longevity – and the evidence to resolve this question is inconclusive – a substantial difference is expected to continue because of the expectation that male-female differences in lifestyle and personal behavior will persist. Several leading analysts posit a continuation of the longevity gap in the United States. Actuaries at the Social Security Administration (U.S. SSA 2008), demographers at the U.S. Census Bureau (2009), Lee and Tuljapurkur (1994), and Ahlburg and Vaupel (1990) all project a continuation of a substantial difference in the life expectancy of the sexes. The SSA projections show a difference in the life expectancy at birth of the sexes of 3.4 years in the year 2050, the Census Bureau series show a difference of 4.3 years for that year, the Lee-Tuljapurkar projections show a difference of 5 years for 2065, and the Ahlburg-Vaupel series show a difference of 5 years for 2080. These figures generally vary little from the observed difference in 2005 (5.2 years).

Morbidity

Women experience much more nonlethal illness than men, and also more lethal illness than men for several conditions. Some chronic conditions of later life are particularly characteristic of women. These include musculoskeletal diseases (e.g., osteoporosis), autoimmune diseases (e.g., arthritis), and urinary incontinence. Type II diabetes, lupus (another autoimmune disease), and irritable bowel syndrome (IBS) are also much more prevalent among women than men, as are depression and

chronic fatigue syndrome. Limitations of activity associated with ADL and IADL disability are about as prevalent among men as women, but at ages 65 years and over, the ages when these conditions are most common, women are affected two-thirds more frequently than men (U.S. NCHS 2006). On the other hand, the incidence and prevalence of HIV/AIDS are much greater among men than women. Women make up only 27% of the new infections (U.S. NCHS 2006) although they develop AIDS with lower viral loads than men. The age-adjusted cancer incidence rate for all sites combined is one-third greater for males than for females.

For many women, working outside the home has the effect of reducing stress and improving health (Verbrugge and Madans 1985). The increasing labor force participation of women and the approximation of the lifestyles of men and women since World War II do not appear to have prejudiced the health and longevity of women greatly, although the smoking practices of women apparently account for a part of the recent convergence of male-female mortality rates. According to Verbrugge and Madans, having the triple social roles of employee, spouse, and parent is associated with better physical health for American women. They report that, despite added pressures on women with multiple roles, increasing labor force participation has had a positive impact on women's health. On the other hand, data from the National Health Interview Survey from the mid-1960s and the late 1970s suggest a pronounced decline in the health condition of women with few roles. Women who work in low-skilled jobs, who have many children, and whose spouses exercise a patriarchal control over the family have the poorest health record.

Differences Between the Races and Hispanics/non-Hispanics

Sources and Quality of Data

Data on the health status of racial groups and persons of Hispanic origin in the United States are available from several sources: National sample surveys, censuses, Medicare claims records, Medicare enrollments, Medicaid records, the State Children's Health Insurance Program, vital registration, and disease registries. The data on race and Hispanic origin available in these sources are not without a number of limitations for purposes of analysis. A leading issue is the correspondence of the data on race and ethnicity with the reality distinguishing the races and ethnic groups, particularly given the considerable degree of race mixture. Moreover, the national surveys do not provide the data needed for analysis of several specific racial or ethnic groups, only the data for broad racial/ethnic groups, because of the small size of these populations in the surveys and the resulting sampling errors. The Medicare data on race and Hispanic ethnicity are incomplete, insufficiently detailed, and probably often inaccurate. The data on race and Hispanic ethnicity in the other public records are not collected in a standardized form. The data in private sources – health insurance plans, hospitals, and medical care practices – do not usually contain information on race or ethnicity or the race/ethnic data are not recorded uniformly.

The National Academy of Sciences (NAS) considered the issue of the limitations of the health data for the races in a study project designed to identify and interpret differences in health and health care among the races and suggest ways of reducing them. In its report, *Eliminating Health Disparities – Measurement and Data Needs,* the NAS (2004) recommended (1) the use of data linkage to maximize the information provided by the existing record systems, (2) the repair of the shortcomings in the designs of sample surveys so as to allow for analysis of specific races and ethnic groups, and (3) the tabulation of the data on race and ethnicity by socioeconomic class.

Official Race and Ethnic Classification in the United States

Race can be measured in several different ways, such as by direct observation, self-report, administrative records data, field tests, and anthropometric tests. Censuses and sample surveys have historically employed direct observation (i.e., an enumerator makes the determination) but more recently self-reporting (i.e., the individual reports his or her own race) has been used. Today self-reporting is the preferred procedure for securing the data in censuses and surveys.

Moreover, if the reports are secured under U.S. government auspices, they must fit into a classification system determined by federal authorities. Race/Ethnicity is now self-reported according to common official guidelines in the decennial census, the surveys conducted or sponsored by the U.S. Census Bureau, the National Center for Health Statistics, the National Institute on Aging, and other federal agencies, and vital statistics records. Specific racial/ethnic categories have been formally established and promulgated by the U.S. government for the reporting of race. These categories are not consistent with academic or scientific (i.e., anthropologists) concepts of race[4] or with the concepts of race held by many private individuals and groups. They are designed to serve the needs of public programs and reflect the view that race is a subjective social construction that may change over time and differ from place to place.

The U.S. government's standards for collecting and tabulating data on race and Hispanic origin determine the type of data generally available for measuring racial and Hispanic/non-Hispanic differences in mortality and morbidity in the United States. Until 1997 the classification used for the collection of data on race and Hispanic origin was governed by the Office of Management and Budget's (OMB) Directive 15. In 1997 OMB promulgated new standards for the classification,

[4]The "scientific" concept of race generally recognizes three major racial groups, with a few smaller groups not formally classified: Caucasoid ("white"), Negroid ("black"), and Mongoloid (American Indians and most Asians and Pacific Islanders). Hawaiian is one of the unclassified groups as is the Veddah of Sri Lanka. The races are distinguished by a battery of physical traits, of which skin color is secondary. For example, most Middle Easterners and most Asian Indians are Caucasoid. It is recognized that the considerable racial mixture is blurring the identity of the races and complicating the task of distinguishing the prototypic races.

collection, and tabulation of such data in Revised Directive 15. The new standards were implemented in the 2000 census. According to this directive, in effect five races – white, black, American Indian, Asian, and Pacific Islander – and two ethnic groups – Hispanic and non-Hispanic – with national subgroups for the races and Hispanics, are recognized. The census questions are:

Item 5. Is person Spanish/Hispanic/Latino?	
No, not Spanish/Hispanic/Latino	Yes, Puerto Rican
Yes, Mexican, Mexican Amer., Chicano	Yes, Cuban
Yes, other Spanish/Hispanic/Latino	[Print group]

Item 6. What is person's race? Mark one or more races to indicate what this person considers himself/herself to be.	
White	
Black, African American, or Negro	
American Indian or Alaska native – [Print name of enrolled or principal tribe]	
Asian Indian	Native Hawaiian
Chinese	Guamanian or Chamorro
Filipino	Samoan
Japanese	Other Pacific Islander – [Print race]
Korean	
Vietnamese	
Other Asian – [Print race]	
Some other race – [Print race]	

In comparison with the old standards, the 1997 standards made two important changes. It subdivided one of the racial categories (Asian and Pacific Islander) into two categories and allowed respondents to choose more than one race.

Calculating vital rates from the U.S. vital registration system requires population data from the Census Bureau for the denominators and data on vital events from state vital statistics offices or NCHS for the numerators. Inasmuch as the 2000 census adopted the 1997 standards on race and Hispanic ethnicity but the state vital statistics offices had not generally adopted them, there is a lack of compatibility between the race classifications used for the vital records and that used for the population counts. Moreover, the data in both sets of records or the aggregate data had to be modified to permit tabulations in accordance with the single-race concept of previous years as well as the calculation of valid race-specific rates. To bridge this gap and to assure historical comparability and comparability between the mortality data and the population data required the modification of the data from one or both of the data sources. Accordingly, NCHS and the Census Bureau developed a bridging mechanism for the race data prepared by each agency that would achieve these ends.

General Health Differences Between Race/Ethnic Groups

Even though in recent decades the overall health of the general population, at least with respect to longevity, appears to have improved, substantial differences in longevity and health outcomes among racial and Hispanic-origin groups in the United States persist. Despite the improvements recorded, non-Hispanic whites, blacks, Hispanics, American Indians, and Asian and Pacific Islanders bear an unequal burden of death and disease. Differences in socioeconomic status are associated with differences in health status, and they explain a large part of the differences between the health of the races and Hispanic-origin groups. Differences in the quality of health care and access to health programs explain an additional as well as overlapping part of the difference, and genetic differences in predisposition to some diseases explain another part. Racial and ethnic differences in health status occur over a wide range of health conditions, even when income, insurance coverage, and other factors related to access are controlled. Hence, part of the difference in health outcomes is due to factors distinguishing the races independent of ability to pay or secure access to health care, such as less healthful lifestyle and behavior, less healthful work and home environments, genetic predispositions, and deficiencies in the quality of health care provided.

Illustrative Black-White Differences

Black men are 35% more likely to die than white men in a current year, according to age-adjusted death rates for 2005 (Table 7.4; U.S. NCHS 2008). Black men have higher prevalence ratios for HIV/AIDS. They are more likely to have heart disease, high blood pressure, and kidney disease, particularly kidney disease associated with hypertension. Young black men in the inner cities have disproportionately high rates of hypertension as a result of unhealthy lifestyles, such as poor nutrition and alcohol abuse. Black men are 34% more likely to die from cancer than white men. They suffer from much higher rates of almost every type of cancer, including lung, prostate, colon, oral, and stomach cancer. The available analyses tend to find that most of the black-white difference in cancer can be explained by the difference in socioeconomic status, but the allocation of "responsibility" between genetic and non-genetic factors remains speculative.

Consider the special case of prostate cancer. A black man has more than a 50% greater risk of incurring prostate cancer than a white man and has more than twice the risk of dying from the disease. Prostate cancer rates have declined in the last several decades, but black men have experienced a decline only about one-half that of white men. Although men of all racial groups show increasing rates of precursor lesions for prostate cancer with advancing age, the lesions progress more rapidly among blacks than whites. In sum, the disease is both more common and more lethal among blacks than among whites. The reasons for this are unknown. Among the risk factors may be chronic infections and poor diet – both of which are involved in contributing to many other types of cancers – as well as inadequate and delayed

Table 7.4 Death rates for non-Hispanic whites, non-Hispanic blacks, and Hispanics, by age and sex, for the United States: 2005

Age and sex(years)	(Rates per 100,000 population)				
	Non-Hispanic			Ratio	
	White	Blacks	Hispanics	Black ÷ white	Hispanic ÷ white
Males					
All ages	970.6	825.6	334.4	0.85	0.34
Under 1	625.7	1,451.4	670.2	2.32	1.02
1–4	29.9	47.6	33.2	1.59	1.11
5–14	17.4	27.8	15.3	1.60	0.88
15–24	105.1	177.7	120.4	1.69	1.15
25–34	134.1	264.0	115.5	1.97	0.90
35–44	236.1	407.5	182.0	1.89	0.86
45–54	517.2	969.9	417.4	1.88	0.88
55–64	1,079.6	1,993.8	875.8	1.85	0.81
65–74	2,584.5	3,814.1	2,029.4	1.48	0.79
75–84	6,420.4	7,771.1	4,856.8	1.21	0.76
85 and over	15,401.3	13,978.1	10,140.5	0.91	0.66
Age-adjusted rate	945.4	1,275.3	717.0	1.35	0.76
Females					
All ages	992.6	727.6	278.2	0.73	0.28
Under 1	496.5	1,198.0	555.4	2.41	1.12
1–4	22.2	37.2	24.5	1.68	1.10
5–14	12.9	20.0	12.0	1.55	0.93
15–24	42.2	52.6	38.6	1.25	0.91
25–34	62.1	114.1	41.1	1.84	0.66
35–44	137.0	258.5	90.6	1.89	0.66
45–54	298.7	582.4	216.4	1.95	0.72
55–64	677.2	1,127.1	493.9	1.66	0.73
65–74	1,729.6	2,383.1	1,291.6	1.38	0.75
75–84	4,579.7	5,338.1	3,365.8	1.17	0.73
85 and over	13,683.1	12,941.6	9,068.4	0.95	0.66
Age-adjusted rate	677.7	860.5	485.3	1.27	0.72

Source: Based on U.S. National Center for Health Statistics (2008)

medical attention. The question must also be posed whether blacks have a greater genetic susceptibility to this and other forms of cancer than whites. Regardless of the answer, given the magnitude and significance of the race differences for prostate and other cancers, public information programs and screening for cancer at an earlier age for black men than white men could reduce the differences in the death rates from these diseases.

Black women are 27% more likely to die than white women in a current year according to age-adjusted death rates for 2005. Black women suffer disproportionately from cancer as compared to white women, and those who have cancer have lower survival rates than white women who have the disease. For example, black women with breast cancer have a three out of four probability of surviving for

Table 7.5 Mean body mass index, by age, sex, and Hispanic ethnicity for the United States: 1988–1994 and 1999–2002

	Male		Female	
Age and hispanic origin	1988–1994	1999–2002	1988–1994	1999–2002
Non-Hispanic white				
20 and over	26.8	27.9	26.1	27.6
20–39	25.9	27.1	24.7	26.7
40–59	27.6	28.7	27.2	28.3
60 and over	27.0	28.3	26.7	28.2
Non-Hispanic black				
20 and over	26.6	27.5	29.1	31.1
20–39	26.3	27.1	27.6	30.2
40–59	27.1	27.7	30.4	32.1
60 and over	26.4	28.0	29.4	31.1
Mexican American				
20 and over	27.3	28.0	28.4	29.0
20–39	26.1	27.1	27.2	27.8
40–59	28.6	28.9	29.7	30.4
60 and over	27.1	28.1	28.7	28.9

Source: U.S. National Center for Health Statistics (2004); Primary source: National health examination survey and National health and nutrition examination survey
Note: Overweight but not obese, $25 =/< \text{BMI} < 30$; obese, $\text{BMI} =/> 30$; healthy weight, $18.5 < \text{BMI} < 25$
Body mass index is calculated as weight in kilograms divided by height in meters squared

5 years, compared with a nine out of ten probability for white women. The higher mortality of black women from breast cancer can be attributed to the higher rates of such accompanying conditions as diabetes and hypertension, their inferior medical treatment, and the more advanced stage of their cancer when diagnosed.

Black women suffer disproportionately from lupus and depression as well as cancer, diabetes, and hypertension. They are much more likely to be obese than white women, as measured by BMIs (while white men and black men have similar obesity levels) (See Table 7.5). About four-fifths of middle-aged (ages 40–59) black women are overweight and about half are obese (BMI = 32.1). Only about half of middle-aged white women are overweight (BMI = 28.3).

Hispanic Paradox

Overall, the recorded age-adjusted death rate of Hispanics is only 74% of the non-Hispanic rate, with the difference between the gap for the Hispanic/non-Hispanic rates for males and females moderately favoring females (Table 7.4). In view of the well-known differences in socioeconomic status of these groups, there has been reasonable concern that the data are in error. The reports of NCHS presenting such data note that the data for Hispanic origin should be interpreted with caution

because of inconsistencies between the reporting of Hispanic origin on death certificates and the reporting of Hispanic origin in censuses and surveys (U.S. NCHS/Hoyert et al. 2006a). Some analysts have confronted this so-called Hispanic mortality paradox by examining other sources of mortality data for Hispanics. A study employing Medicare data shows a smaller, but still substantial, advantage of Hispanics relative to non-Hispanic whites (Palloni and Arias 2004).

Studies of the Hispanic mortality paradox by Markides and Eschbach (2005) also do not dispute the lower death rates and higher life expectancy of Hispanics. They reviewed the vital statistics data and studies using national sample surveys linked to the National Death Index, Medicare data linked to applications for Social Security cards ("Numident" files), and other sources. They found that the mortality advantage of Hispanics was much smaller than in the vital statistics data and that it occurred principally among older persons, persons of Mexican origin, and men. They hypothesized that the advantage could be attributed to selective return migration of less healthy immigrants to Mexico.

In spite of the apparent mortality advantage, Hispanics are less healthy than whites as suggested by self-reports of health status (National Research Council 2004a). Behavioral risk factors, such as smoking, overeating, lack of exercise, and excessive alcohol use, are more common among Hispanics than whites. Markides and Eschbach (2005) reported higher prevalence ratios for chronic conditions among elderly Hispanics, such as diabetes (especially Mexican-Americans and Puerto Ricans), disability (also Mexican-Americans and Puerto Ricans), obesity, and depression. White and Mexican-American men have similar obesity levels (as measured by BMIs), but Mexican-American women are more likely to be obese than white women (Table 7.5). Markides also noted that Hispanics have lower percentages of Medicare coverage and usage and that they are less likely to utilize physician care. These findings are consistent with those of Crimmins et al. (2007). The latter researchers used data on adults from the National Health and Nutrition Examination Survey (NHANES) to compare the blood pressure and the metabolic and inflammatory risk profiles of whites and Hispanics. They did not find evidence of a Hispanic paradox on the basis of biological risk profiles. They found poorer risk profiles for foreign-born Hispanics, as compared with whites, but these disadvantages disappeared when the results were controlled for socioeconomic status. At the same time, higher biological risk scores were obtained for native Mexican-Americans as compared with both whites and foreign-born Mexican-Americans.

Further Interpretation of Race/Ethnic Differences

The differences in health outcomes among racial and ethnic groups in the United States can be attributed to a range of factors, both social/cultural and genetic. Associated with these factors, various processes of selection play a part through differences in survival of the various component groups, as suggested below in the discussion of "racial crossover."

Social and cultural factors. Factors like socioeconomic status clearly play an important role in the differences of the health of racial and ethnic groups. Marmot (2004), like many other analysts, believes that racial differences in mortality are largely or wholly due to socioeconomic factors. More affluent groups routinely receive better and more extensive medical treatment than poorer, marginalized groups. Nonwhite racial groups and Hispanics are likely to receive poorer care because of less favorable public policies for allocating resources to them, their disproportionate representation in restrictive health plans, their greater lack of health insurance, and linguistic, cultural, and educational barriers (National Research Council 2004a).

Cultural differences along with the other factors cited affect the way in which patients seek medical care and comply with treatments prescribed by their doctors. Because of mistrust or misunderstanding, members of minority groups are more likely to refuse treatment or to fail to follow prescribed treatment. There may be a failure of communication between a U.S. medical system informed primarily by Western science, as compared with the cultural practices and expectations of different racial and ethnic groups. Such explanations fit the Hispanic/non-Hispanic differences in mortality and morbidity reasonably well. A reconciliation of the discrepancy between such explanations and the Hispanic mortality advantage and Hispanic morbidity disadvantage is needed, however.

Discrimination on the part of health providers, in addition, has been assumed to contribute to racial/ethnic differences in health. The channel by which this factor could play a part is through inequitable treatment in access to health care and in the quality of health care provided. Geiger (2001) specifically points to differences in the diagnostic work-up and treatment of minority patients already in the health care system, including lesser opportunities to receive advanced treatment and poorer clinical care. Yet, the NAS Panel Measuring Race Discrimination (2004) observed that it would be hard to address the question whether an adverse outcome for a group would have been different if the group had been assigned to a different race. The panel concluded that, in view of the many ways of measuring race and the complexity of the causes of the differences in the health of the races, the determination of the role of discrimination among the causes of death would be both exceedingly complex and problematic.

Genetic factors in racial/ethnic differences. As the National Research Council (2004a) reminds us, all sociocultural factors should be considered in terms of their links to stress and the biopsychological and biophysical factors that lead to impaired health. This is a reminder too that socioeconomic and cultural factors do not fully explain the differences in health outcomes. Genetic factors also play a role in the racial/ethnic variations in health. An undetermined part of the differences between the rates of sickness and death among the racial and ethnic groups has genetic origins. The different race/ethnic groups experience in different degrees the gene mutations that render them more or less vulnerable to various diseases. Some of these mutated genes are not expressed until late in life and some, with their associated diseases, target particular racial or ethnic groups early in life.

Among the latter are sickle-cell anemia, a degenerative blood disease, which affects mainly blacks, and Tay-Sachs disease and Gaucher's disease, both degenerative metabolic/neurological diseases, which strike mainly Jews of East European origin.

Jews of Ashkenazi (Eastern European) descent are at increased risk for carrying the genes for several genetic diseases and for transmitting them to their children. The genes are inherited as recessive traits. The disproportionate frequency of these diseases among this ethnic group appears to be related to its practice of endogamy. The so-called Jewish genetic diseases, ranked in order of frequency in the Jewish Ashkenazi population, are as follows[5]:

1. Gaucher disease Type I
2. Cystic fibrosis
3. Tay-Sachs disease
4. Familial dysautonomia
5. Canavan disease
6. Glycogen storage disorder Type 1a
7. Maple syrup urine disease
8. Franconi anemia type C
9. Niemann-Pick disease type A
10. Bloom's syndrome
11. Mucolipidosis IV

These are mostly neurological or metabolic diseases with poor prospects for survival unless early treatment is initiated.

Racial Crossover

I have already discussed the phenomenon of the crossover of age-specific death rates of paired populations and the common explanations for it in Chap. 3. The crossover phenomenon is by no means restricted to race differences in mortality, but it came to the attention of demographers in the United States first because of the race crossover and has been analyzed principally with respect to the races. I include a further note here on this phenomenon.

The recorded age-specific death rates of blacks exceed those of whites at every age until the very high ages, when the recorded death rates for whites exceed those for blacks. The age of the recorded black/white crossover has risen over time so that in 2005 the crossover occurred at about age 87 (Table 7.6). In earlier decades the age of the crossover was much lower. For example, it was 75 in 1950 and 79 in 1970. The crossover of white and black death rates occurs for each of the sexes and the sexes have parallel crossover histories. In 2005 the age for the race crossover was 87 for men and 87 for women. Back in 1950 the ages were 74 and 76.

[5]Information provided by the Jewish Genetic Disease Consortium of New York City.

Table 7.6 Estimated age of crossover of death rates of whites and blacks, by sex: 1929–1931 to 2005

Period or year	Both sexes	Male	Female
1929–31	77	80	75
1939–41	73[a]	73	73
1949–51[b]	75	74	76
1959–61[b]	76	75	77
1969–71	79[a]	78	80
1979–81	85	84	85
1989–91	86	86	87
1999–2001	88	89	88
2005	87	87	88

Source: Based on mortality rates in unabridged U.S. life tables published by the U.S. National Center for Health Statistics or predecessor agencies

[a]Use of data for all races other than white, instead of data for blacks only, gives the same results

[b]Data for all races other than white are used to represent black

The persistent question has been, is the race crossover real? The evidence is not consistent and different analysts come down on opposite sides of this question. Manton and Stallard (1981) consider the crossover real and attribute it to the selective "survival of the fittest," or the "heterogeneity of frailty." This process depends on the fact that any cohort is heterogeneous with respect to the health of its members and involves the selective removal by mortality of the less healthy members at the younger ages, leaving the healthier members to survive to the older ages. It would apply to the whites as well as blacks, so that the results imply that the surviving blacks are more fit than the surviving whites, presumably because of a greater elimination of weaker blacks than weaker whites at earlier ages. Kestenbaum (1992) concluded that the crossover is real on the basis of improved Medicare enrollment data. Another group of analysts consider the racial crossover a statistical artifact resulting from the poor quality of the data for blacks. They maintain that the considerable bias in the reporting of age of death for blacks at the higher ages tends to cause an underestimate of the death rates at these ages. This view is held by Coale and Kisker (1986) and Preston et al. (1996). The widespread occurrence of such a crossover between populations of different socioeconomic statuses and different ethnic and demographic groups supports the position that the crossover is real (Nam 1995) and the steadily rising age of the crossover supports the position that it is an artifact. I believe that the preponderance of evidence falls on the side of the "real-ists."

Marital Status and Family Arrangements

Marital Status and Mortality

Married life, as compared with single, divorced, or widowed life, has been hailed as a great protector of physical and mental health and a prime contributor to increased length of life. There is considerable evidence of the health advantages of the married state. Indeed the data for the United States show that married persons have much lower death rates than persons in the other marital classes. This is true for both men and women, with the survival advantage of the married state generally being slightly greater for women than men. Table 7.7 gives index numbers relating the death rate of each marital class to the death rate for all marital classes combined, for the older ages, in 2003, 1979–1981, and 1949–1951. For example, in 2003 at ages 65–74 the death rate of married men was 17% lower than the death rate of all men at these ages, whereas for single (i.e., never-married), widowed, and divorced men the death rate tended to be about 44% to 71% higher. The corresponding figures for women aged 65–74 are one-quarter lower (for married women) and one-third to two-thirds higher (for other marital classes). This general pattern has been characteristic of marital mortality in the United States during the second half of the twentieth century, as shown by the figures for 1949–1951 and 1979–1981 at these and other older ages in Table 7.7. These data also suggest that the gap between the mortality of married persons and the other marital groups has been widening.

Table 7.7 Comparative mortality of older persons according to marital status, by age and sex: 1949–1951 to 2003

	Male					Female				
Age and year	All classes	NM	M	W	D	All classes	NM	M	W	D
2003										
55–64	1.00	1.98	0.73	2.16	1.94	1.00	1.69	0.74	1.64	1.37
65–74	1.00	1.60	0.83	1.44	1.71	1.00	1.64	0.73	1.29	1.37
75 and over	1.00	1.33	0.80	1.50	1.32	1.00	1.38	0.51	1.17	1.04
1979–1981										
55–64	1.00	1.63	0.86	1.88	1.98	1.00	1.33	0.84	1.42	1.31
65–74	1.00	1.35	0.89	1.49	1.62	1.00	1.11	0.82	1.19	1.14
75 and over	1.00	1.23	0.87	1.40	2.15	1.00	1.08	0.63	1.11	0.90
1949–1951										
55–64	1.00	1.46	0.89	1.44	1.75	1.00	0.98	0.90	1.24	1.21
65–74	1.00	1.32	0.89	1.18	1.56	1.00	0.97	0.90	1.06	1.31
75 and over	1.00	1.15	0.84	1.16	1.45	1.00	1.02	0.75	1.05	1.27

2003: Based on U.S. National Center for Health Statistics (2006a)
1979–1981: Based on unpublished tabulations of the U.S. National Center for Health Statistics
1949–1951: Based on U.S. National Center for Health Statistics, Table 7.2, May 8 (1956)
Note: Death rates are shown as index numbers based on the death rates of the total male or female population at each age. Marital status not stated in 2003 was not distributed
NM never married, *M* married, *W* widowed, *D* divorced

The differences in marital mortality would translate directly into differences in life expectation according to marital status. However, few life tables for the marital classes have been constructed, either in the United States or elsewhere, to suggest the magnitude of the gap between the life expectancies of the marital classes.[6]

The general findings described here for the United States characterize other countries as well. A study of mortality differences according to marital status among 16 developed countries, conducted by Hu and Goldman (1990), found that unmarried persons, especially men, have greater mortality than married persons, and that divorced persons, especially divorced men, have the highest death rates among the unmarried groups of each gender. Using log-linear models to assess the effect of marital status on mortality over the adult ages, the researchers also found that the excess mortality of unmarried persons has generally been increasing over the decades from 1950 to 1980.[7]

The great importance of this difference, as noted by Martikainen et al. (2005), is that in many countries differences in mortality for the marital classes are greater than differences for socioeconomic classes although there is clearly an interaction between these factors. They found that, for Finland in 1996–2000, much of the difference in mortality for marital classes can be explained by socioeconomic status and household composition, in particular by the economically disadvantaged status of the nonmarried groups. The greater affluence of married persons than those in other marital statuses is associated with their better health as well as greater longevity. Psychological factors in addition to socioeconomic factors may influence the direction and magnitude of the differences. Other studies have shown that marital-status differences persist even when the effects of socioeconomic status are controlled.

The differences in mortality rates according to marital status play out in many specific causes of death, especially those relating to social and psychological adjustment (Gove 1973). Suicide rates for single men are much higher than those for married men. Single and divorced men have higher death rates from

[6] A tabulation of deaths for marital classes in the United States is available annually from NCHS and the age detail is sufficient (7 age groups 15+) to make approximate life tables. Grabill constructed a life table for single women in the United States for 1940 from which I extracted the following approximate life expectancies for several older ages:

Age	Never married	All marital classes
55	16.1	18.2
60	13.8	15.0
65	11.4	12.1

Source: Based on Grabill (1945) and the official U.S. Census Bureau life table for 1939–1941.

[7] From a nuptiality life table constructed for Belgium, 1970–1971, Willekens et al. (1982) found that the life expectancy of married women at age 20 exceeded the figures for the other marital statuses at age 20 (56.3 years vs. 54.3 to 56.1).

cirrhosis of the liver and lung cancer, causes linked to excessive use of alcohol and tobacco, respectively. The marital-status differences for women are the same but less extreme. In fact, for nearly all causes of death, married persons have substantially lower death rates than nonmarried persons (Martikainen et al. 2005).

Marital Status and Health

Data from the National Health Interview Surveys for 1999–2002 show that regardless of the health indicator or risk factor – whether fair or poor health, limitations in activities, low back pain, headaches, psychological distress, smoking, leisure-time physical inactivity – married adults were generally found to be healthier than adults in other marital-status categories (U.S. NCHS/Schoenborn 2004). Marital-status differences in health were found in each of the three age groups studied (18–44 years, 45–64 years, and 65 years and over) but were most striking among adults 18–44 years.

In addition to describing the extent of the differences, such data are likely to suggest causation to many. Marriage can improve health in a variety of ways, whether by imposing structure and stability in the partner's life, providing social support and hence reducing stress, and serving as an inducement for maintaining good health (Lillard and Panis 1996; Waite and Gallagher 2000). It is also possible, however, that people who get married are a selected group with superior health and remain married because of their superior health and ability to expand their income. From the available data and studies, we cannot ascertain whether marriage or health is the primary causative factor; rather, both appear to be cause and effect of one another. Or they both may have a common cause. A special longitudinal analysis would be needed to clarify and possibly establish the causation/selection role of marriage in health/mortality.

The Family Environment and Health

Family membership affects the stresses we experience while bestowing on us our genes and our home environment. It is not easy to specify how much variation in individual health is explained by the family environment, but here are several illustrations of the family effect. The spouses of people who suffer from such diseases as asthma, depression, hypertension, or ulcers are at sharply increased risk of developing the very same condition; this may result from sharing the same diet, environment, and lifestyle. A parent's death is associated with the depression and anxiety symptoms of their children, and the children experience these conditions when a parent is beset with them. The rate of suicide increases several times for persons whose spouses have taken their own lives. The loss of a child leads to a major increase in a woman's risk of death in the several years following the death of the child.

Effect of spouse's death on longevity of the survivor. Actuaries as well as family researchers have long been interested in the survival of widowed persons in the event of the death of the spouse. Studies consistently show that caring for a sick spouse can have negative health consequences for the caretaker and that a spouse's death can significantly increase the survivor's risk of death. The first of these effects is the so-called "caregiver's burden" and is especially heavy for women (Pruchno and Resch 1989). The second of these effects is the so-called "widower effect" and is especially severe on men (Schaefer et al. 1995; Lillard and Waite 1995). The death of a spouse doubles the likelihood that the other spouse will die within a year. The premature death of a surviving spouse may result from the loss of the physical and emotional support of the ailing spouse and the stress of caring for that ailing spouse during the period just before death.

Effect of quality of marriage and of death of spouse on coping experience. There is evidence that individuals who were in happy marriages are much less likely to experience long-term emotional difficulties after the death of their spouse. A stable emotional base in a marital relationship allows the survivor to cope better with the loss of a marital partner. It seems reasonable that, if stress is inimical to the survival of bereaved spouses, methods of reducing this stress could contribute to the longevity of the spouses. There is evidence that the "quality" of the death of a spouse can have a positive effect on the mental and physical health of widowed persons. Christakis and Iwashyna (2003) have shown that the quality of health care of a dying spouse, as represented by choice of hospice care, may have positive health effects on the decedent's family members. The impact of hospice, as a particularly supportive type of end-of-life care, on the spouses of patients who succumb to some disease, is especially beneficial for bereaved wives. Christakos and Iwashyna came to this conclusion by conducting a matched cohort retrospective study using the propensity-score method (see Appendix A).

Blacks fail to make adequate use of hospice care and palliative-care services. This is so for two principal reasons, cost and preference for private informal family support. Use of these formal programs by their dying family members could vastly improve pain management and the overall quality of their lives in their final days, according to the Robert Wood Johnson Foundation online report, "Disparities at the End of Life" (www.rwjf.org/news/eoldisparities).

The family's social network and health care. In addition to the vast formal network of public and private health providers, family members, friends, and neighbors provide a large informal network of social and health support. A survey conducted for the National Alliance for Care giving and the AARP yielded an estimate of about 45 million caregivers providing unpaid care to another adult. Almost six in ten of these caregivers either work or have worked while providing care and over three-fifths have had to make some adjustments in their work life, from reporting late to work to giving up work entirely. Almost two-fifths of the caregivers are men and three-fifths of them are working full-time. Families differ greatly in the kind and degree of informal support they can provide when a member becomes

dependent. The composition of the family (e.g., the number of living children) and the geographic dispersion of its members are important factors affecting the kind, amount, and quality of family support.

Socioeconomic Status

Basic Considerations

In turning to the relation of socioeconomic status and health, it is useful at the outset to set forth some basic considerations bearing on this relation. Socioeconomic status is not normally an operational concept in itself, but is operationalized in terms of educational level, income, wealth, occupation, or some combination of these specific characteristics. It is possible, however, to employ a subjective evaluation of socioeconomic status, where it is self-reported by the respondent. Hence, research studies may differ in the proxy used for socioeconomic status, and different researchers may use different definitions of these proxies or do their analyses with different levels of classification detail. Commonly in the United States the definitions of these concepts employed in the decennial census, Current Population Survey, or American Community Survey are accepted for research in this area, but other large national sample surveys or administrative records may employ other definitions. The MacArthur Foundation, for example, for its Research Network on Socioeconomic Status and Health has developed a subjective social status scale. Each criterion of socioeconomic status may be expected to produce somewhat different results for the association of socioeconomic status and health (Grundy and Holt 2002).

To describe the socioeconomic status of families and households, married couples can be considered as a unit for income, wealth, and occupation. Those women who do not have a career occupation may be assigned the occupation of their husbands. Children can be assigned the status of their parents.

The indicators of socioeconomic status differ with respect to their tendency to change during an individual's lifetime. Formal education does not change much for a cohort or the individual members of it after age 30, but income, wealth, and occupation do change, sometimes frequently and considerably, as individuals grow older. Inasmuch as these changes are common and may influence the interpretation of the relation of socioeconomic status and health, any analysis of this relation should try to include data on the differing statuses. Data on educational level cannot be used for this purpose even though they are more accessible and easier to apply.

The reader may recall the discussion of the role of causation in the relation between marital status and health. It is apparent that socioeconomic status and health can influence one another, but the magnitude of the influence of each variable in each direction is a question to be resolved. The effect of health events on socioeconomic status may be as strong as the influence of socioeconomic status on health. This effect is both direct and indirect. Health events have an influence on

household income and wealth through direct out-of-pocket expenditures for health care, and they may also affect income and wealth more drastically through their effect on participation in the labor force, including disability retirement, part-time employment, and loss of the income of the second-earner. In addition, health and income may both change in response to other common factors, e.g., marital status and education.

One issue to be resolved is the relative importance of the socioeconomic indicators – income, wealth, occupation, or education – in health outcomes. Which has an independent and dominant influence on health? Another issue is the relative influence on health in later life of economic circumstances early in the life course and economic circumstances late in the life course. Then there is the question of the influence of demographic forces on the health history of a cohort. They may be expected to vary as the cohort grows older and it's socioeconomic and health composition shift from childhood to old age. Socioeconomic status is a factor in modifying the health composition of the cohort as it ages. People of lower socioeconomic status are more likely to get sick and die early, so that the cohort tends to have a larger share of healthier people of higher socioeconomic status at the older ages and socioeconomic differences in mortality tend to be less pronounced at the older ages (Crimmins 2005; Hoffman 2005; Smith 2004).[8]

Specific Findings

Findings from numerous studies show a strong negative association between socioeconomic status and mortality/ill-health (e.g., Marmot et al. 1984; Preston and Taubman 1994; Crimmins 2005; Smith 2004, 2004; Feinglass et al. 2007). The social gradient in health is continuous and significant (e.g., Adler et al. 1999; Marmot 2004; Marmot et al. 1995). Socioeconomic differences in lung cancer mortality have been observed in many country studies, and the results typically show that persons in the lower status have higher lung cancer mortality than persons in the higher status. To understand the gradient, both biological and environmental factors have to be taken into account (National Research Council 2004a; Crimmins and Seeman 2004). Generally the various dimensions of socioeconomic status, namely income, wealth, education, and occupation, appear to influence health outcomes to some degree independently; that is, even where income is the same, differences in education are associated with differences in health outcomes. Occupational differences also add to the gradient, particularly between blue-collar and white-collar workers.

[8]This theory needs to be reconciled with the crossover of the mortality of blacks and whites in the United States at the higher ages, given that blacks are concentrated in the lower socioeconomic segment of the population. At the oldest ages the mortality of blacks is lower than that of whites but the mortality of the highest and lowest socioeconomic classes appears to be nearly the same. There is a conflict between the view that a harsh existence in childhood builds vitality and contributes to survival to old age, and the view that higher socioeconomic status in one's younger years contributes to greater survival. Both forces may be in play, with one, presumably the second, exerting far greater influence.

Seeman et al. (2004) have examined mortality differences in socioeconomic status on the basis of biomarkers. They found that a measure of cumulative biological risk, i.e., cumulative allostatic load, explained 35% of the difference in mortality between those with higher SES and those with lower SES (as measured by less-than-high-school education versus high-school-or-greater education). Pappas et al. (1993) maintain that socioeconomic status has a greater effect on mortality than smoking, hypertension, and high cholesterol combined. The Pappas et al. and the Seaman et al. findings may be viewed as consistent in that a high allostatic load can be largely explained by the negative lifestyle and the unhealthful environment associated with low socioeconomic status.

The differences in mortality between socioeconomic groups have been growing (Glied and Lleras-Muney 2008; Pappas et al. 1993; Feldman et al. 1989). Socioeconomic differences in mortality were strong during the early part of the twentieth century when communicable diseases predominated among the causes of death. With the decline in these diseases and the growing importance of chronic diseases, socioeconomic differences in mortality appeared to diminish. The evidence is that the differences have become substantial again.

Education

There is an inverse relation between health (including mortality) and educational attainment. Educational attainment is classified in the United States, on the basis of a question on last year of school completed, into such categories as no schooling, elementary school only, some high school, completed high school, and some college. Figure 7.3, drawn from a study by Smith (2004), depicts the sharp inequality in the percent reporting fair or poor health according to three broad educational-attainment classes – 0–11 years of schooling, 12–15 years, and 16+ years – based on the National Health Interview Surveys for 1991–1996. The health disadvantage of not completing a high school education is starkly evident throughout the age scale. Smith found that, even after controlling for an extensive array of current health conditions, persons with less schooling were much more likely to experience a major negative health shock than those with more schooling, and this effect persists into old age.

Here are a few other illustrations of the education/health relation. The National Health and Nutrition Examination Survey for 1988–1994 showed a general inverse gradation in the years of education with having certain infectious diseases, including hepatitis A, B, and C, and herpes simplex virus (U.S. NCHS/Kruszon-Moran 2005). Feldman et al. (1989) found that, in 1960 and in 1971–1984, U.S. white men with no more than an elementary-school education were at substantially elevated risk for succumbing to heart disease compared to those who had completed high school or higher. The relative risks for these two education classes according to age in 1971–1984 ranged from 1.38 for men aged 65–74 years to 2.27 for men aged 45–64. The women also showed a strong inverse relation between education and mortality from heart disease. The relative risks ranged from 1.48 at ages 65–74 to 1.97 at ages 45–64.

In a major study of the variations in mortality and health with socioeconomic status in 22 countries of Europe in the 1990s and early 2000s, it was found that in almost all countries, death rates and the proportions with poorer self-assessments of their health were substantially higher in lower socioeconomic-status groups (Mackenbach et al. 2008). The measures used were the relative index of inequality and the slope index of inequality and the main variable used to represent socioeconomic status was educational attainment. The relative index of inequality in mortality between groups of more education and less education varied from less than two to over four. In all countries mortality was greater among those with less education. The inequalities in mortality were small in some southern European countries and very large in most countries in the eastern and Baltic regions. The magnitude of inequalities in self-assessed health according to education also varied substantially among the other countries of Europe. Among the causes of death, cardiovascular diseases accounted for a substantial share of the socioeconomic-class differences in mortality found among the countries. Considering underlying causes and possible interventions, the variations among countries appeared to be attributable in large part to causes of death related to smoking, alcohol use, and access to health care. Given the egalitarian welfare policies in the countries of northern Europe and their demonstrated socioeconomic inequalities in health, it may be inferred that, although a reasonable level of social security may be required to reduce health inequalities, that is not sufficient. Lifestyle and behavioral factors distinguishing the socioeconomic classes still play an important role in the differences in mortality in high-income countries (Mackenbach et al. 2008). This finding is reported in separate studies for Great Britain, Canada, Denmark, Norway, and Sweden – all countries with systems of universal health care.

There is evidence of the independent effect of educational status on health and of the dominance of education over the economic measures of socioeconomic status. Figure 7.3 provides such evidence in showing that the education/health gradient, unlike the income/health gradient, persists to the oldest ages (Smith 2004). This finding is supported by a study of Finnish data. Martikainen et al. (2000) compared the mortality from lung cancer during 1985–1988 for a group of 50–69-year-old Finnish male smokers who had received only a basic education (i.e., elementary school or less) with a group of Finnish male smokers of the same ages who had completed high school or more. Lung cancer mortality was 32% higher for the less educated than for the more educated men. The excess lung cancer mortality among the less educated male smokers was about 40% higher than for the general male Finnish population of the same ages. These results remained valid after adjustment for duration of smoking, occupational exposure to asbestos, and similar factors.

We should not be surprised that educational level is dominant among the socioeconomic factors affecting health outcomes. Higher education provides the intellectual and psychological background for taking advantage of newer developments in medical knowledge, adhering to instructions regarding the treatment for illness (some being rather complicated to follow, e.g., diabetes), seeking to

maximize the level of one's health, and anticipating the consequences of failure to secure treatment and follow the prescribed treatment (e.g., Glied and Lleras-Muney 2008; Goldman and Smith 2002).

Income, Poverty, and Wealth

In studies linking poverty/income and health in the United States, the official definitions for poverty and household income are usually employed. The poverty index is calculated by dividing total family income by the U.S. Census Bureau's poverty thresholds, adjusted for family size at the year of the interview.

Poverty and health. The National Health and Nutrition Examination Survey (U.S. NCHS 2005) showed a "gradation" in the poverty index (i.e., below, at, or above poverty) for the same group of infectious diseases as shown above for education. The Annual Housing Survey data for 2003 showed that 9% of the households with incomes below the poverty level in the United States reported seeing signs of mice in their housing units in the last 3 months, as compared with 6% for all occupied housing units; and 4.4% of the poor households reported major accumulations of trash in their streets as compared with 2.5% for all households (Table 7.8). (Note that the percents of housing units with black householders and with a variety of negative indicators of health risks closely resembled the percents for the households with incomes below the poverty level.) As a result of such living conditions, the poor have a higher risk of infectious diseases, as well as cardiovascular and other chronic diseases, than the nonpoor. In addition, those poor persons who are already burdened with serious chronic illnesses may be pressed to limit their use of prescription drugs because they cannot afford to pay for them, and thus put themselves at risk of experiencing more adverse health events.[9]

Income/wealth and health. People who have higher incomes and greater wealth tend to be healthier and to live longer. Smith (1998) has shown, with data from the Health and Retirement Study (HRS), that the income of persons between ages 51 and 61 in 1992–93 reporting themselves to be in excellent health is more than three times as great as the income of persons in poor health, and that their wealth is more than five times as great (Fig. 7.4). Hurd et al. (1998) used data for 1993 and 1995 from the Survey of Assets and Health Dynamics among the Oldest Old (AHEAD) to examine the variation of death rates of persons aged 70 and over during the 2-year period 1993–1995 with their level of wealth. They found that persons in the lowest

[9] An NIA-sponsored study covering middle-aged and older Americans who used prescription drugs found that 7% among them had taken less medication than prescribed because of the cost (Heisler et al. 2004). Those with heart disease who cut back on their prescribed medications because of the cost were 50% more likely to suffer a heart attack, stroke, or angina than those who did not report such cost-induced restriction on their use of medications. About one-third of the respondents who restricted their medication reported a significant decline in health status compared to one-fifth of the respondents who did not restrict their medication.

Table 7.8 Percent of occupied housing units with selected indicators representing health risks, according to characteristics of the householder: 2003

Indicator	Total	Black	Hispanic	65 years and over	Households below poverty level
Signs of rats in last 3 months	0.8	1.4	2.1	0.6	1.6
Signs of mice in last 3 months	6.0	9.1	7.7	4.4	9.0
Holes in floors	0.9	1.8	1.7	0.5	2.0
Exposed wiring	0.6	0.5	0.8	0.5	1.0
Water leakage from inside structure[a]	8.0	10.7	8.7	4.5	9.4
Water leakage from outside structure[a]	10.4	11.1	8.5	7.8	10.0
Major accumulation of trash on street	2.5	4.7	4.4	1.8	4.4
No public transportation	41.8	25.7	24.8	45.2	37.4
No telephone	2.8	3.9	3.6	2.2	4.2

Source: U.S. Census Bureau (2006)
[a]During the 12 months prior to the survey

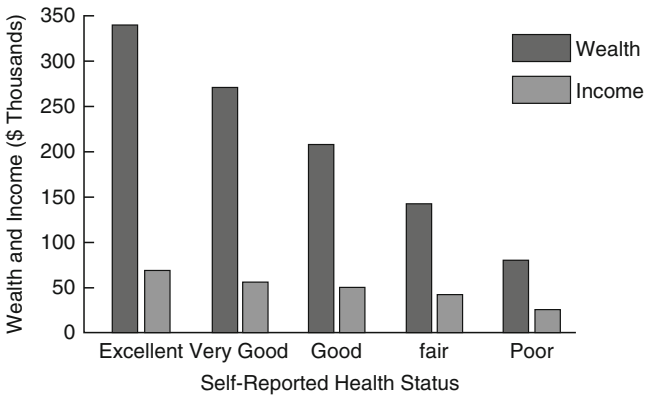


Fig. 7.4 Wealth and income according to self-reported health status, for the U.S. population 51–61 years of age: 1992 (Source: U.S. National Institute on Aging (1999); From: Smith (1998); Reprinted with permission of the America Economic Association. See also U.S. National Institute on Aging (1999). Primary source: Health and Retirement Study)

wealth quartile were twice as likely to die during this period as individuals in the highest wealth quartile. Similar results were found by Attanasio and Hoynes (2000) on the basis of SIPP data (Fig. 7.5).

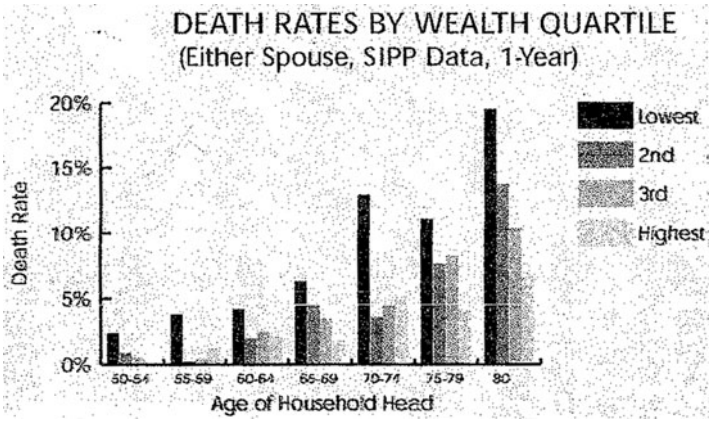


Fig. 7.5 Death rates of one or both spouses in a couple according to wealth quartile, for the population 50 years of age and over, by age: United States, 1984 and 1987 combined (Source: Attanasio and Hoynes (2000). © 2000 by the Board of Regents of the University of Wisconsin Press. Primary source: Survey of Income and Program Participation (SIPP). See also National Institute on Aging (1999)

Interactive relation of economic status and health. The relation between economic status and health is interactive and this interaction goes on throughout the life course. Low economic status limits access to good health care, and poor health diminishes one's economic status and prospects (Smith 2004). The reduction in income and wealth to which adverse health events can lead is a consequence not so much of the out-of-pocket costs of the health care, but of their impact on the labor force participation of the breadwinner (U.S. National Institute on Aging 1999; Smith 1998). Smith found that persons who experience serious adverse health events have much larger reductions in their total wealth than result merely from their out-of-pocket medical expenditures. This difference is explained by the effect of the adverse health event in bringing about earlier retirement or other reductions in work of the ill-person and his/her caretaker.

There is also a psychological aspect to the interrelation of economic status and health. The work situation itself, specifically the level of the job and the degree of independence in the job, affects the worker's health. Lack of autonomy in a job, or over one's life in general, can cause serious stresses. Michael Marmot (2004), the well-known British epidemiologist, has devoted considerable attention to the analysis of this phenomenon, which he calls the "status syndrome." Stress occurs when people work in dead-end but demanding jobs that provide little reward for their efforts. Persons worried about difficulties in their jobs and homes are not likely to attend to their basic health, but are likely to engage in unhealthful and risky behavior that may bring immediate pleasure, such as smoking, overeating, and speeding on the road. An association has been found between socioeconomic status and mental health, particularly depression. Here again the issue of selection and causation is hard to resolve; both influences apply in varying degrees throughout the life course.

Some international interrelations. A wide difference in income within the countries of Latin America is associated with differences in life expectancy, according to PAHO (2002). The income gap was calculated in these countries by comparing quintiles of population and income, i.e., the average income of the top fifth of a population and the average income of the bottom fifth. The PAHO study found that low-income nations with relatively equal distributions of income often had more favorable health indicators than better-off countries with wide income gaps. For example, according to PAHO, Brazil has a relatively high average income but great income disparity; its life expectancy is relatively low (68 years) and its infant mortality is relatively high (38 infant deaths per 1,000 births). Peru, on the other hand, has a relatively low median income but a smaller income spread; it has a higher life expectancy (70 years) and a lower infant mortality rate (37 per 1,000 births).

Marmot (2004) and Marmot et al. (1995) have elaborated the basis of the socioeconomic gradient in health and mortality in Great Britain and other countries. Marmot believes that the national level of per capita income is a primary factor in the poorest countries in accounting for poor health and high mortality but that, as PAHO found, inequalities with respect to income within countries are even more important for societies other than the poorest ones. This finding is suggested by the fact that many countries that have lower gross domestic products per person than the United States but less socioeconomic inequality have higher life expectancies. According to Marmot, extreme inequalities in social status, as represented by inequalities in education, occupation, degree of autonomy, job security, quality of the job, degree of social participation, environmental conditions, and similar factors, undermine public health, while relative equality in these areas promotes the general health of the population.[10]

Occupation

Like income, occupation can be a useful indicator of changing socioeconomic status throughout the life course because occupations may change over the working lifetime of the individual. Life-course occupational data can be obtained in a single survey or over a series of dates in a longitudinal survey. On the other hand, compilation of occupation data in comparable categories is more complicated than compilation of comparable income and education data, comparable data over the life course are often not available and are hard to secure, and useful data are not available for the many persons who have not pursued regular working careers in the

[10]Marmot concludes on this basis that favorable health outcomes for blacks in the United States are more likely to be achieved if programs such as maternal and child health programs are targeted toward poor, less educated adults, the clients are trained to make more effective use of these services, child immunization drives are targeted toward those who are less likely to become immunized without the program, and clinics and mobile health facilities are established in locations convenient for poor clients.

labor force. Different results may be obtained on the basis of longest occupation and current or last occupation. Moore and Hayward (1990) showed that mortality variations based on current or last occupation differed substantially from mortality variations based on longest occupation, when they controlled for education, income, health status, and other sociodemographic factors.

Great Britain has a long history of compiling vital statistics for occupational groups. These studies have shown a consistent history of variation of mortality with social class based on occupation. Tietze (1943) constructed life tables for males in seven broad social classes based on occupational groups for England in 1930–1932 and found that life expectation ranged from 63 years for the highest class and agricultural workers to less than 56 years for unskilled workers and coal miners. More than a half century later Moore and Hayward (1990) arrived at similar results for a cohort of males aged 55 years and older in the United States for the period 1963–1989. Life expectancy at age 55, for persons classified according to the longest occupation held, varied from 22.7 years for the professional class to 18.3 years for the service class (Fig. 7.6).

In sum, socioeconomic status may be represented by different variables, namely, education, income/wealth, or occupation, and notably disparate results may be obtained for the various socioeconomic variables. Hence, it is desirable to evaluate at least two of the three indicators of socioeconomic status in any program to determine the direction and magnitude of socioeconomic differences in health, if the data and resources are available to do so.

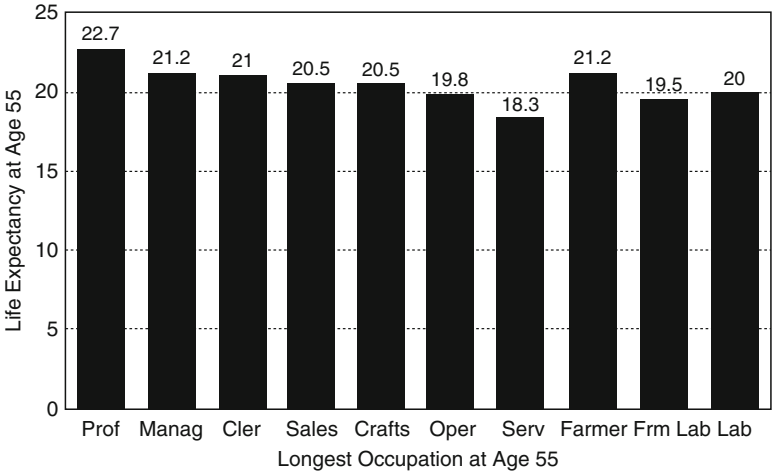


Fig. 7.6 Life expectancy at age 55 according to longest occupation at age 55, for the U.S. male population 55 years and over: 1966–1983 (Source: Moore and Hayward (1990). Reprinted with permission of the Population Association of America. Primary source: National Longitudinal Survey of Mature Men)

Religious Participation[11]

The available research suggests that the links between religion and health are quite strong. The studies suggest that many aspects of religious involvement (e.g., attendance at religious services) are beneficial to health (Idler 2002). Attendance at religious services is associated with better physical health (e.g., lower blood pressure, stronger immune function, improved physical functioning), better mental health (e.g., lower levels of depression and anxiety), and better perceived health (Ellison and Levin 1998; Idler and Kasl 1992 1997). Accordingly, we can expect religious participation to reduce mortality rates and increase life expectancy. Many types of studies – national health surveys, epidemiological surveys, and clinical studies – provide consistent and ample evidence linking religious attendance with lower mortality (Hummer et al. 1999; George 2002). Religious attendance is inversely associated with poor subjective health, and poor subjective health is a good indicator of a greater mortality risk (Idler and Kasl 1992). George (2002) offers the arguably exaggerated estimate of a life-expectancy advantage of 7–8 years for those who attend religious services regularly over those who attend religious services rarely or never.

There are several possible explanations for the association between attendance at religious services and a reduced mortality risk (Idler 2002). By encouraging group activity, particularly church or temple membership, religion serves to reduce personal isolation. Personal isolation has been associated with higher risks of mortality. Religious participation imbues a sense of community with other members of the congregation. Religious congregations mix all generations, bringing children, adults, and older persons together in social contact. Religious attendance encourages wider connections, more extensive family ties, and greater civic participation, so that those attending have more social support from family and friends. As explained in Chap. 6, greater social support is associated with reduced stress, improvements in health, and reduced mortality risks.

Religious persons are more likely than others to pursue a healthful lifestyle. In general, religions teach moderation and discourage risk-taking behavior. Religious attendance is associated with less alcohol use, less alcoholism, less use of tobacco, and greater cessation rates of smoking. As set forth in Chap. 6, heavy drinking and smoking have been found to be important risk factors for morbidity and mortality. Religious attendance is associated with greater peace of mind, contentment, optimism, sense of purpose, and a fatalistic acceptance of adverse outcomes. These all are known to contribute to better mental health and greater mental well-being. Religious attendance may positively affect cognitive functioning in addition to

[11] Religion as a social institution may be defined as encompassing social relationships, physical and social structures, and rules of ethical behavior linked to sets of beliefs about humans' relation to the supernatural. It is operationalized as a survey concept in terms of attendance at religious services, membership in a religious group, and religiosity, i.e., having a strong set of religious beliefs.

emotional health (Van Ness and Kasl 2003). Like emotional ill-health, cognitive dysfunction is associated with an increased risk of mortality.

Use of Small-Area Data for Analysis of Health Variations

I treat this subject summarily here; additional attention is given to it in Chaps. 10 and 16. Vast numbers of studies have been made describing the variations in mortality among the primary and secondary geographic divisions of countries (e.g., Barbi et al. 2003; Langford and Bentham 1996; Luy 2004). Such studies may deal with interesting analytic issues beyond simple geographic comparisons. Barbi et al., for example, considered whether mortality differences among elders observed in various Italian regions may be linked to different mortality levels at earlier ages and whether any associations observed may be explained by different selection mechanisms.

Subnational Variations in Mortality According to Gender, Race, and Ethnicity

The subnational studies of mortality usually provide detail for males and females. In countries with notable differences in the mortality of the sexes at the national level, these differences are reflected in the subnational areas, with the subnational figures for males and females tending to vary in nearly the same pattern. When the national difference between the sexes is negligible and the regional total figures fluctuate greatly, the male and female figures for the regions tend to fluctuate together and differ little from one another. An example of the latter situation is the case of India (Table 7.9). In 1999–2003 life expectancies at birth were above 70 in some Indian states such as Kerala (74) and below 60 in others such as Madya Pradesh (57) and Orissa (59). The figures for the selected states given in the table, both for males and females, show considerable variation, but the male and female figures for each state are close to one another (with one exception) and move in "sync" from one state to another.

Most of the studies on subnational variations of mortality according to race have been done in the United States because few foreign censuses and surveys tabulate data on race. Such data are illustrated here with life expectancies for the states of the United States for race groups in 1989–1991 (Table 7.10). Note that there is substantial variation in life expectancies when the all-races figures for states are considered. This variation is somewhat reduced when the variation among geographic divisions is examined. However, most of the variation in the state values can be explained by the differences in life expectation between the races within the states. In other words, the dispersion of the figures around their national averages is modest when the races are considered separately. Such a finding has an important implication for making estimates and projections of the population of the states and

Table 7.9 Life expectancy at birth in years, for India and selected states of India, by sex: 1999–2003

State	Both sexes	Male	Female
India	62.7	61.8	63.5
Kerala	73.6	70.9	76.0
Madya Pradesh	57.1	57.2	56.9
Maharashtra	66.4	65.2	67.6
Orissa	58.7	58.6	58.7
Punjab	68.6	67.6	69.6
Rajasthan	61.3	60.7	61.8
Tamil Nadu	65.4	64.3	66.5
Uttar Pradesh	59.3	59.6	58.7

Source: Haub, C., & Sharma, O. P. (2006). India's population reality: Reconciling change and tradition. *Population Bulletin*, *61*(3). Washington, DC: Population Reference Bureau; Primary source: Registrar General, India (2006), 2006: statement 2

Table 7.10 Life expectancy at birth, by race, for each state in the United States: 1989–1991

	Total[a]	White	All other	Black
United States	75.37	76.13	71.25	69.16
Northeast				
New England				
Maine	76.35	76.35	68.99	68.62
New Hampshire	76.72	76.68	–	–
Vermont	76.54	76.50	–	–
Massachusetts	76.72	76.90	75.08	72.45
Rhode Island	76.54	76.80	–	–
Connecticut	76.91	77.44	72.31	70.84
Middle Atlantic				
New York	74.68	75.61	71.53	69.33
New Jersey	75.42	76.46	70.73	68.47
Pennsylvania	75.38	76.15	69.34	68.27
Midwest				
East North Central				
Ohio	75.32	75.93	70.86	70.15
Indiana	75.39	75.82	70.76	69.80
Illinois	74.90	76.16	69.25	67.46
Michigan	75.04	76.18	69.22	68.49
Wisconsin	76.87	77.18	72.37	70.96
West North Central				
Minnesota	77.76	77.97	73.05	–
Iowa	77.29	77.38	–	–
Missouri	75.25	76.02	69.65	68.81
North Dakota	77.62	77.99	–	–
South Dakota	76.91	77.91	–	–

(continued)

Table 7.10 (continued)

	Total[a]	White	All other	Black
Nebraska	76.92	77.21	71.14	–
Kansas	76.76	77.07	72.77	71.22
South				
South Atlantic				
Delaware	74.76	75.75	70.06	69.26
Maryland	74.79	76.30	70.76	69.69
District of Columbia	67.99	76.09	64.97	64.44
Virginia	75.22	76.34	71.17	70.05
West Virginia	74.26	74.37	71.20	69.75
North Carolina	74.48	75.89	69.83	69.38
South Carolina	73.51	75.33	69.09	68.82
Georgia	73.61	75.24	69.21	68.79
Florida	75.84	76.82	69.82	68.77
East South Central				
Kentucky	74.37	74.65	70.79	70.16
Tennessee	74.32	75.27	69.43	68.97
Alabama	73.64	75.01	69.59	69.23
Mississippi	73.03	74.78	69.54	69.41
West South Central				
Arkansas	74.33	75.20	69.63	68.93
Louisiana	73.05	74.87	68.99	68.62
Oklahoma	75.10	75.21	74.81	70.85
Texas	75.14	75.75	71.25	69.79
West				
Mountain				
Montana	76.23	76.72	–	–
Idaho	76.88	76.89	–	–
Wyoming	76.21	76.34	–	–
Colorado	76.96	77.06	75.71	72.41
New Mexico	75.74	76.08	73.41	–
Arizona	76.10	76.42	72.76	70.84
Utah	77.70	77.77	–	–
Nevada	74.18	74.44	72.74	–
Pacific				
Washington	76.82	76.92	76.09	71.34
Oregon	76.44	76.51	75.24	–
California	75.86	75.92	75.79	69.65
Alaska	74.83	75.83	71.67	–
Hawaii	78.21	77.92	78.40	–

Source: Adapted from U.S. National Center for Health Statistics (1999)

States are listed according to geographic division

–Figure does not meet standards of reliability or precision (based on fewer than 20 events)

[a]Includes other nonwhite races in addition to blacks. In Hawaii such groups constitute a substantial share of the population and there are few blacks

their subareas, namely that much of the state variation in mortality can be captured by using separate national life tables for the races in designing the calculations for the states.

Studies of mortality variations for subnational areas based on "ethnic nationality" or language can be found more commonly in Europe and Canada. For example, Saarela and Finnäs (2006) analyzed mortality variations of older persons for regions of residence and regions of birth in Finland through a study of the mortality of the two principal ethnic groups in Finland, the Swedish-speaking and Finnish-speaking residents. After restricting their analysis to regions with both Swedish-speaking and Finnish-speaking settlements, they found a mortality advantage for the former group. Even after controlling for region of birth and a number of other covariates, they found that a mortality advantage for the Swedish-speaking group remained. They hypothesize that this advantage is due in important part to circumstances operating before or in childhood rather than to present-day environmental conditions.

Type-of-Residence Area and Health Variations

Mortality differences for type-of-residence areas (i.e., urban and rural areas) in the United States can be considered most effectively by assigning each county to one of the following categories:

Metropolitan area[a]
Central county
Other county
Nonmetropolitan area[a]
Urban county
Rural-nonfarm county
Rural-farm county

[a]Metropolitan areas are composed of counties, with few exceptions.

Death statistics are readily available for counties from NCHS and state departments of health, and the counties can be assigned to type-of-residence classes on the basis of decennial census data. The interpretation of the data is complicated by the differences among counties in such characteristics as age and sex composition, race-ethnic composition, and availability of health services and transportation and communication facilities.

Rural areas in the United States are generally disadvantaged in comparison with urban areas in a number of respects that can have an impact on health levels. The populations of rural areas tend to be older, poorer, and less educated, and have less access to health services. Access is limited in rural areas largely because fewer health services are available, especially in isolated and mountainous areas and very

small communities. Resources are inequitably distributed between urban and rural areas, and health providers, even general practitioners, are not inclined to work in rural areas. These factors are known to be, or considered to be, associated with poorer health and higher mortality levels. Computer connections and rapid transfer of patients (e.g., via helicopter service) between small hospitals or health agencies in some rural areas and large medical centers have been initiated, but the scope of these programs remains limited.

Much has been written about the relative health of urban and rural areas. These studies have shown a survival advantage for urban residents over rural residents in the United States (e.g., House et al. 2007; Smith et al. 1995; Hayward et al. 1997). Urban-rural differences in social and economic conditions apply generally in other countries, especially the less developed countries, with consequences for the health of urban-rural areas, but few studies have examined urban-rural differences in mortality in the less developed areas. The available studies also tend to show that, currently, urban residence in the less developed countries brings with it a survival advantage (e.g., National Research Council 2003; Zimmer et al. 2007). The Zimmer et al. study found, for example, that rural mortality in China was 30% higher than urban mortality among adults aged 50 and older after controlling for age and gender. These authors attribute this huge mortality gap to differences in economic and social conditions, as illustrated by poorer health services and the inadequate investment in rural areas, and the "cadre" or privileged status, higher wage structure, and greater number of amenities in the cities.

Small-Areas as Observation Units for Analysis of Variations in Health

An alternative approach to the use of microdata from sample surveys in the analysis of the relation of socioeconomic status and health for small areas is to use aggregate data for small areas. In other words, small geographic areas, such as the 3,200 or so counties or county equivalents in the United States, or the census tracts of a metropolitan area, are taken as the observation units. For these areas we secure data on a health variable (e.g., personal care disability) and a socioeconomic-status variable (e.g., per capita income) from the census. Statistical analysis linking the health-status variable and the socioeconomic-status variable for each county in the country, or each census tract in the selected metropolitan area, is employed to determine the degree of association between these variables. In this case, we are trying to determine if there is an association between per capita income and personal-care disability. The data may also be used to draw inferences about how this association varies from one part of the country to another, or from one part of the metropolitan area to another.

Consider a variation in the scenario just described. Medicare data on claims are available for the counties from Medicare records, but socioeconomic-status data are lacking from this source. The latter can be imported from the census, and each county can be coded as belonging to one of a half dozen or so per-capita-income categories. Thus, each county is assigned a particular socioeconomic status.

The use of small geographic areas as units of observation to make inferences about the characteristics of individuals, called ecological correlation, is not uncommon in social science analysis because of its advantages, but serious concerns may be raised about the validity of the findings. Using aggregated census data may greatly simplify the task of accessing the data and reduce vastly the volume of data that have to be managed in the analysis. Such aggregated data also represent directly a much larger number of individuals than the national sample surveys. The magnitude of the task may be reduced even further by taking a sample of the geographic units. No matching of records for individuals is required, as when microdata from different sources are used. On the other hand, this method can sometimes give misleading results for individuals. The likelihood of such an outcome is diminished to the extent that the number of observation units is increased, the size of the units is decreased, and the homogeneity of the population within the units with respect to the variables being correlated is maximized. Further information on ecological correlation is given in Appendix B.

Variation in Quality of Health Care and Access to Health Care

Nominally most Americans have public or private health insurance. Nearly all persons 65 years of age and over have Medicare, the federal program of health insurance for the elderly, the poorest segment of the population has Medicaid or Children's Health Insurance, the state-administered health programs for the poor, and millions secure private insurance through their employer or otherwise. However, tens of millions persons under age 65, mostly of the "middle" classes, do not have any health insurance and the health insurance of most of the insured population is severely limited and not portable. Access to health care proves to be unequal for the members of our population. Some nonwhite racial groups, the Hispanic population, persons of lower socioeconomic status, and the rural population report problems with access to health care (U.S. AHRQ 2005a). These same groups are less likely to have a usual source of care. They use preventive services less than other groups. These groups are less likely to have taken health screening tests, such as those for cancer and cardiac risk factors, and to have received recommended childhood or other immunizations and dental care. Rates of hospital admission for conditions that are usually treatable with ambulatory care are generally higher for persons who live in low income areas compared with those who live in high income areas. Such persons are likely to have higher hospitalization rates for chronic conditions (e.g., diabetic complications, asthma). On the other hand, for most quality-of-care measures Asians report a quality of care similar to or better than that reported by whites.

The other Western industrialized countries in general do not have patterns of socioeconomic differences in access to health care and quality of treatment like the United States because they have universal health care. They still show socioeconomic differences in the levels of health and mortality, however, because of

differences among the classes in lifestyle and behavior. The situation in the United States with respect to access is more like that in the Less Developed Countries (LDC) that also lack systems of universal health care. Like the poor in the United States, the poor in the LDC have less access to health care and are more likely to receive late and less effective treatment. (See Chap. 11 for further discussion of this subject.)

Concluding Note

Several methods have been developed for dealing with relationships between health, on the one hand, and demographic and socioeconomic variables, on the other, in non-experimental (i.e., observational) situations. One is the use of multivariate regression, particularly logistic regression or, when economic variables are involved, econometric models. To analyze longitudinal event data, an extension of logistic regression, namely proportional hazard models are applied. Other methods are stochastic model-building and simulation. These are appropriate where we have theoretical models linking health and other variables, statistical data representing these variables, and electronic computers that can handle numerous variables and complex relationships.

In countries with well developed statistical systems, knowledge of the interrelationships between health and mortality, on the one hand, and demographic, socioeconomic, and other factors, on the other, has grown tremendously in the last half century. In spite of the complexity of the relationships, often involving numerous variables, the lack of the required data on certain variables, and the tendency of some variables to confound the role of others, the development of new statistical methods and of powerful electronic computers and the considerable attention devoted to this effort have been productive in clarifying many interrelationships involving health variables. Nevertheless, it will always be difficult to prove causal relationships even where associations between variables are strong (Moffitt 2005).

References and Suggested Readings

Concepts, Basic Data, and Measurement of Group Differences

Adler, N. E. (2006). Appendix D, Overview of health disparities. In National Research Council, Board of Health Sciences Policy, *Examining the Health Disparities Research Plan of the National Institutes of Health: Unfinished Business* (pp. 121–174). Washington, DC: The National Academies Press.

Carter-Pokras, O., & Baquet, C. (2002). What is 'health inequality'? *Public Health Reports*, *117*(5), 426–434.

Ferraro, K. F., & Shippee, T. P. (2009). Aging and cumulative inequality: How does inequality get under the skin? *Gerontologist*, *49*(3), 333–343.

Harper, S., Lynch, J., Meersman, S.C., Breen, N., et al. (2008). An overview of methods for monitoring social disparities in cancer, with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992–2004. *American Journal of Epidemiology*, *167*(8), 889–899.

Jackson, P. B. (2005). Health inequalities among minority populations. *Journal of Gerontology Series B, 60B* (Special Issue II), 63–67.

Moffitt, R. (2005). Remarks on the analysis of causal relationships in population research. *Demography*, *42*(1), 91–108.

National Academy on an Aging Society. (2003). Disparities in health status and health care. *Public Policy and Aging Report 13*(3). Articles by K. E. Whitfield, L. M. Burton, M. Hayward, J. L. Angel, R. J. Angel, R. B. Hudson, and T. A. LaVeist. Washington, DC: Gerontological Society of America.

Pamuk, E., Lentzner, H., & Brackbill, R. (1993). Measuring socioeconomic inequality in health: An update on methodological issues. In *Proceedings of the 1993 Public Health Conference on Records and Statistics*. Hyattsville, MD: U.S. National Center for Health Statistics.

Pan-American Health Organization (PAHO). (2002). *Health in the Americas*. Washington, DC: Pan-American Health Organization.

Population Reference Bureau. (2005). *World population data sheet, 2005*. Washington, DC: Population Reference Bureau.

Regidor, E. (2004a). Measures of health inequalities: Part 1. *Journal of Epidemiology and Community Medicine, 58*, 858–861.

Regidor, E. (2004b). Measures of health inequalities: Part 2. *Journal of Epidemiology and Community Medicine, 58*, 900–903.

Rogers, R. G., Hummer, R. A., & Nam, C. B. (2000). *Living and dying in the USA*. San Diego, CA: Academic.

Scotland, Government of. (2007). *Measuring socio-economic inequality in health: A practical guide*. By R. Muñoz-Arroyo, & M. Sutton. Edinburgh, UK: Scottish Public Health Observatory Collaboration, Public Health Information for Scotland.

U.S. Agency for Healthcare Research and Quality (AHRQ). (2005a). *The national healthcare disparities report*. Rockville, MD: U.S. Agency for Healthcare Research and Quality.

U.S. Agency for Healthcare Research and Quality (AHRQ). (2005b). *The national healthcare quality report*. Rockville, MD: U.S. Agency for Healthcare Research and Quality.

U.S. Agency for Healthcare Research and Quality (AHRQ). (2008). *The national healthcare disparities report*. Rockville, MD: U.S. Agency for Healthcare Research and Quality.

U.S. National Center for Health Statistics. (2000). *Healthy people, 2010*. Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2002). Summary health statistics for U.S. adults: National Health Interview Survey: 1998. By J.R. Pleis, & R. Coles. *Vital and Health Statistics, 10*(209).

U.S. National Center for Health Statistics. (2002). U.S. life tables, 2000. By E. Arias. *National Vital Statistics Reports, 51*(3).

U.S. National Center for Health Statistics. (2006). *Health, United States, 2006: With chart book on trends in the health of Americans*. Hyattsville, MD. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2007). U.S. life tables, 2004. By E. Arias. *National Vital Statistics Reports, 56*(9).

U.S. National Center for Health Statistics. (2008). Deaths: Final data for 2005. By H. -C. Kung, D. L. Hoyert, J. Xu, & S. L. Murphy. *Vital and Health Statistics, 56*(10).

U.S. National Center for Health Statistics. (2009). Deaths: Leading causes for 2005. By M. Heron, & B. Tejada-Vera. *National Vital Statistics Reports, 58*(8).

Zarit, S., Pearlin, L. I., & Hendricks, J. (Eds.). (2005). *Health Inequalities Across the Life Course. Journal of Gerontology: Series B, Psychological and Social Sciences*, Special Issue II, Vol. *60*B.

Sex Differences

Ahlburg, D., & Vaupel, J. (1990). Alternative projections of the U.S. population. *Demography, 27*(4), 639–652.

Aviv A., Sha., J., Christensen, K., & Wright, W. (2005). The longevity gender gap: an telomeres the explation?, *Sci. Aging Knowledge Environ.* (2005)(23), p.16.

Carey, J. R. (2002). The importance of including biodemography in the demography curriculum. *Genus, 58*(3–4), 189–200.

Carey, J. R., & Judge, D. S. (2001). Principles of biodemography with special reference to human longevity. *Population: An English Selection, 13*, 9–40. Paris: INED.

Frankenhaueser, M., Wright, M. V., Collins, A. et al. (1978). Sex differences in psychoneuroendrocrine reactions to examination stress. *Psychosomatic Medicine, 40*, 334–343.

Frankenhaueser, M., Lundberg, U., & Chesney, M. (Eds.). (1991). *Women, work, and health: stress and opportunities*. New York: Plenum Press.

Haub, C., & Sharma, O. P. (2006). India's population reality: Reconciling change and tradition. *Population Bulletin*, *61*(3). Washington, DC: Population Reference Bureau; Primary source: Registrar General, India (2006), 2006: statement 2

Harris, B. (Ed.). (1989, April). Differential female mortality and health care in South Asia. *Journal of Social Studies, 44*, 1–123.

Holliday, R. (1997). *Aging: The paradox of life – Why we age*. Dordrecht, the Netherlands: Springer.

Lee, R., & Tuljapurkur, S. (1994). Stochastic population projections for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association, 89*(428), 1175–1189.

Lopez, A. D., & Ruzicka, L. T. (Eds.). (1983). *Sex differentials in mortality: Trends, determinants and consequences*. Canberra, Australia: Department of Demography, Australian National University.

Luy, M. (2003). Causes of male excess mortality: Insights from cloistered populations. *Population and Development Review, 29*(4), 647–678.

Madigan, F. C. (1957). Are sex mortality differentials biologically caused? *Milbank Memorial Fund Quarterly, 35*(2), 202–203.

Moen, P., Dempster-McCain, D., & Williams, R.M. (1989). Social integration and longevity: An event history analysis of women's roles and resiliency. *American Sociological Review, 54*(4), 635–647.

Ory, M. G., & Warner, H. R. (Eds.). 1990. *Gender, health, and longevity: Multidisciplinary perspectives*. New York: Springer.

Patrick, C. H., Palesch, Y. Y., Feinleib, M., & Brody, J. A. (1982). Sex differences in declining cohort death rates from heart disease. *American Journal of Public Health, 72*, 161–166.

Ramey, E. R. (1982). The natural capacity for health in women. In P. W. Berman & E. R. Ramey (Eds.), *Women, a developmental perspective*. Washington, DC: U.S. National Institutes of Health.

Retherford, R. D. (1972). Tobacco smoking and the sex mortality differential. *Demography, 9*(2), 203–216.

United Nations. (1983). Patterns of sex differentials in mortality in less developed countries. In A. D. Lopez & L. T. Ruzicka (Eds.), *Sex differentials in mortality: Trends, determinants and consequences* (pp. 7–32). Canberra, Australia: Department of Demography, Australian National University.

U.S. Census Bureau. (2009). U.S. Population projections: Methodology. United States population projections by age, sex, race, and Hispanic orgin: July 1, 2000–2050. Accessed July 2009, from www.census.gov./population/www/projections

U.S. Social Security Administration, Actuary's Office. (2008). Accessed April 28, from www.ssa, gov/OACT/index.html

Verbrugge, L. (1987). Sex differences in health behavior, morbidity, and mortality. In *NIH conference on gender and longevity, September 17–18, 1987*. Bethesda, MD: National Institute on Aging and National Institute of Child Health and Human Development.

Verbrugge, L. M., & Madans, J. H. (1985). Social roles and health trends of American women. *Milbank Memorial Fund Quarterly*: Health and Society, *63*(4), 691–735.

Waldron, I. (1976). Why do women live longer than men? Part I. *Journal of Human Stress*, *2*(1), 2–13.

Waldron, I. (1983). Sex differences in human mortality: The role of genetic factors. *Social Science and Medicine, 17*(6), 321–333.

Waldron, I. (1993). Recent trends in sex mortality ratios for adults in developed countries. *Social Science and Medicine, 36*, 451–462.

Waldron, I., & Johnston, S. (1976). Why do women live longer than men? Part II. *Journal of Human Stress, 2*(2), 19–30.

Race and Hispanic Ethnicity

Abraido-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The Latino mortality paradox: A test of the 'salmon' bias' and the healthy migrant hypotheses. *American Journal of Public Health, 89*(10), 1543–1548.

Coale, A. J., & Kisker, E. E. (1986). Mortality crossover: Reality or bad data? *Population Studies, 40*, 389–401.

Cooper, R. S., & David, R. (1986). The biological concept of race and its application to public health and epidemiology. *Journal of Health and Politics, Policy and Law, 11*, 97–116.

Crimmins E. M., Kim, J. K., Alley, D. E, et al. (2007). Hispanic paradox in biological risk profiles. *American Journal of Public Health, 97*(7), 305–310.

Elo I. T., Turra, C. M., Kestenbaum, B., & Ferguson, R. F. (2004). Mortality among elderly Hispanics in the United States: Past evidence and new results. *Demography, 41*, 109–128.

Geiger, H.J, (2001). A review of the evidence: Summary of the peer-reviewed literature on racial and ethnic disparities in medical care. Presentation to the Physicians for Human Rights Distinguished Panel on Racial Disparities in Care, February 7, 2001, Washington, D.C.

Geronomus, A. T., Bound, J., Waldmann, T. A., Colen, C. G., & Steffick, D. (2001). Inequalities in life expectancy, functional status, and active life expectancy across selected black and white populations in the United States. *Demography, 38*, 227–251.

Hummer, R. A., Benjamins, M. R., & Rogers, R. G. (2004). Racial and ethnic disparities in health and mortality among the U.S. elderly population. In R. A. Bulatao, & N. B. Anderson (Eds.), *Understanding racial and ethnic differences in health in late life*: *A research agenda.* (pp. 35–44). Washington, DC: National Academy Press.

Johnson, N.E. (2000). The racial crossover in comorbidity, disability, and mortality. *Demography, 37*(3), 267–284.

Kestenbaum, B. 1992. A description of the extreme aged population based on improved Medicare enrollment data. *Demography, 29*(4), 565–580.

Manton, K. G., & Stallard, E. (1981). Methods for evaluating the heterogeneity of aging processes in human populations using vital statistics data: Exploring black/white mortality crossover by a model of mortality selection. *Human Biology, 53*, 47–67.

Manton, K. G., & Stallard, E. (1997). Health and disability differences among racial and ethnic groups. In L. G. Martin, & B. J. Soldo (Eds.), *Racial and ethnic differences in the health of older Americans* (pp. 43–104). Washington, DC: National Academy Press.

Markides, K. S., & K. Eschbach. (2005). Aging, migration, and mortality: Current status of research on the Hispanic Paradox. *Journal of Gerontology: Series B, 60B* (Special Issue II): 68–75.

Nam, C. B. (1995). Another look at mortality crossovers. *Social Biology, 42*(1–2), 133–142.

National Research Council. (1996). *Changing numbers, changing needs: American Indian demography and public health*. G. D. Sandefur, R. R. Rindfus, & B. Cohen (Eds.). Washington, DC: National Academy Press.
National Research Council, Institute of Medicine. (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Smedley, B. D., Stith, A. Y., & Nolson, A. R. (Eds.) Washington, DC: National Academies Press.
National Research Council. (2004a). *Understanding racial and ethnic differences in health in later life: A research agenda*. R. A. Bulatao & N. B. Anderson (Eds.). Washington, DC: National Academies Press.
National Research Council. (2004b). *Critical perspectives on racial and ethnic differences in health in later life*. N. B. Anderson, R. A. Bulatao, & B. Cohen (Eds.). Washington, DC: National Academies Press.
National Research Council, Board of Health Sciences Policy. (2006). *Examining the health disparities research plan of the National Institutes of Health: Unfinished business*. Washington, DC: The National Academies Press.
Palloni, A., & Arias, E. (2004). Paradox lost: Explaining the Hispanic adult mortality advantage. *Demography, 41*, 385–415.
Preston, S. H., Rosenwaike, L., & Hill, M. E. (1996). African-American mortality at older ages: Results of a matching study. *Demography, 33*, 193–209.
U.S. Agency for Healthcare Research and Quality (AHRQ). (2005a). *The national healthcare disparities report*. Rockville, MD: U.S. Agency for Healthcare Research and Quality.
U.S. National Center for Health Statistics (NCHS). (1999). Quality of death rates by race and Hispanic origin: A summary of current research. By H. M. Rosenberg, J. D. Maurer, P. D. Sorlie, et al. *Vital and Health Statistics, 2*, 128.
U.S. National Center for Health Statistics. (2004). Mean body weight, height, and body mass index, United States 1960–2002. By C. L. Ogden, C. D. Fryar, M. D. Carroll, & K. M. Flegal, *Advance Data from Vital and Health Statistics*, 347.
U.S. National Center for Health Statistics. (2009). Prevalence of metabolic syndrome among adults 20 years of age and over, by sex, age, race and ethnicity, and body mass index: United States, 2003–2006. By R. B. Ervin. *National Health Statistics Reports*, 13.
White-Means, S. I. (2000). Racial patterns in disabled elderly persons' use of medical services. *Journal of Gerontology: Social Sciences, 55B*(2), S76–S89.

Marital Status and Family Arrangements

Christakis, N., & Iwashyna, T. J. (2003). The impact of health care on families: A matched cohort study of hospice use by decedents and mortality outcomes in surviving widowed spouses. *Social Science and Medicine, 57*, 465–475.
Grabill, W. G. (1945). Attrition life tables for the single population. *Journal of American Statistical Association, 40*, 364–375.
Gove, W. (1973). Sex, marital status, and mortality. *American Journal of Sociology, 79*(1), 45–67.
Hu, Y., & Goldman, N. (1990). Mortality differentials by marital status: An international comparison. *Demography, 27*, 233–250.
Lillard, L. A., & Panis, C. W. A. (1996). Marital status and mortality: The role of health. *Demography, 33*, 313–327.
Lillard, L. A., & Waite, L. J. (1995). Til death do us part: Marital disruption and mortality. *American Journal of Sociology, 100*, 1131–1156.
Martikainen, P., Martelin, T., Nihtilä, E., Majamaa, K., & Koskinen, S. (2005). Differences in mortality by marital status in Finland from 1976 to 2000: Analyses of changes in marital-status distributions, socio-demographic composition, and cause of death. *Population Studies, 59*(1), 99–115.

Pruchno, R. A., & Resch, N. L. (1989). Husbands and wives as caregivers. *Gerontologist, 29*, 159–165.
Schaefer, C., Quesenberry, C. P., & Soora, W. (1995). Mortality following conjugal bereavement and the effects of a shared environment. *American Journal of Epidemiology, 141*, 1142–1152.
U.S. National Center for Health Statistics. (1956). *Vital Statistics – Special Reports, 39*(7).
U.S. National Center for Health Statistics. (2004). Marital status and health: United States, 1999–2002. By C. A. Schoenborn (Ed.) *Vital and Health Statistics: Advance Data*, No. 351.
U.S. National Center for Health Statistics. (2006a). Deaths: Final data for 2003. By D. L. Hoyert, M. P. Heron, S. L. Murphy, & H-C. Kung. *Vital and Health Statistics, 54(*13).
Waite, L. J., & Gallagher, M. (2000). *The case for marriage*. New York: Doubleday.
Willekens, F. J., Shah, I., Shah, J. M., Ramachandran, P. (1982). Multi-state analysis of marital status life tables: Theory and application. *Population Studies, 36*, 129–144.

Socioeconomic Status

Adams, P., Hurd, M. D., McFadden, D., Merrill, A., & Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics, 112*(1), 3–56.
Adler, N. E., Marmot, M., McEwen, B. S., & Stewart, J. (Eds.). (1999). *Socioeconomic status and health in industrial nations*. New York: New York Academy of Sciences.
Attanasio, O. P., & Hoynes, H. W. (2000). Differential mortality and wealth accumulation. *Journal of Human Resources, 35*(1), 1–29.
Banks, J., Marmot, M., Oldfield, Z., & Smith, J. P. (2006). Disease and disadvantage in the United States and England. *Journal of the American Medical Association, 295*, 2037–2045.
Crimmins, E. M. (2005). Socioeconomic differentials in mortality and health at the older ages. *Genus, 61*(1),163–176.
Crimmins, E. M., & Seeman, T. E. (2004). Integrating biology into the study of health disparities. In L. J. Waite (Ed.), *Aging, health, and public policy: Demographic and economic perspectives, Supplement to population and development review* (Vol. *30*, pp. 89–107). New York: Population Council.
Duleep, H. O. (1995). Mortality and income inequality among economically developed countries. *Social Security Bulletin*, 58, 34–50.
Elo, I. T., & Preston, S. H. (1996). Educational differentials in mortality: United States, 1979–1985. *Social Science and Medicine, 42*, 47–57.
Feinglass, J., Lin, S., Thompson, J., et al. (2007). Baseline health, socioeconomic status, and 10-year mortality among older middle-aged Americans: Findings from the Health and Retirement Study, 1992–2002. *Journal of Gerontology: Social Sciences, 62B*(4), 209–217.
Feldman, J. J., Makuc, D. M., Kleinman, J. C., & Cornoni-Huntley, J. (1989). National trends in educational differentials in mortality. *American Journal of Epidemiology, 129*(5), 919–933.
Firebaugh, G. (1999). Empirics of world income inequality. *American Journal of Sociology, 104*, 1597–1630.
George, L. K. (2005). Socioeconomic status and health across the life course: Progress and Prospects. *Journal of Gerontology: Series B, 60B* (Special Issue II), 117–124.
Glied, S., & Lleras-Muney, A. (2008). Technological innovation and inequality in health. *Demography, 45*(3), 741–761.
Goldman, D., & Smith, J. P. (2002). Can patient self-management explain the SES health gradient? *Proceedings of the National Academy of Sciences, 99*(16), 10929–10934.
Grundy, E., & Holt, G. (2002). The socioeconomic status of older adults: How should we measure it in studies of health inequalities? *Journal of Epidemiology and Community Health, 55*(12), 895–904.
Heisler, M., Langa, K. M., Eby, E. L., Fendrick, A. M., et al. (2004). The health effects of restricting medication use because of cost. *Medical Care, 42*(7), 526–634.

Hoffman, R. (2005). Does the impact of socioeconomic status on mortality decrease with increasing age? *Journal of Gerontology: Social Sciences, 62B*(4), S209–217.

Hurd, M., McFadden, D., & Merrill, A. (1998). Healthy, wealthy, and wise? *Socioeconomic status, morbidity, and mortality among the elderly* (Working paper), Cambridge, MA: National Bureau of Economic Research.

Mackenbach, J. P., Stirbu, I., Roskam, A-J. R., Schapp, M. M. et al. (2008). Socioeconomic inequalities in health in 22 European countries. *New England Journal of Medicine, 358*(23), 2468–2481.

Marmot, M. G. (2004). *The status syndrome: How social standing affects our health and longevity*. New York: Henry Holt.

Marmot, M. G., & Shipley, M. J. (1996). Do socioeconomic differences in mortality persist after retirement? 25-year follow-up of civil servants from the first Whitehall Study. *British Medical Journal, 313*, 1177–1180.

Marmot, M.G., Shipley, M., & Rose, G. (1984). Inequalities in death: Specific explanations of a general pattern? *Lancet, 321*, 1003–1006.

Marmot, M. G., Bobak, N., & Davey Smith, G. (1995). Explanations for social inequalities in health. In B. C. Amick, S. Levine, A. Tarlov, & D. C. Walsh (Eds.), *Society and health* (pp. 172–210). Oxford, UK: Oxford University Press.

Martikainen, M., Lahelma, E., Ripatti, S., Albanes, D., & Virtamo, J. (2000). Educational differences in lung cancer mortality in male smokers. *International Journal of Epidemiology, 30*, 264–267.

Moore, E., & Hayward, M. (1990). Occupational careers and mortality of elderly men. *Demography, 27*(1), 31–53.

Pappas, G., Queen, S., Hadden, W., & Fisher, G. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *New England Journal of Medicine, 329*, 103–109.

Preston, S., & Taubman, P. (1994). Socioeconomic differences in adult mortality and health status. In L. G. Martin & S. H. Preston (Eds.), *Demography of aging* (pp. 279–318). Washington, DC: National Academy Press.

Rodgers, G. B. (1979). Income and inequality as determinants of mortality: An international cross-section analysis. *Population Studies, 33*(2), 343–351.

Seeman, T. E., & Crimmins, E. (2001). Social environment effects on health and aging: Integrating epidemiologic and demographic approaches and perspectives. *Annals of the New York Academy of Medicine, 954*(1), 88–117.

Smith, J. P. (1998). Socioeconomic status and health. *American Economic Review, 88*(2), 192–196.

Smith, J. P. (2004). Unraveling the SES – health connection. In L. J. Waite (Ed.), *Aging, health, and public policy: Demographic and economic perspectives*, Supplement to *Population and Development Review* (Vol. 30, pp. 108–132). New York: Population Council.

Smith, J. P., & Kington, R. (1997). Demographic and economic correlates of health in old age. *Demography, 34*(1), 159–170.

Tietze, C. (1943). Life tables for social classes in England. *Milbank Memorial Fund Quarterly, 21*(2), 1–6.

U. S. Census Bureau. (2006). *Current Housing Reports, Series* 150/03 (Annual Housing Survey for the United States).

U.S. National Center for Health Statistics. (2005). Seroprevalence of six infectious diseases among adults in the United States by race/ethnicity: Data from the third National Health and Nutrition Examination Survey, 1988–94. By D. Kruszon-Moran, & M. McQuillan. *Vital and Health Statistics: Advance Data*, No. 352.

U.S. National Institute on Aging. (1999). Socioeconomic status, health, and longevity. *Research Highlights, Demography and Economics of Aging*. Issue No. 4.

Waite, L. T. (Ed.). (2004). *Aging, health, and public policy: Demographic and economic perspectives. A Supplement to Population and Development Review* (Vol. *30*). New York: Population Council.

Religious Participation

Ellison, C., & Levin, J. (1998). The religion-health connection: Evidence, theory, and future direction. *Health Education and Behavior, 25*, 700–725.
George, L. K. (2002). The links between religion and health: Are they real? *Public Policy and Aging Report, 12*(4). Washington, DC: National Academy on an Aging Society.
Hummer, R., Rogers, R., Nam, C., & Ellison, C. (1999). Religious involvement and U.S. adult mortality. *Demography, 36*, 273–285.
Idler, E. (2002). *The many causal pathways linking religion to health* (Public policy and aging report). Washington, DC: National Academy on an Aging Society, Gerontological Society of America.
Idler, E., & Kasl, S. (1992). Religion, disability, depression, and the timing of death. *American Journal of Sociology, 97*, 1052–1097.
Idler, E., & Kasl, S. (1997). Religion among disabled and non-disabled persons. II. Attendance at religious services as a predictor of the course of disability. *Journal of Health and Social Behavior, 38*, 21–37.
Van Ness, P., & Kasl, S. (2003). Religion and cognitive dysfunction in an elderly cohort. *Journal of Gerontology: Social Sciences, 58B*, S21–29.

Regional and Type-of-Residence Comparisons

Barbi, E., & Caselli, G. (2003). Selection effects on regional differences in survivorship in Italy. *Genus, 49*(2), 37–62.
Elman, C., & Myers, G.C. (1999). Geographic morbidity differentials in the late nineteenth-century United States. *Demography, 36*(4), 429–444.
Hayward, M. D, Pienta, A. M., & Mclaughlin, D. K. (1997). Inequality in men's mortality: The socioeconomic status gradient and geographic context. *Journal of Health and Social Behavior, 38*, 313–330.
House, J. S., Lepkowski, J. M., Williams, D. R., et al. (2007). Excess mortality among urban residents: How much, for whom, and why? *American Journal of Public Health, 90*, 1898–1904.
Langford, I. H., & Bentham, G. (1996). Regional variations in mortality rates in England and Wales: An analysis using multi-level modeling. *Social Science and Medicine, 42*, 897–908.
Luy, M. (2004). Mortality differences between Western and Eastern Germany before and after reunification: A macro and micro level analysis of developments and responsible factors. *Genus, 60*(3–4), 99–141.
National Research Council. (2003). *Cities transformed: Demographic change and its implications in the developing world*. Washington, DC: National Academies Press.
Registrar General, India. (2006). *SRS based abridged life tables, 1999–2003*.
Saarela, J., & Finnäs, F. (2006). Regional mortality variation in Finland: a study of two populations groups. *Genus, 62*(2), 169–211.
Smith, M. H., Anderson, R.T., Bradham, D.D., & Longino, C.F. (1995). Urban and rural differences in mortality among Americans 55 years and older: Analysis of the national longitudinal mortality study. *Journal of Rural Health, 11*, 274–285.
U.S. National Center for Health Statistics (NCHS). (1999). Some trends and comparisons of United States life table data: 1900–1991. U.S. *Decennial Life Tables for 1989–1991*, *1*(3).
Zimmer, Z., Kaneda, T., & Spess, L. (2007). An examination of urban versus rural mortality in China using community and individual data. *Journal of Gerontology: Social Sciences, 62B*(5), S349–S357.

Chapter 8
Gauging Population Health: Measures Linking Mortality and Morbidity

As countries progress through the more advanced stages of the epidemiologic transition, the traditional indicators of population health, based on mortality rates alone, are expected to change little. Therefore, the incidence, prevalence, duration, and disabling effects of morbid conditions are becoming increasingly important measures in the evaluation of population health. The most informative summary measures of population health, the type developed in this chapter, combine those for mortality and morbidity.

Relation Between Changes in Mortality and Changes in Morbidity

Various scholars, including analysts at the World Health Organization, have graphically modeled the relation between morbidity, disability, and mortality in the form of survival curves. A stylized representation of these relations over the age cycle is shown in Fig. 8.1. More detailed diagrams can insert curves for degrees and types of disability and dysfunctionality between morbidity and death, such as IADL limitations, ADL limitations, and frailty. An individual can skip stages and may not experience any stage other than death, or can experience change in the reverse direction. Typically, however, disease occurs first, then functioning loss or disability, and finally, death.

It has become evident that, to assess population health adequately, it is necessary to develop summary measures of health that combine mortality and morbidity. The issues affecting the measurement of mortality are well understood and well established, but this cannot be said of morbidity. Health status can be defined and measured in a variety of ways, as we saw, and there is no agreement on the relative merits of the various measures. Hence, there is no general agreement on the measures combining mortality and morbidity. The various measures of morbidity, and hence of the combination of morbidity and mortality, give different indications of the level and trend of population health.

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_8, © Springer Science+Business Media B.V. 2012

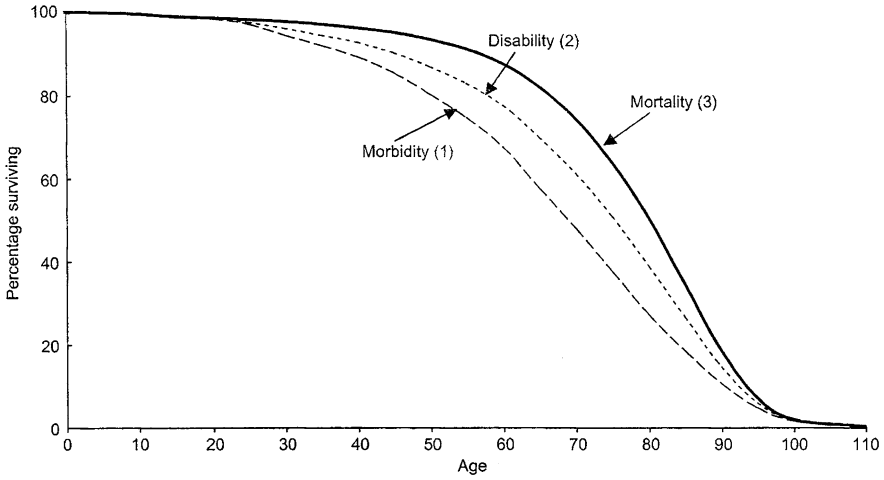


Fig. 8.1 Stylized representation of the relation between the age curves for total survivors, disability-free survivors, and morbidity-free survivors (Note: Percentage of the initial birth cohort who (1) survive to the indicated ages without at least one major chronic disease and without a major disability, (2) survive to the indicated ages without a major disability but with at least one major chronic disease, or (3) survive to the indicated ages with at least one major disability and with at least one major chronic disease)

Measurement of the relation between changes in morbidity and changes in mortality is also complex, therefore. The relation varies in different circumstances – over the age scale, over time, among birth cohorts, and among population groups. More basic is the fact that there are different levels and dimensions of morbidity and, therefore, there is no unitary trajectory of morbidity. The level of ADL-disability can decline while the level of chronic diseases can rise; severe disability may move at a different rate or even direction than moderate disability. Moreover, the trajectories of morbidity and mortality can move in different ways depending on the circumstances.

The initial assumption regarding mortality and morbidity trends has been that the two were closely related, and therefore followed close trajectories over time. Thus, if mortality declines, then morbidity was also thought to decline (Fries 1980). Alter and Riley (1989) and Riley and Alter (1996), however, have demonstrated, using historical data, that morbidity levels tended to rise after mortality declines, rather than the reverse. While this change may have reflected a rise in morbidity prevalence ratios, it also reflected reductions in case-fatality rates. This means that persons saved from dying experienced a more extended period of illness or frailty than would otherwise have been the case (Gruenberg 1977). When the health profile of a population is dominated by chronic conditions at the older ages, rather than infectious diseases, which affect mainly the younger ages, it is possible to have both longer life and deteriorating population health. On the other hand, neither of the theories just suggested, the compression of morbidity or the expansion of morbidity,

may be at play at present. Although chronic diseases are becoming more prevalent, their expression may be becoming milder and the deterioration of health due to chronic disease may be moderating. This has been named the concept of dynamic equilibrium (Manton 1982).

To evaluate the possible relation between mortality and morbidity further, consider a few specific diseases. If progress in the treatment of heart disease results in a decline in death rates from this cause, more people will survive but, at the same time, more may be expected to suffer from the disease. Under these conditions we are likely to observe more cases of the disease in the advanced stages and, as a result, higher levels of morbidity and frailty. Under this scenario the heart-disease prevalence ratio will rise unless the incidence rate for the disease can be successfully reduced through preventive interventions. The association between health change and mortality change can be seen as even more convoluted when we consider the many causes of chronic degenerative morbidity that do not usually result in death. For instance, osteoarthritis is among the leading causes of disability among older people, but it is rarely listed as an underlying cause of death. Even if the age-specific prevalence of osteoarthritis and similar diseases declines, the death rate is hardly affected. Summary measures of health status that combine mortality and morbidity attempt to reflect these more complex aspects of health status.

Measurement of Healthy/Active/Disability-Free Life Expectancy

Summary measures of population health linking morbidity and mortality may be classified into three broad groups, but the groups are not easily distinguished or even mutually exclusive on close examination. The first group includes measures of health expectancy (HE). These measures are a direct extension of conventional life table calculations that merge data on morbidity into the life table. They are variously labeled disability-free life expectancy (DFLE), active life expectancy (ALE), healthy life expectancy (HLE), or health-adjusted life expectancy (HALE). The second group includes measures that are a composite of a life table measure and externally applied health measures. They typically measure the health gap, the difference between the actual health of a population and some preset goal(s) to be achieved for the population‘s health. The group includes measures of quality-adjusted life years (QALY). The specific measures are labeled disability-adjusted life years (DALY), health-adjusted life-years (HALY), and healthy life years (HELY or HeaLY). The third group includes measures not based on life tables or their extensions, and may not even include a mortality component. The first two types of measures now virtually dominate the field of population-health measurement. Accordingly, for this discussion, the third group has been combined with the second. Each of these several measures can have numerous variations because of the different concepts of health, ways of securing data on it, and ways of designing the measure.

To anticipate the later discussion with a general illustration, consider a composite measure that can be classified as a complex variation of the first type of measure. It was developed for *Healthy People* 2000: *Final Review* (U.S. NCHS 2001). It combines a life table measure, a disability measure, and the individual's assessment of his or her own health, with the latter two components being merged and then incorporated into the life table. Thus, a single measure of healthy life expectancy is derived from the life table calculations.

Data on age-specific disability (or other unhealthy state) for the United States can be derived from the National Health Interview Survey (NHIS), the Longitudinal Study on Aging (LSOA), the National Long-Term Care Survey (NLTCS), the Medicare Current Beneficiary Survey (MCBS), the Health and Retirement Study (HRS), the Established Populations for Epidemiologic Studies of the Elderly (EPESE), other similar health surveys, and the decennial census. The surveys are all longitudinal surveys that can provide data for a number of years.

Prevalence-Ratio Method

Much of the analysis linking mortality and morbidity in recent years has involved calculations of tables of active life. As was indicated in Chap. 4, such tables are an extension of the standard life table that allows jointly for deaths and illness. Their main product is known by various names – active, disability-free, independent, and healthy life expectancy – but the choice of the name does not usually indicate precisely the nature of the health data employed in constructing the table. Most tables of this kind have used data for ADL-disabled persons, although they have been called tables of active or healthy life.

Tables of active life can be constructed by any of several methods – the prevalence-ratio method, the event-exposure method, the multistate method, and microsimulation. All of these methods rely on life-table techniques and all are multiple-increment-decrement methods, although they are not all characterized this way. That is, they allow for increments and decrements of the health variables, either implicitly or explicitly. Health changes in the two directions are only implicit in the prevalence-ratio method. The event-exposure method has limited ability to model transitions in both directions explicitly, and with health data, this usually means that the changes in each age are net shifts out of the healthy state. The other two methods can explicitly model all possible transitions separately between the health states under consideration.

The idea of combining the measurement of health and mortality in a single life table was originally proposed by Sanders (1964) and a life-table method for estimating active life expectancy, specifically the prevalence-ratio method, was first developed by Sullivan (1971). These tables provide measures of the average remaining years of independent and dependent life at each age, and the shares of remaining life at each age that are independent and dependent. Tables of independent life for a succession of years can answer questions such as: Has health status, as measured by dependency, improved among the elderly? Has health status improved in tandem

with life expectancy? How are the extra years of life expectancy of women over those of men divided between dependent years and independent years? What is the comparative proportion of independent years for men and women?

The prevalence-ratio method of constructing a table of active life, the most common method used, has a number of advantages over the other methods of measuring mortality and morbidity jointly. First, the required data are relatively easy to secure and the method is relatively simple to apply as compared with constructing tables of active life by the other methods. Next, the composite measure, active life expectancy, does not require any special weighting procedure in its calculation since the parts come from a common source. Like life expectation, the results are unaffected by the size and age distribution of the observed population and are comparable from table to table, although, as explained in Chap. 4 , they are affected by the shift in the internally generated age distributions from table to table.

Method of Calculation

The Sullivan method is a common basis for the HE group of measures and has been widely used to produce estimates of disability-free (or healthy) life expectancy. It employs a combination of age-specific disability (or unhealthy state) prevalence ratios, that is, the percentages of persons in each age interval who are disabled (or unhealthy), and a standard life table (incorporating the mortality rates used for constructing a standard life table).

In constructing a table of active life by the prevalence-ratio method (Table 8.1), the complements of the disability (or unhealthy) prevalence ratios (col. 5) are applied to the L_x function of the basic life table, to obtain the person-years of life lived in a disability-free (or healthy) condition in each age interval (col. 6):

$$L_{x(hs)} = (1 - DPR_x) * L_x \tag{8.1a}$$

$$L_{x(hs)} = (1 - \pi_x) * L_x \tag{8.1b}$$

where the subscript *hs* denotes disability-free status and DPR_x (or π_x) is the age-specific disability (unhealthy state) prevalence ratio. The person-years of disability-free life for each age interval ($L_{x(hs)}$) are summed from age x forward to the end of the life table, to obtain the total person-years that are disability-free from age x to the end of the age cycle ($T_{x(hs)}$; col. 7):

$$T_{x(hs)} = \sum_{a=x}^{a=\infty} L_{[x(hs)=a]} \tag{8.2}$$

The expectancy measure, average remaining years of disability-free (healthy) life for all survivors remaining at age x, is obtained by dividing total disability-free person-years at age $x (= T_{x(hs)})$ by the l_x value at that age in the basic life table (col. 8):

$$e_{x(hs)} = T_{x(hs)} \div l_x \tag{8.3a}$$

Table 8.1 Calculation of healthy life expectancy by the prevalence-ratio (Sullivan's) method, for white females in the United States: 1995

Age interval (x to x + n)	Number living at beginning of age interval of 100,000 born alive (l_x)	Stationary population in the age interval ($_nL_x$)	Proportion of persons in age interval in state considered unhealthy ($_n\pi_x$)	Proportion of persons in age interval in healthy state ($1 -_n \pi_x$)	Number of healthy years lived in age interval ($_nL'_x$)	Number of years lived in health state in this and all subsequent age intervals (T'_x)	Average number of years in healthy state remaining at beginning of age interval (e'_x)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0–5	100,000	497,211	0.0185	0.9815	488,012	6,981,686	69.8
5–10	99,321	496,412	0.0196	0.9804	486,682	6,493,474	65.4
10–15	99,247	496,020	0.0189	0.9811	486,645	6,006,992	60.5
15–20	99,156	495,294	0.0435	0.9565	473,740	5,520,347	55.7
20–25	98,938	494,163	0.0490	0.9510	469,949	5,046,598	51.0
25–30	98,720	492,962	0,0617	0.9383	462,546	4,576,649	46.4
30–35	98.455	491,441	0.0614	0.9386	461,268	4,114,103	41.8
35–40	98,094	489,247	0.0773	0.9227	451,426	3,652,837	37.2
40–45	97,580	486,191	0.0890	0.9110	442,920	3,201,409	32.8
45–50	96,861	481,715	0.1094	0.8906	429,015	2,758,489	28.5
50–55	95,754	474,612	0.1506	0.8494	403,136	2,329,473	24.3
55–60	93,969	463,278	0.1919	0.8061	374,375	1,926,338	20.5
60–65	91,152	445,546	0.2031	0.7969	355,055	1,551,963	17.0
65–70	86,772	419,113	0.2257	0.7743	324,520	1,196,908	13.8
70–75	80,441	381.366	0.2364	0.7636	291,211	872,388	10.8
75–80	71,408	328,775	0.2782	0.7218	237,310	581.177	8.1
80–85	59,051	257,187	0.3298	0.6702	172,367	343,867	5.8
85 and over	42,880	255,399	0.3285	0.6715	171,501	171,501	4.0

Source: U.S. NCHS/Molla et al. 2001. Primary source of prevalence ratios: National Health Interview Survey. U.S. NCHS life table for 1995.

Average years of disabled life for all survivors remaining at age x is obtained as the difference between total life expectancy and disability-free life expectancy.

$$e_{x(hs)} = e_x - e_{x(hs)} \tag{8.4}$$

In Table 8.1 health has been defined on the basis of respondent-assessed health, which is obtained from health interview surveys using a question with five response categories: Excellent, very good, good, fair, and poor. Healthy life years are years in a good or better state of health. Table 8.1 shows that white women in the United States in 1995 had a healthy life expectancy of 69.8 years; in that year the life expectancy of white women, without regard to their health status, was 79.6 years (not shown).

Two values for average remaining years of disability-free life at age x can be obtained, one for all survivors to age x, the so-called population-based value, and the other for disability-free survivors, the so-called status-based value. We obtained the first of these in Formula (8.3a). To derive the average years of disability-free life remaining at age x for the disability-free survivors to age x, we need estimates of the survivor function of disability-free persons at the initial ages of each age group. One way to derive them is to interpolate the $L_{x(hs)}$ function of the new table to obtain $l_{x(hs)}$ values.[1] Another is to take the differences of the $L_{x(hs)}$ values (equivalent to 5-year $d_{x(hs)}$ values uncentered), representing net shifts in the disability-free population, and interpolate these differences so as to obtain net shifts in the disability-free population between exact single ages. Now you can derive values of $l_{x(hs)}$ at every fifth year and divide $T_{x(hs)}$ by $l_{x(hs)}$, to derive $e_{x(hs)}$:

$$e_{x(hs)} = T_{x(hs)}/l_{x(hs)} \tag{8.3b}$$

The symbol $e_{x(hs)}$ represents then the average years of life remaining in a disability-free state at age x for persons surviving to age x in a disability-free state.

Limitations and Extensions

The prevalence-ratio method may be viewed as being consistent with the model of a double-decrement table in that two factors, death and net shift from disability-free life (net shift from good health), reduce the original cohort of disability-free (healthy) survivors. Probabilities of becoming (net) disabled at each age are not explicitly derived and applied in the method, but approximate estimates of these probabilities can be derived indirectly from the other functions in the table if they are desired.

[1] For interpolation to derive single-year-of-age data, the following approximation is satisfactory: $l_{x(hs)} = .50L_{x-1(hs)} + .50L_{x(hs)}$. For grouped data, osculatory interpolation, such as was described in Chap. 4, can be used. Once single ages are derived by such interpolation, the formula shown here can be applied to convert L_x to l_x.

The prevalence-ratio method of life table construction has some major limitations. First, the method does not allow explicitly for a transition from disabled status to nondisabled status, i.e., recovery from disability, although the prevalence ratios do represent currently disabled persons and hence allow implicitly for returns to health as well as for shifts to ill health and death. Second, the method cannot efficiently allow for more than one other decrement than death. Crimmins et al. (1993) found in their evaluation of the prevalence-ratio method that the estimates obtained by this method are only a rough reflection of the health levels of a population at a particular time when the estimates are evaluated against multistate tables, the present gold-standard for such measurement. Palloni et al. (2005), however, found in a simulation study carefully designed to evaluate the method that it gave results quite close to those of multistate tables except at the highest ages (90 years and over).

Many of the tables constructed so far have tended to conceptualize morbidity, or define inactive or unhealthy life, in terms of dependency, i.e., having ADL limitations. As is evident, health status and healthy life expectancy can be defined and measured in other ways than in terms of ADL-dependency. There are different dimensions of health that need to be measured and no single measure can adequately incorporate all aspects of health. In the original formulation of the Sullivan method, the total number of years of expected life is partitioned into only two health categories. This formulation can be modified with respect to the definition of health that is employed and the number of health categories. Multiple states of health, for example three or more rather than two, may be delineated. In this case, the prevalence ratios for each of these states must be applied separately to the life table stationary population in the construction of the table (if they do not overlap).

Following a design with several categories rather than only two categories makes it possible to track more than one dimension of the health life cycle. For example, in the calculation of disability-free life expectancy, we can further subdivide life expectancy with disability into life with severe disability and life with moderate disability, and carry these categories along with life without disability, as do Robine et al. (2001). Alternatively, tables of active life can be constructed for each of the ADLs or for combinations of them, so that the years lived free of each ADL and the average years of life remaining for each of them can be measured separately (Lynch and Brown 2005).

Life tables can be constructed by the prevalence-ratio method for a variety of health-defined groups, such as persons with/without any functioning problems (ADLs or IADLs), with/without a major chronic disease, with/without an impairment, with poor, fair, good, very good, or excellent health, with/without risky health behavior (e.g., smoking), having/lacking regular health care, or with/without a chronic disease that prevents participation in a major activity. Other tables of healthy life expectancy can be based on prevalence ratios characterizing health in terms of the presence or absence of a specific condition (e.g., cancer, heart disease, cognitive impairment).

A measure described by U.S. NCHS/Erickson et al. (1995) is based on an extension of the life table that produces the summary measure HLE, healthy life expectancy. Healthy life expectancy for a specified population indicates the

Table 8.2 Index values for health-related quality of life, with health states defined jointly in terms of self-perceived health states and activity limitation, for the United States: 1990 (Health status is indexed in relation to "Activity not limited" and "Excellent perceived health status" jointly)

	Perceived health status				
	Excellent	Very good	Good	Fair	Poor
Activity limitation					
Not limited	1.00	0.92	0.84	0.63	0.47
Limited – other	0.87	0.79	0.72	0.52	0.38
Limited – major	0.81	0.74	0.67	0.48	0.34
Unable – major	0.68	0.62	0.55	0.38	0.25
Limited in IADL[a]	0.57	0.51	0.45	0.29	0.17
Limited in ADL[b]	0.47	0.41	0.36	0.21	0.10

Source: U.S. NCHS/Erickson et al. 1995. Primary source: U.S. National Health Interview Survey and U.S. Medicare Current Beneficiary Survey
[a]*IADL* instrumental activities of daily living
[b]*ADL* activities of daily living

average number of years expected to be lived in a healthy state. This particular measure incorporates mortality and two aspects of morbidity in a single statistic (Fryback 1998; Institute of Medicine 1998). The method measures healthy remaining years by combining mortality in the life table with age-specific health-related quality-of-life scores that measure both functioning and self-reported health.

More specifically, the morbidity component in the NCHS/Erickson et al. table combines two measures, (a) respondent-assessed health and (b) functional limitations due to chronic conditions. Individuals are classified into 30 groups according to measures (a) and (b), and these groups are then assigned a quality-of-life score indicating the quality of life for that group relative to a healthy state. The conventional life table is then extended to include the age-specific health-related quality-of-life score, to derive the average years of healthy life a person would live. The scores for the morbidity component are obtained by weighting the score for six categories on functioning and the score on five self-perceived health statuses – 30 categories in total – as shown in Table 8.2. Since the survey data covered only the noninstitutional population, adjustments had to be made in the scores to allow for the institutional population – prisoners, nursing homes, military stations, long-term hospitals, and other residential care facilities. The final average scores, ${}_nQ_x$, (Table 8.3, col. 4) are multiplied by the number of person-years in each age interval, ${}_nL_x$(col. 3), to obtain the quality-adjusted person-years in each age interval (col. 5). The rest of the table of healthy life is completed with the usual steps.

The main result, average years of healthy life, or healthy life expectancy, resembles the measure derived from the simpler table of active life expectancy, i.e., Table 8.1. This more complex life table, like the simpler one, serves to divide the conventional figure for life expectancy into unhealthy and healthy years. The morbidity data needed to construct such tables of healthy life can usually be obtained from one or more of the national health surveys (either published data or the public-use microdata files).

Table 8.3 Calculation of healthy-and limitation-free life expectancy by the prevalence-ratio method, for the United States: 1990

Age interval Period of life between two exact ages in years (x to x + n)	Number living at beginning of age interval of 100,000 born alive (l_x)	Stationary population in the age interval ($_nL_x$)	Average health-related quality of life of persons in the age interval (Q_x)	Quality-adjusted stationary population		Years of healthy life remaining (e'_x)	Life years remaining (e_x)
				In the age interval ($Q^*_xL_x$)	In this and all subsequent age intervals (T'_x)		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0–5	100,000	495,073	0.94	465,369	6,403,748	64.0	75.4
5–10	98,890	494,150	0.93	459,560	5,938,379	60.1	75.1
10–15	98,780	493,654	0.93	459,098	5,478,819	55.5	71.2
15–20	98,653	492,290	0.92	452,907	5,019,721	50.9	66.3
20–25	98,223	489,794	0.91	445,713	4,566.814	46.5	61.3
25–30	97,684	486,901	0.91	443,080	4,121,101	42.2	56.6
30–35	97,077	483,571	0.90	435,214	3,678,021	37.9	51.9
35–40	96,334	479,425	0.89	426,688	3,242,807	33.7	47.2
40–45	95,382	474,117	0.88	417,223	2,816,119	29.5	42.6
45–50	94,179	466,820	0.86	401,465	2,398,896	25.5	38.0
50–55	92,420	455,809	0.83	378,321	1,997.431	21.6	33.4
55–60	89,735	439,012	0.81	355,600	1,619,101	18.0	29.0
60–65	85,634	413,879	0.77	318,687	1,263,510	14.8	24.8
65–70	79,590	378,369	0.76	287,560	944,823	11.9	20.8
70–75	71,404	330,846	0.74	244,826	657,263	9.2	17.2
75–80	60,557	270,129	0.70	189,090	412,437	6.8	13.9
80–85	47,168	197,857	0.63	124,650	223,347	4.7	10.9
85 and over	31,892	193,523	0.51	98,697	98.697	3.1	8.3

Source: U.S. NCHS/Erickson et al. 1995. Primary source of prevalence ratios: Table 8.2 and similar calculations for the institutional population given in U.S. NCHS/Erickson et al. 1995. U.S. official (NCHS) life table for 1990

Availability of HE Tables and Some Findings

The general use of the measure, healthy life expectancy (HLE) or health-adjusted life expectancy (HALE), goes back to the 1980s. The World Health Organization (1985) has used this measure since 1985 as one of the indicators for assessing national progress toward its Health for All goals. The organization of *Réseau Espérance de Vie en Santé* (REVES), The International Network on Health Expectancy, in the late 1980s gave considerable impetus to the development of methods and data for national production of tables of active life. REVES is an international organization that promotes research on the use of health expectancy as a population health indicator and on the definition, measurement, and comparison of disability and other health concepts used in the measurement of health expectancy.

Robine et al. (1999) report that by 1999 estimates of healthy life expectancy, mainly disability-free life expectancy (DFLE), were available for 49 countries, each based on its own definition of disability and developed from data in separate surveys. To achieve a greater degree of comparability in the results, in 2001 Robine et al. (2001) calculated DFLEs for 12 member states of the European Union (EU) for 1994 on the basis of disability data derived from a single survey using a single set of questions on disability. The disability questions concerned freedom from activity limitation and had two levels of severity. The data were collected in the European Community Household Panel (ECHP). Such results could serve as a firmer basis for international comparisons of health expectancy and generalizations about it than the previous individual studies.

To assess whether longer life implies better or worse health, Robine et al. (2001) compared the variation in life expectancy with the proportion of remaining life spent free of disability for the 12 EU states in 1994. For males there was little relation between these two series. For women, however, there was a strong negative relation between life expectancy at birth and the proportion of remaining life spent free of severe disability, indicating proportionately less time spent free of severe disability as life expectancy rises. At age 65 also, there was a tendency toward a smaller proportion of severe-disability-free years as life expectancy rose. The data also showed that, both at birth and at age 65, females spend a smaller proportion of their total life expectancy free of disability or severe disability than males. Previous studies for eight EU member states plus Norway, tracking the joint trends of life expectancy and health expectancy, suggested that the increase in life expectancy in Europe was accompanied by an increase in the time spent with moderate disability, but not severe disability (Robine et al. 1998).

The WHO published estimates of health-adjusted life expectancy (HALE) for 191 member states for the year 2000 with the collaboration of a panel of epidemiologists. Later, WHO published estimates of HALEs for all WHO member states for 2002 in the *World Health Report* for 2004. The WHO figures for the United States, as for the WHO member countries in general, show that a larger share of total life expectancy was lost to disability by women than men, so that active life expectancy for the two sexes was closer than total life expectancy. In the

more developed countries about 7.5 years were lost by males to disability and about 8.5 years were lost by females. As a result, the share of total life expectancy at birth lost as unhealthy years was somewhat higher for females than males.

Model calculations indicate that it takes smaller percent reductions in age-specific ratios of activity limitation to increase years free of activity limitation by a specified amount than general percent reductions in age-specific mortality, and even smaller percent reductions in both mortality and activity limitation (U.S. NCHS/Wagener et al. 2001). On the other hand, the contribution of decreases in self-assessed poor health status to increases in years of good health is less than the contribution of decreases in age-specific mortality. In general, however, the greatest increase in healthy life expectancy can be achieved by reducing morbidity and mortality at the same time.

Tables of Working Life and Disability

We can illustrate a further application of the prevalence-ratio method by measuring the effect of disability on average future lifetime in the labor force for males. The general plan of our method is as follows: Construct two double-decrement tables for males by the prevalence-ratio method, the first being a conventional labor-force life table and the second a labor-force life table that is designed to include in the labor force those who retired on disability. Then take the differences between the corresponding summary functions of the two tables. First, a table of working life is constructed as a type of double-decrement table that, in addition to death, incorporates the effect of net rates of labor force entry and net rates of labor force exit (Siegel 2002). This table is similar in structure to the table of healthy life shown as Table 8.1. The second table, a variation of the first, is a special hypothetical construct that assumes that no one retires on disability.

To construct a labor-force life table by the prevalence-ratio method,[2] secure labor force participation ratios (i.e., percentages of the population in the labor force at each age) from the U.S. decennial census or the U.S. Current Population Survey. These are applied to the L_x function of a standard U.S. life table, to obtain the life-table stationary population in the labor force ($L_{x(lf)}$). The expectation of life in the labor force at age x can be derived by dividing total years in the labor force, $T_{x(lf)}$, i.e., the sum of the $L_{x(lf)}$ column from age x forward, by the number of survivors in the standard life table:

$$e_{x(lf)} = T_{x(lf)} \div l_x \tag{8.5}$$

[2]Like tables of healthy life, tables of working life can also be constructed by the event-exposure method and by the multistate methods, to be discussed later.

I do not consider here the calculation of the average remaining years in the labor force for those who are still in the labor force, only for the entire survivor population. This is another option for securing a measure of the effect of disability on average years in the labor force.

Next we want to modify the standard labor-force life table by restoring to the labor force the persons who dropped out because they became disabled. The simplest assumption we can make is that the net withdrawals between age 35 and age 61 are permanent withdrawals due to disability. Those under age 35 and those over age 61 who withdraw from the labor force are assumed to withdraw for other reasons; net changes in the labor force below age 35 are all positive and the negative changes above age 61 are all assumed to be for retirement. Age 35 is selected because it is observed to be the age of the peak labor-force participation ratio; the actual data may suggest a different age for the beginning age of net exits for disability. We now assume that the net withdrawals at these ages remain in the labor force. To keep the disabled workers in the stationary labor force, we will have the new labor-force life table hold the labor-force participation ratio at age 35 constant from age 35 to age 61. The difference between labor force life expectancy in the two tables, the one allowing net withdrawals of persons aged 35 to 61 and the other not allowing such net withdrawals, is next obtained. This calculation provides a rough, maximal estimate of the effect of disability on labor force life expectancy.[3]

These calculations tend to overstate the effect of disability on the labor force because persons leave the labor force at the ages cited for other reasons than work disability. Retirement is an increasingly common basis of withdrawing from the labor force, even at ages below 61. Some leave the labor force because of age discrimination, plant closure, and family reasons. At the same time, a substantial number of persons with a work disability remain in the labor force for many years. At the older ages, work disability has a considerable impact on the labor-force participation ratio. For example, about three-quarters of the men aged 55–59 with a work disability are not in the labor force. Such data can be another basis for adjusting labor force participation ratios to determine the effect of work disability on labor force life expectancy.[4]

Event-Exposure Method

Alternatively, analysts can employ the event (incidence)-exposure method, following the double-decrement model or the increment-decrement model, to measure

[3] Another rough procedure for adjusting the stationary labor force in the life table involves augmenting the change in the stationary labor force from age to age by the estimated loss from withdrawals due to disability: (1) Obtain the differences between the stationary labor force at successive ages; (2) next, estimate the deaths at each age in the labor-force population by applying the observed death rates to the stationary labor force, (3) take the difference between the net changes from age to age in the stationary labor force in (1) and the deaths in (2); and (4) combine the results in (3) with the original labor-force stationary population.

[4] One way of making an adjustment using this type of data is to secure from the most recent census or Current Population Survey, for each age-sex group, (1) the sum of persons not in the labor force with a work disability and (b) the labor force; (2) divide the result in (1) by the labor force; and (3) multiply the factor in (2) by the labor force participation ratio in the first of the two labor force life tables being prepared; and (4) use the adjusted ratios in (3) to construct the second labor force life table.

healthy or disability-free life expectancy. Such a table is structured more like the standard (single-decrement) life table. Allowing for additional decrements, such as the probability of becoming disabled or of becoming disability-free, is a natural extension of the standard method of life table construction. The method appears to have been first applied to measure health expectancy by Katz et al. (1983) (See Table 8.4). They used it to measure the shift from an ADL-independent state to an ADL-dependent state (at least one ADL) and from the community to an institution. They could then measure life expectancy with and without an ADL-dependency.

In the simplest case of multiple decrements, the case of only two decrements, probabilities of dying and incurring a health condition, may be employed. Construction of a life table by the event-exposure method with two decrements is more complex than construction of a life table by the prevalence-ratio method, but it provides more products directly that are useful in analyzing the health status of a population. The Katz et al. table combines the probability of dying and the probability of losing independence (i.e., acquiring an ADL disability) in one rate (col. 2) and hence the decrements also are presented jointly (col. 4).

The event-exposure method differs from the prevalence-ratio method in its explicit use of probabilities of incurring an unhealthy condition, and in its explicit allowance for losses to the healthy survivor population of those who incur such conditions. With the event-exposure method, the basic data are generally obtained from longitudinal (panel) surveys, but a single cross-sectional survey with some retrospective questions will suffice. On the other hand, the prevalence-ratio method simply extracts the required health data from a single survey or census.

Neither method allows for movement from a healthy condition to an unhealthy condition and return to a healthy condition of these same individuals in an age interval. Both methods can deal with the one-way *net* movement from a healthy state to an unhealthy state in the age interval, but only the event-exposure method can handle the component increments and decrements separately, i.e., the shift of disability-free survivors to the disabled state and the return of disabled survivors to a disability-free state. Both methods can handle multiple decrements, but the calculations become messy and complex if there is more than one decrement in addition to death. Furthermore, usually the same mortality rates are applied to all health groups in both methods because mortality data are not available from within the system of calculations for these methods.

Some of these limitations can be overcome in constructing a multiple increment-decrement table by the event-exposure method when the necessary data can be secured from a panel survey. Mortality rates, age-specific incidence rates of disablement, and rates of recovery from disability can be derived from longitudinal (panel) health surveys. Such a table is still not the most efficient way to produce information on multiple-status shifts with regard to health and the range of measures that can be derived from it easily and accurately is limited. These problems can be overcome by employing the multistate life-table method, to which I will turn in a later section.

Table 8.4 Calculation of a truncated life table by the event-exposure method for the noninstitutional population 65 years and over in Massachusetts: 1974

Age (x to x + n)	Central rate of losing independence or dying in age interval ($_nm_x$)	Probability of losing independence or dying in age interval ($_nq_x$)	Survivors free of ADL (l_x)	Number losing independence or dying in age interval ($_nd_x$)	Person-years of survivors free of ADL from age x on (T_x)	Active life expectancy (e_x)
	(1)	(2)	(3)	(4)	(5)	(6)
65–69	0.07	0.29	100,000	29,152	999.463	10.0
70–74	0.10	0.40	70,848	28,304	570,767	8.1
75–79	0.10	0.41	42,544	17,418	287,724	6.8
80–84	0.16	0.57	25,126	14,353	118,940	4.7
85+	0.34	1.00	10,773	10,773	31,420	2.9

Source: Katz et al. (1983). Reprinted with permission of the Massachusetts Medical Society. Primary source of basic data: First and second waves of Massachusetts Health Care Panel Study

Note: Independence denotes free of ADL, an inability to perform one or more of the Activities of Daily Living, and not living in an institution. Active life expectancy is the number of years remaining before loss of independence or death

Chances of Ever Incurring Breast Cancer and Related Functions

Suppose we want to know the chances of a health event ever occurring from some age forward during the lifetime of an individual, whether at birth or some older age. Examples include experiencing a first heart attack, being diagnosed for the first time as a diabetic or as having breast cancer, and entering a nursing home for the first time. The same general demographic device can be used to deal with all of these events, a double-decrement life table constructed by the event-exposure method.

The lifetime chance of a woman's ever incurring breast cancer can be derived by life table calculations similar to the calculations of the lifetime chance of ever dying from a disease, as described in Chap. 4. A double-decrement life table for women constructed by the event-exposure method might display most or all of the following types of functions:

Age-specific probabilities of dying
Age-specific probabilities of incurring breast cancer for the first time
Survivors at each age who have never had breast cancer
Deaths in each age interval
Persons incurring breast cancer in the age interval
Number ever incurring breast cancer from age x on
Probability of ever incurring breast cancer from age x on
Years lived in age interval free of breast cancer
Years lived in age interval and later ages free of breast cancer
Average remaining years free of breast cancer
  Among all survivors
  Among the survivors free of breast cancer
Average remaining years with breast cancer
  Among all survivors
  Among survivors who have had breast cancer

Age-specific incidence rates for breast cancer in the United States can be obtained from the SEER survey. I leave to the reader the interpretation of these functions and turn next to a special issue regarding the double-decrement table described.

A Special Issue in Constructing a Double-Decrement Table for Breast Cancer Risk

When a decrement in addition to death is applied, the issue of double jeopardy in the calculation of survivors arises. The functions for deaths and incurrence of breast cancer represent events but the survivor function represents persons. It is important to reconcile the events column with the persons column in the table. If the probability columns allow for all deaths of women and all new cases of breast cancer in the age interval, some persons will be subjected to both events. An adjustment must be made to exclude the cases of both occurring to the same person in the same age interval and to avoid an excessive diminution of the cancer-free survivors column.

If we let q'_x represent the mortality rate of all women in a year and q''_x represent the risk of incurring breast cancer in the year, then the formulas for adjusting the rate for the incidence of breast cancer, computed independently, for the double counting of persons affected by both events are,[5]

$$\text{Adjusted } q''_x = q''_x(1 - 1/2q'_x) \tag{8.6a}$$

$$q''_x = \frac{\text{Adjusted } q''_x}{1 - 1/2q'_x} \tag{8.6b}$$

Chances of Ever Entering a Nursing Home

Measures of the utilization of nursing homes are needed to plan for the personnel and facilities required in future years. Some measures can be developed from an extension of standard life tables. The life table to be constructed goes beyond the conventional life table in including such functions as the probability of entering a nursing home for the first time during a year, the probability of ever entering a nursing home, and the average years of residence in a nursing home. One measure of nursing-home utilization, namely, the chance of ever entering a nursing home, is especially helpful in public planning, policy formation, and even individual planning for long-term care.

We want to construct a double-decrement table with deaths and first admissions to nursing homes as the two decrements. Cross-sectional data, such as those given in censuses or surveys (e.g., U.S. National Nursing Home Survey), can be employed to develop annual age-specific rates of net admissions or net discharges, from which the other functions in the table can be derived. A procedure for deriving approximate probabilities of net admissions to nursinghomes between age a and age $a + 1$ from census or other cross-sectional data on the percent of the population resident in nursing homes in single ages is to,

1. Subtract the proportion resident in nursing homes at age a from the proportion resident at age $a+1$:

$$\begin{gathered}\text{Diff. in proportions} = \left[\left(P^n_{a+1} \div P^t_{a+1}\right)\right] - \left[P^n_a \div P^t_a\right] \\ \text{For example, } (.118) - (.102) = .016\end{gathered} \tag{8.7a}$$

[5]The adjustment in the total number of deaths and in the total number of new cases of breast cancer is the same and can be made in either function.

If mortality rates can be determined separately for women who are free of breast cancer and those who are not free of it from two successive panel surveys, the adjustment for double jeopardy can be avoided. In such a case the functions obtained would be the deaths for women who have not incurred breast cancer, the number of new breast-cancer cases excluding any deaths, and the deaths of the new breast-cancer cases. The two main functions, adjusted deaths and new breast cancer cases, will then represent nonoverlapping numbers of persons to be used for reducing the function, female survivors free of breast cancer.

where $(P^n_a \div P^t_a)$ represents the proportion of the population resident in nursing homes at age a.

2. Divide the result in (1) by the proportion at age a:

$$\begin{gathered} Pr_a = (1) \div \left[\left(P^n_a \div P^t_a\right)\right] \\ (.016 \div .102) = .157 \end{gathered} \tag{8.7b}$$

The result in step (2), .16, is the estimated probabilityof net admissions to nursing homes between age a and a + 1. It is not the first-time admission rate, however. Such a measure can be derived only from survey data on first-time nursing-home residents at different dates, or from administrative-records data on first-time admissions and discharges from a sample of nursing homes.

We also need a set of probabilities of dying. One possibility is to use the mortality rates for the general population, but to adjust them for understatement of the true mortality levels. We could arbitrarily increase them by, say, 15%, unless actual information on the excess mortality of nursing-home residents is available.

Having the two sets of probabilities, one for net admissions and one for mortality, we can construct the remainder of the double-decrement table. The two decrements, deaths and net admissions (adjusted for net admissions that die within the year), are obtained by multiplying the probabilities by the survivors (which exclude nursing home admissions as well as deaths). Next, secure the net admissions ever occurring from age a on by summing the net admissions from age a forward to the end of the life table. Finally, divide these cumulative net admissions by the survivors at the starting age of any cumulation:

If A_a = surviving nonresidents of nursing homes at age a
and B_a = net admissions to nursing homes at age a,
then $C_a (= \Sigma_a B_a)$ = cumulative net admissions from age a forward
and $D_a (= C_a / A_a)$ = Proportion at age a who will ever enter a nursing home.

These are the synthetic estimates of the lifetime chances of ever entering a nursing home.[6]

[6]We may choose to take an indirect route to derive this measure by using the prevalence-ratio method of constructing the life table. We would apply the percentages of persons at each age who are not residents of nursing homes from a survey or census to the stationary population in a standard life table (L_x), to derive the person-years of non-residence in nursing homes at each age. The rest of the life table, including the function for the chances of ever entering a nursing home, can be derived by working backward to derive the prior functions for the number of deaths and the number of net admissions at each age. The net admissions to nursing homes and the deaths in the population excluding admitted persons are bound together in the age-to-age changes in the stationary population of nonresidents of nursing homes. The analyst has the task of unbinding these age-to age changes in the stationary population to extract the components. This is difficult because of the lack of information on the relative mortality of nursing-home residents and community-dwellers. Another problem is that this procedure yields data on all (net) entries into nursing homes, not only first entries.

Many studies providing estimates of the lifetime chances of entering a U.S. nursing home in older age have been made. They employ a variety of life table methods and the results differ substantially, as illustrated by the following figures: Kemper and Murtaugh (1991) estimated, on the basis of data on persons who became 65 in 1990, that a 65-year-old has a 43% chance of ever entering a nursing home. Of those who enter nursing homes, 55% will have a total lifetime residence of at least one year. McConnel (1984) concluded that the risk of first-time admission approaches, and may exceed, 50% while Liang and Tu (1986) estimated the risk at 36%. In general, the probability of nursing home use increases sharply with age.

A Disability Life Table Derived by the Event-Exposure Method

As a final illustration of the event-exposure method, I show a portion of a disability life table published by the Social Security Administration, relating to the cohort born in 1985, of insured male workers of the United States (Table 8.5). The table shows active (i.e., disability-free) survivors from age 20 to each successive age, the probability of survival and the probability of becoming disabled from age 20 to each successive age, and decrements (deaths, disablements) as well as increments (recoveries) at each age. To construct these tables, data on persons who have died or withdrawn from the labor force because of disability, available in the records of the Social Security Administration, are employed to determine the separate probabilities of death and disability on which the table is built. Table 8.5 presents selected functions from the original published table, in particular the active surviving insured workers and the decrements for these insured workers.

The Multistate Life Table

With advances in mathematical demography and the availability of more powerful computers, the multistate life-table model was introduced in the last quarter of the last century as a more flexible and elegant tool for analyzing life-course changes in multiple demographic statuses. Techniques of multistate demography can be used to analyze health changes over the lifetime of a birth cohort by explicitly allowing for the transitions among all the health states and from each health state to death, within each age group. (See, for example, Rogers et al. 1989; Crimmins et al. 1994; Land et al. 1994). Multistate tables can also be used to measure health-status changes for cohorts that incur specific severe chronic illnesses, e.g., heart disease, diabetes, and cancer, as well as for those that incur a specific type of disability or impairment. Moreover, the multistate model can deal with multiple starting states and follow each in its many transitions along the age trajectory. For this purpose one needs information on occurrence and recovery rates, and sets of death rates, for those afflicted with the specific health conditions under study. Data from at least two waves of a longitudinal (panel) survey are required, permitting the matching of reports for the same individual at two dates.

Table 8.5 Segment of disability life table derived by the event-exposure method for insured male workers 20–30 Years of age and other selected ages born in 1985: United States Social Security Area

Age	Probability of disability from age 20 to age x	Probability of death while active from age 20 to age x	Active survivors at beginning of year	Deaths of active persons in year	Newly disabled active persons	Newly recovered in year
	(percent) (1)	(percent) (2)	(3)–(4)–(5)= (3)	(4)	(5)	(6)
20	–	–	1,000,000	1,130	1,469	15
21	0.1	0.1	997,401	1,189	1,466	56
22	0.3	0.2	994,746	12,222	1,530	104
23	0.4	0.4	991,994	1,208	1,532	165
24	0.6	0.5	989,254	1,145	1,548	280
25	0.8	0.6	986,561	1,051	1,632	421
26	0.9	0.7	983,878	956	1,786	536
27	1.1	0.8	981,136	967	1,998	620
28	1.3	0.9	978,271	804	2,104	680
29	1.5	1.0	989,557	783	2,019	724
30	1.7	1.0	972,561	786	2,199	725
40	2.0	6.8	931,925	1,340	4,085	707
50	3.8	13.8	861,803	2,032	8,272	576
60	5.9	27.8	721,600	2,309	16,970	321
66	7.9	37.0	630,170	6,485	2,645	0
67	8.6	37.9	621,040	–	–	–

Source: U.S. Office of the Chief Actuary, Social Security Administration/Baldwin (2006)

Note: The probability of disability in col. (1) and the probability of death while active in col. (2) are mutually exclusive since deaths of disabled persons in col. (1) are excluded from the probabilities of death in col. (2). Active survivors in col. (3) at age $x + 1$ are derived by subtracting deaths of active persons in col. (4) at age x and newly disabled active persons in col. (5) at age x from active survivors in col. (3) at age x. Deaths in col. (4) relate only to deaths of active persons, and persons newly disabled in col. (5) relate only to active persons

Exhibit 8.1 Types of products of a multistate life table on ADL-disability

Number of (ADL) persons disabled for first time during the life course
Proportion of persons who ever become (ADL) disabled
Expectation of life at birth in an (ADL) disability-free state
Proportion of life expectancy at birth in (ADL) disability-free state
Proportion of life expectancy at age 25 in a (ADL) disability-free state
Expectation of life at age 65 in an (ADL) disability-free state
Proportion of life expectancy at age 65 in a (ADL) disability-free state
Probability that an (ADL) disability-free person reaching his/her xth birthday will be disabled by his/her $x + n$th birthday
Expectation of life at birth in an (ADL) disabled state
Proportion of life expectancy at birth in an (ADL) disabled state
Expectation of life at age 65 in an (ADL) disabled state
Proportion of life expectancy at age 65 in an (ADL) disabled state
Proportion of persons at age x who will die in an (ADL) disabled state
Proportion of persons at age x who will die in an (ADL) disability-free state
Probability at birth that a disablement will end in death
Probability at age 65 that a disablement will end in death
Difference between average number of years in an (ADL) disability state at birth and at age 65
Number of (ADL) disablements that ever become disability-free
Proportion of (ADL) disablements that ever become disability-free
Proportion of (ADL) disablements that become disability-free and then become (ADL) disabled again
Number of (ADL) disabled persons at age x who become disability-free in the next 5 years
Proportion of (ADL) disabled persons at age x who become disability-free in the next 5 years
Change in expectation of (ADL) disability-free life at age x depending on disability status at age x

Note: The products would be greatly increased if we were dealing with additional disability and morbidity states. The multistate life table is distinctive in allowing for returns to prior states, applying separate mortality levels to the different states, and being able to "survive" cohorts in the different states from birth. On the other hand, it normally does not allow the same person to have two changes of health status in 1 year

The multistate model is distinctive in allowing efficiently for exits from given states and returns to prior states in each age group, applying separate mortality rates for the different states, and being able to carry forward cohorts in the different states from birth. The latter ability makes the computation of population-based and status-based measures readily possible. The focus is on gross, not net, changes, and on aggregate data compiled from microdata based on panel sample surveys or matched census samples. Unless the survey questions are specially structured, however, only the initial and final statuses of individuals are recorded for the year's comparison of statuses. Hence, if individuals make repeated changes during a year, the intermediate changes are lost.

This type of table can answer questions such as those listed in Exhibit 8.1. Among these are: What is the average number of remaining years of life that will be spent free of disability or major chronic disease? What proportion of remaining years will be spent free of disability or major chronic disease? What percent of persons will

ever incur a disability or major chronic disease? What is the average age of onset of disability or major chronic disease? What is the expected duration of stay in each health state? What proportion of persons die while in any state? What is number of transitions to each state per person (allowing only one per age group)? A series of such tables for different dates can answer questions such as the following: How did the average years of healthy life/disability-free life change between the dates? How did the proportion of unhealthy/disabled persons at age 65 and over change over time? How did the risk of ever becoming disabled change between dates? How did the age of onset of disability or major chronic disease change over time?

General Theory of Multistate Tables

The multistate life table is based on the assumption that the probabilities follow a Markov process and so the equations for constructing such tables are consistent with a Markov-process design. A Markov process is essentially specified when the probability that an individual will leave his/her state depends only on the state and a person's age. More fully defined, the Markov-chain model has the following characteristics:

1. The probability that a person will move between states depends only on these states and the person's age.
2. Hence, the probability is not affected by the duration in the particular state, the person's states prior to the present transition, and any covariates that might otherwise affect the transition.
3. The states are mutually exclusive and discrete states.
4. Persons in a state have an equal probability of moving to other states.
5. At least two of the states are available to exchange occupants so that there are entrants and exits in at least one case.
6. Some states may not be experienced by persons in another state.
7. Changes of states do not occur in a fixed order.

Multistate models accommodate more than one active state in addition to death, and the occupants of any active state can move from one active state to any of the others or to death during any transition period. In other words, the multistate model is assumed to involve $k + 1$ states, the $k + 1$th state being an absorbing state or death. Occupants of at least two of the k states must be able to move between one another, so that there are entrants into at least one active state during the transition period. Note that the standard life table, with only two states, living and dead, is not normally considered a multistate table inasmuch as it does not involve two or more active states and employ the multistate design.

Because of the number and complexity of the computations required in multistate modeling, the formulas are expressed in matrix notation, a shorthand algebraic notation for representing and mathematically processing data in blocks. As seen below, matrices are represented by capital letters in bold type. Multistate methods

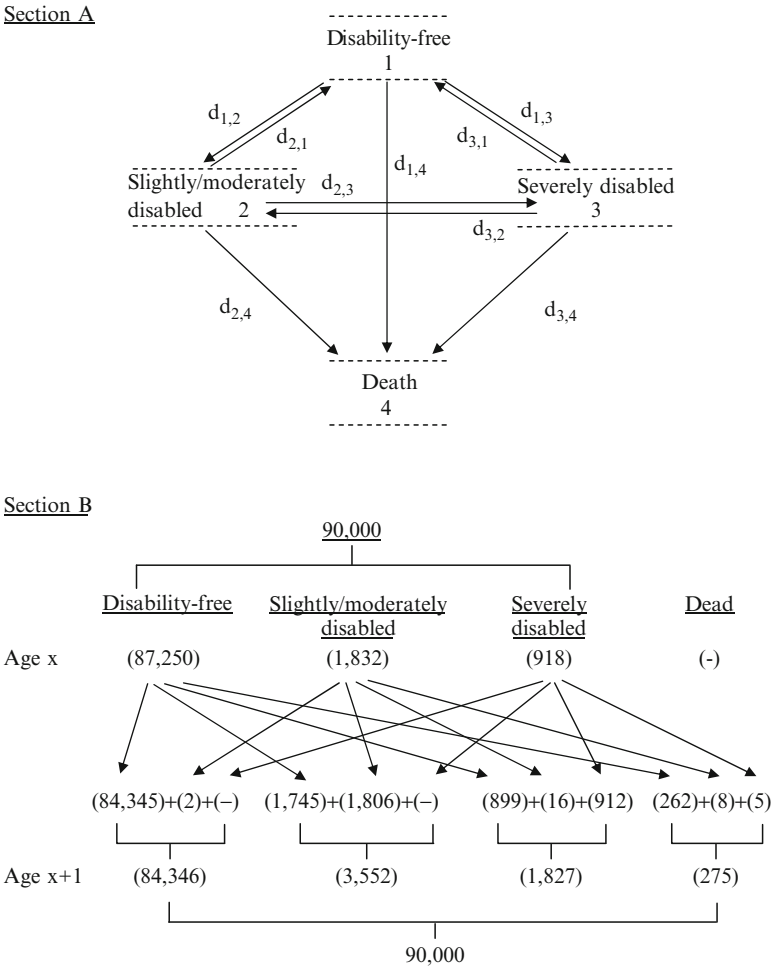


Fig. 8.2 Transitions between various states of ADL-disability from age x to age $x + 1$ (Note: There are rounding errors in section B. (-) denotes zero; Source: Table 8.6)

were not possible or manageable until the advent of powerful computers and calculation with matrix algebra made the task easier.

A multistate life table relating to disability is based on transition probabilities between pairs of states identified as disability-free, disabled, and death. Transition probabilities represent the likelihood that an individual of a given exact age and health status (disability-free, not disability-free) will be in the three possible health states (e.g., disability-free, not disability-free, and dead) one year later. For each age and health status the transition probabilities sum to one. Figure 8.2 depicts these varied transitions for three disability states and death. Such transitions may be, for example, a shift from a disability-free state to a severely disabled state or a

shift from a moderately disabled state to a disability-free state. As noted, transition probabilities at each age depend only on the immediately prior state – an assumption that ignores other factors that could influence movement between disability states, such as the duration in a disability state, the disability state of the person prior to the current transition period, and the socioeconomic variables influencing disability status, such as education, income, and marital status.

Sources of Data

The basic data for the construction of a table of healthy, active, or disability-free life come from longitudinal (*i.e.*, panel) health surveys that provide observations of the health status of a person at two or more dates (*e.g.*, a year apart). A single health survey inquiring about current and prior health status could provide the basic data (except for death), but the retrospective reports on health status would clearly be less accurate and less complete than current reports for two dates, and could not provide estimates of mortality rates. The list of recent U.S. national health surveys that provide data on health conditions over a series of years was enumerated in an earlier section of the chapter. In addition to providing data on health status at each date, and the basis for measuring the change in health status between the dates, the surveys provide the basis for measuring differential mortality according to health status. Inasmuch as the sample sizes may be quite small for some groups, the observed mortality rates may be quite irregular, and considerable smoothing of these rates may be necessary before constructing the multistate tables. In addition, it is wise to evaluate the losses through death to the sample between waves of the surveys by comparing the number of these deaths with estimates of deaths based on available vital statistics and life tables.

Construction of Multistate Tables

A detailed illustration of the steps in constructing a multistate table on healthy expectancy is not given here since the process is quite complex, but the basic concepts and formulas will be set forth, and references to appropriate computer software for carrying out the computations will be given. An illustration of the calculation of a multistate table for the labor force is given in Siegel and Swanson (Eds.)/Kintner (2004), Chap. 13.

A basic preparatory step in the construction of such a table is the determination of the health transitions according to age between survey dates. The transition rates from state i to state j in the age interval x to $x + 1$ (or x to $x + n$) are initially derived in the form of observed central transfer rates (M_x), that is, the number of transfers (e.g., number of disablements between exact ages x and $x + n$) per 1,000 persons aged x to $x + n$ in the observed population (i.e., the person-years lived in state i between age x and age $x + n$). Because of irregularities in the age pattern of these rates, resulting from small sample sizes for some categories, the transfer rates

may require smoothing to remove sampling fluctuations. A log-linear regression equation such as the following or other smoothing method may be used for this purpose (Crimmins et al. 1994; Rogers et al. 1990):

$$M_x = e^{(\alpha+\beta x)} \tag{8.8a}$$

or

$$\ln M_x = \alpha + \beta x \tag{8.8b}$$

The smoothed transfer rates, represented in matrix form by **M**(x, n), next have to be converted to a transition probability matrix, **P**(x, n). The elements in this matrix, $p^{ij}(x)$, represent the probability that a person at exact age x and state i at the beginning of an interval will be in state j at the end of the interval at exact age $x + 1$ (or $x + n$). The following numerical approximation has been proposed for the conversion from **M**(x,n) to **P**(x,n) by Rogers and Ledent (1976):

$$\mathbf{P}(x, n) = [\mathbf{I} + n/2\,\mathbf{M}(x, n)]^{-1}[\mathbf{I} - n/2\mathbf{M}(x, n)] \tag{8.9}$$

where **P**(x,n) is the transition-probability matrix, **M**(x,n) is the matrix of transfer rates, and **I** is the identity matrix. (See also Crimmins et al. 1994; Schoen 1988: 70; Willekens et al. 1982). Where the transfer rate is low (*e.g.*, 0.02), the transition probability roughly approximates its complement (e.g., 0.96); where it is high (e.g., 0.30), the transition probability may differ substantially from its complement (e.g., 0.54). The health transition probabilities may be refined by applying a multinomial logistic regression of the form,

$$\ln[p^{ij}/p^{ii}] = \beta_{ij0} + \beta_{ij1}x \tag{8.10}$$

to the preliminary estimates, where the initial health status is i and the terminal health status is j.

In the simplest situation for applying the multistate method to the analysis of health/disability states, there are two origin states and three destination states, including deaths, representing three different states, four possible transitions between the different states, and six transitions in total, including transitions from a state to the same state, as indicated by the following paradigm:

Transition probabilities				
State of origin, exact age x	State of destination, exact age x + 1			
	Independent life	Dependent life	Death (3)	Total
Independent life	X	X	X	1,000
Dependent life	X	X	X	1,000

These relations can be structured in the following 3 by 3 matrix of health transition probabilities:

$$\mathbf{P}(x,n) = \begin{vmatrix} p(x,n)^{11} & p(x,n)^{12} & p(x,n)^{13} \\ p(x,n)^{21} & p(x,n)^{22} & p(x,n)^{23} \\ 0 & 0 & 1 \end{vmatrix} \tag{8.11}$$

where **P**(x,n) is the matrix of health transition probabilities, $p(x,n)^{ij}$ is a particular transition probability, x is age, n is one or more years, and the superscripts represent health statuses. The superscripts are given in the order of the transition. Superscript three represents death, and so row three shows two zeroes for the two impossible transitions and the number one for the absorbing state of death, $(p(x,n)^{33} = 1)$.

The remaining life-table functions may be derived from the transition-probability matrix or matrices. The survivors at exact age x (or $x+n$) and the person-years lived between exact age x and exact age $x+1$ (or $x+n$), in each health state, may be calculated by the following matrix formulas:

$$\mathbf{l}(x+n) = \mathbf{l}(x)^{*}\mathbf{P}(x,n) \tag{8.12}$$

$$\mathbf{L}(x,n) = n/2\,[\mathbf{l}(x) + \mathbf{l}(x+n)] \tag{8.13}$$

The second equation assumes that l(x) changes linearly within the age interval and that transfers are distributed rectangularly within the interval. An element in the matrix **l**(x+n) is l^{ij}(x+n) and an element in the matrix **L**(x+n) is L^{j}(x+n). Table 8.6 illustrates the matrix calculations for the transitions of the cohort from one disability state to another between ages x and $x+1$, using four states – disability-free, slightly or moderately disabled, severely disabled, and dead. The 90,000 surviving members of the cohort x years of age are redistributed among the four states from the numbers shown for year y for age x, on the basis of the transition probabilities for that age, to the numbers shown for year $y+1$ for age $x+1$. Figure 8.2 diagrams the many transitions represented by Table 8.6. We can obtain the absolute number of life table transfers of each type between health states for the stationary population at each age, **D**(x, n), by the formula,

$$\mathbf{D}(x,n) = \mathbf{L}(x,n)^{*}\mathbf{M}(x,n) \tag{8.14}$$

This assumes that, within each age group, the life-table rates are equal to the observed central event-exposure rates. More specifically, it implies that within each age interval and each health status of the life table, the population is distributed as in the observed population.

Finally, the expectation of life for each health state at each exact age may be derived by the following formulas:

$$\mathbf{T}^{a}(x) = \Sigma_{x}^{\omega}\mathbf{L}^{a}(x,n) \tag{8.15}$$

and

$$\mathbf{e}^{a}(x) = \mathbf{T}^{a}(x) \div \mathbf{l}(x) \tag{8.16a}$$

Table 8.6 Calculation of survivors from age x to age x + 1 in various states of disability using transition probabilities

	At age x				Survivors	
At age x + 1	Disability-free	Slightly/ moderately disabled	Severely disabled	Dead	At age x	At age x + 1
Disability-free	.9667	.0010	.0005	–	87,250	84,346
Slightly/moderately disabled	.0200	.9860	.0005	–	1,832	3,552
Severely disabled	.0103	.0085	.9940	–	918	1,827
Dead	.0030	.0045	.0050	1.0000	–	275
Sum	1.0000	1.0000	1.0000	1.0000	90,000	90,000

Source: Hypothetical data

or

$$^{a}\mathbf{e}^{a}(x) = \mathbf{T}^{a}(x) \div \mathbf{l}^{a}(x) \tag{8.16b}$$

or

$$^{i}\mathbf{e}^{a}(x) = \mathbf{T}^{a}(x) \div \mathbf{l}^{i}(x) \tag{8.16c}$$

where $\boldsymbol{T}^{a}(x)$ is the number of person-years lived in state a (e.g., healthy /active state) at age x and beyond, derived by summing the $\boldsymbol{L}^{a}(x)$ values in a given state, and $\boldsymbol{l}(x)$ is the number of survivors to age x. The first formula for $e(x)$, (8.16a), the so-called population-based formula, gives the average years of healthy/active life of all survivors (regardless of health status) at exact age x, while the second and third formulas, the so-called health-status-based values, give the average years of healthy/active life at age x for healthy/active survivors at that age (8.16b) and of unhealthy/inactive survivors at that age (8.16c), respectively.

To estimate the health-status-based measures, separate survivor populations according to health status must be carried forward. The status-based measures can be of two types, depending on the age at which health status is considered: it may be the current age or a fixed age, such as the lowest age in which there are some occupants of every health status under study. If we focus on the current age, then we accumulate survivors of each health state with advancing age, through the application of the transition probabilities, allowing the healthy cohort(s) to contribute new members to the unhealthy cohort(s), the unhealthy cohort(s) to restore members to the healthy cohort(s), and the healthy and unhealthy cohort(s) to shed their dead members. Accordingly, at each age, the combined cohorts of survivors reflect the numbers of healthy/active persons and the numbers of unhealthy/inactive persons. Each of these numbers are available as divisors for computing alternativeestimates of life expectancy.[7]

Computer Programs Available

The computer program most widely used for producing multistate tables is the IMaCh Program designed by the Institut National d'Etudes Démographiques (INED) in Paris and jointly sponsored by INED and Euro-REVES. The input data

[7] An alternative route for carrying out the calculations of multistate tables may be employed. Transition probabilities can be derived directly from the survey data. To do this, first, an n by $n + 1$ contingency table giving health status at the earlier date and health status and deaths at the later date is set up. Age-specific transition probabilities (P_x) can be calculated directly from this table. For example, the probability of transferring from the healthy survivor population at year y to an unhealthy survivor population at year $y + 1$ between age x and age $x + 1$ is calculated as the proportion of such transfers out the total number of healthy survivors in year y. These transition probabilities can be converted into central age-specific transfer rates (M_x) by appropriate formulas. (See *The Methods and Materials of Demography*, 2004, p. 333). The age-specific transfer rates can be multiplied against the stationary populations to derive the transfers of each kind, including death, i.e., $D_x = M_x^{*}L_x$.

of the INED/Euro-REVES program are the dates of birth and death, health status at each interview with their dates, and any covariates of interest. With these data, the program calculates transition probabilities from one health state to another for the period of observation. The program can calculate transition probabilities for intervals of one year, intervals longer than one year, and intervals shorter than one year. The output data are a group of multistate tables and illustrative charts, including total life expectancy and life expectancies in the different health states, i.e., both population-based and status-based life expectancies, and the standard errors of the life and health expectancies.

Others have also published computer programs for calculating multistate tables. Among them are: University of Michigan Population Studies Center (*www.psc.isr.umich.edu;* 2007); University of Wisconsin Social Science Computing Cooperative (*www.ssc.wisc.edu/∼mweden;* 2007); Lynch and Brown (2005); Hayward and Grady (1990), and Laditka and Wolf (1998).

Availability of Multistate Tables and Some Findings

Tables of healthy (disability-free, active) life have been calculated by the multistate method for only a few countries, including the United States, Canada, Italy, Japan, the Netherlands, and France. The U.S. tables have been used to track the trend in disabled or unhealthy life expectancy in the United States in the last several decades (Crimmins et al. 1989; Crimmins et al. 1996; Crimmins et al. 1997; Manton et al. 2006; U.S. NCHS/Molla and Madans 2008). The multistate method has also been used to analyze the relationship between disability and life expectancy for the sexes, socioeconomic groups, especially educational groups, and racial and Hispanic-origin groups, and between obese and non-obese persons, smokers and non-smokers, and other complementary groups.

With increasing age, the number of disability-free or healthy years falls while the number of disabled or unhealthy years may rise or fall depending on the age and measure of health employed (Tables 8.7 and 8.8). The share of disabled or unhealthy years of total life expectancy tends to rise. During the 1970s, years of disabled life increased in the United States, but during the 1980s and 1990s years of disabled life fell (Crimmins et al. 1997; Manton et al. 2006, Table 8.9). While the number of years of disabled life decreased after 1980, the relative share of disabled years out of all expected years has been increasing in this period, especially at ages 85 and over but also at ages 65 and over, and is expected to continue on this track during this century (Manton et al. 2006, Table 8.9).

Women at age 70 in the United States in 1984–1990 had a substantially greater total life expectancy than men, 3.6 years, but the years of independent life expected for the two sexes was substantially closer, 2.2 years, and the percent of total remaining years to be spent in a dependent state was much greater for women than for men, 20% vs. 14% (Crimmins et al. 1996, Table 8.7). This pattern appears in other tables of healthy/active life as well, even though the concept of health varies from table to table. In 1995 total life expectancy at birth in the United States was 73

Table 8.7 Years of expected life by functioning status for females and males in the United States at ages 70, 80, and 90: 1984–1990

		Expected independent life[a]			Expected dependent life[a]			
Sex and age	Total expected	Total	No functioning problems	Some functioning problems	Total	Unable to manage independent living	Unable to manage personal care	Percent dependent years
Female								
70 years	13.9	11.1	4.3	6.8	2.8	1.0	1.8	20
80 years	8.4	5.5	1.6	3.9	2.9	1.0	1.9	35
90 years	4.8	1.8	0.3	1.5	2.9	0.9	2.0	60
Male								
70 years	10.3	8.9	4.1	4.8	1.4	0.6	0.8	14
80 years	6.0	4.4	1.6	2.8	1.6	0.7	0.9	27
90 years	3.3	1.6	0.4	1.2	1.8	0.7	1.1	55

Source: Crimmins et al. (1996), Table 8.3

[a]The source denotes these as active and inactive years but the terms independent and dependent describe these conditions more exactly

Table 8.8 Total life expectancies and expected years in good and poor health, for males and females at ages 70, 80, and 90, derived by the multistate method: United States, 2001–2002

	Male expected years			Female expected years		
	Total	In good health	In poor health	Total	In good health	In poor health
Age						
70	13.0	10.1	2.9	15.7	11.8	3.9
80	7.7	5.6	2.1	9.0	6.3	2.7
90	4.2	2.9	1.3	4.6	2.9	1.7

Source: U.S. NCHS/Molla and Madans (2008). Private e-mail communication from M. Molla to the author, September 9, 2008. Primary source for health data: Center for Medicare and Medicaid Services, Medicare Current Beneficiary Survey
Note: Years in "good health" include years in "good," "very good," and "excellent" health, and years in "poor health" include years in "fair" and "poor" health

Table 8.9 Life expectancy (LE) and non-disabled/active life expectancy (ALE) at ages 65 and 85 for the United States: Estimates, 1935–1999, and projections, 2015–2080

	At age 65				At age 85			
Year	LE	ALE	Difference (disabled years)	ALE/ LE (%)	LE	ALE	Difference (disabled years)	ALE/ LE (%)
Estimates[a]								
1935	11.9	8.8	3.1	74	3.0	0.7	2.3	23
1965	15.0	10.9	4.1	73	5.4	1.5	3.9	28
1982	16.9	12.3	4.6	73	6.2	2.1	4.1	34
1999	17.7	13.9	3.8	79	6.4	3.0	3.4	47
Projections								
Scenario 1[b]								
2015	18.9	15.6	3.3	83	7.0	4.1	2.9	59
2022	19.4	16.4	3.0	85	7.3	4.6	2.7	63
2080	23.6	20.8	2.8	88	9.6	7.2	2.4	75
Scenario 2[c]								
2015	18.9	15.1	3.8	80	7.0	3.6	3.4	51
2022	19.4	15.5	3.7	81	7.3	3.9	3.4	53
2080	23.6	20.1	3.5	85	9.6	6.6	3.0	69

Source: Manton et al. (2006). Copyright © (2006) National Academy of Sciences, U.S.A. Reprinted with permission. Primary source: National Long-Term Care Survey
[a]Annual rates of decline in disability ratios: 1935–1965, 0.6%; 1982–1999, 1.7%
[b]Annual rates of decline in disability ratios: 1999–2022, 1.7%; 2022–2080, 0.8%
[c]Annual rates of decline in disability ratios: 1999–2080, 0.8%

years for males and 79 years for females, and the figures for expected years in good or better health were 65 and 69, respectively (U.S. DHHS 2003:Table 3.2). While females had a longevity advantage of 6 years over males, their advantage of years in good or better health was only 4 years because they spent more years in poor

or fair health than males (10 vs. 8). This type of outcome was true at every age. Similarly Molla and Madans (U.S. NCHS 2008) show that for 2000–2001 the difference between the total life expectancy of the sexes at age 70 was 2.7 years but the difference in life expectancy in good health was only 1.7 years (Table 8.8). In sum, disabled years make up a large share of the excess years of life expectancy that women have over men at the advanced ages. As a result, the numbers of disability-free years of men and women are more alike than the total years of life expectancy of the two sexes.

There are also race and class differences in total life expectancy and disability-free life expectancy. Guralnik et al. (1993) found that 65-year-old black men had a lower total life expectancy (11.4 years) and disability-free expectancy (10 years) than white men (12.6 years and 11.2 years), but that the figures for 65-year-old black women (18.7 and 15.9) were similar to those for white women. At age 75, black women had higher values than white women. The crossover of these figures reflects a phenomenon of advanced age identified earlier; now it is shown for disability-free expectancy. Guralnik et al. also found that education had a substantially stronger relationship to total life expectancy and active life expectancy than did race. At age 65 those with 12 or more years of education had an active life expectancy 2.4–3.9 years longer than those with less education, depending on sex-race group. Hayward and Heron (1999) have shown that some groups (e.g., American Indians) have longer life but worse health than the non-Hispanic white population, while others (e.g., Asian Americans) have longer life and better health than this population.

Limitations of Multistate Methods

The results on expected years of healthy life obtained from the prevalence-ratio method, the event-exposure method, and the multistate method are not comparable because of differences in the data, methods, and assumptions employed (Rogers et al. 1990). The results from the multistate tables are viewed as more realistic and valid. However, inasmuch as the data on mortality and health transitions used in constructing the multistate life tables may be unstable because of the great disaggregation of the data and the small number of cases in some cells, even the summary estimates of healthy life expectancy derived from these tables should be viewed with caution. The complexity and data requirements of the multistate method preclude its adoption for most countries of the world at this time, particularly the statistically less developed countries.

Multistate life tables share the fundamental limitations of life tables as descriptors of reality. They are hypothetical constructs: The cohorts are synthetic and the populations are stationary. They are not indicators of past or future changes, only summary indicators of the current state of population health, useful for comparison with comparable figures for other populations similarly derived.

Microsimulation and Health Status Transitions

The fourth method of measuring health expectancy is denominated microsimulation and, as proposed by Laditka and Wolf (1998), is an extension of, and an advance over, the multistate life table method. The method models health transitions through microsimulation. It has several advantages. It can readily model all the data on changes in health states that were obtained in the original sample survey. While a method can describe only as many health changes as the respondent reports in a survey, to the extent that the respondent reported every change, not just the situation a year earlier, microsimulation has the ability to incorporate all of these reports readily into the analysis. Microsimulation also permits the ready calculation of the sampling variability of the health expectancies as well as the sampling variability of the parameters used in estimating them. Finally, microsimulation permits the researcher to incorporate covariates into the estimation process more readily than a macrosimulation approach such as the multistate life table method. Multistate life table models have been used as a way of modeling covariates in complex life cycle processes such as health, but microsimulation extends multivariate multistate methods by making this process more flexible and more expansible (Laditka and Hayward 2003).

The time unit used in the transitions in the microsimulation method is a month. The information used in the Laditka/Wolf study was obtained from the Longitudinal Study of Aging (LSOA). Transition probabilities were calculated with race and education as covariates and further estimated by multinomial logistic regression. The microsimulation method produces a distribution of the population in the various health states and information on the length of time spent in each health state, the number of health transitions, and the distribution of the episodes in each health state by the duration of the episodes.

Measurement of Health-Related Quality of Life

In this section the range of measures describing the state of health of a population is expanded to encompass measures that commonly include in a single number the years of life lost or gained based on conventional life-table data and the general health profile of a population based on other sources. The latter is the key component, and in some cases the mortality component is omitted and a life table is not involved. Many such measures have been developed. The various mortality and morbidity components of the health measure are combined by weights so as to produce a multidimensional index of population health. Different health conditions may also be combined by weights that vary according to the severity of the conditions. These measures are called quality-adjusted life years (QALY). In this group are the measures called health-adjusted life years (HALY), healthy life years (HeaLY), and disability-adjusted life years (DALY). The measures are distinguished by how they define health, measure health conditions, and weight the health conditions with one another and with life years lost or gained.

To calculate these measures and to achieve comparability from population to population, a single measurement unit is required. Time can serve this purpose, whether we are measuring healthy or disabled life or gains or losses in life expectancy. The weighting schemes for health conditions may involve assigning a number from 0 to 1 to reflect the quality of life or the social utility of the health condition experienced (Fryback 1998). The selection of the weights requires considerable care because of their potential effect on the resulting measure of population health. The health status or disability score may then be combined with years of life lost or gained to produce quality-adjusted life years (QALY) or some specific type of QALY (i.e., HALY, HeaLY, or DALY).

Quality-Adjusted Life Years

Quality-adjusted life years (QALY) is then a generic term referring to a group of measures that normally combine in a single figure the impact of years gained by avoiding death and years of healthy life achieved by avoiding ill-health or disability. In thinking about this measure, bear in mind that there are many variants, just as there are many definitions of health, measures of health, and ways of combining statistical components into single measures. One special variant gives the life-years gained by implementation of some diagnostic or treatment protocol and does not involve use of life table data. Weinstein (2005) employs this version of QALY in measuring the cost-effectiveness of selected methods for diagnosing and treating major health conditions. In calculating his QALYs, he assigns relative weights to different levels of health or health conditions, ranging from perfect health (1.0) to death (0.0). Between these extremes are various levels of health or health conditions, ranging from a mild headache (.99) to serious disabilities such as a stroke (.15). He derived these weights from reports of the relative values that people place on relief from these health conditions. The weights are multiplied against the durations of the health conditions, and the products are then added to obtain the number of quality-adjusted life years (QALYs) gained by the use of the medical procedure being evaluated.

Here is a specific example suggested by Weinstein (2005). Let us assume that a man is a candidate for treatment of asymptomatic coronary artery disease (i.e., CAD without breathlessness or angina on exertion). For the first 3 years after treatment is started, he continues to be asymptomatic (weight 1.0); then he experiences angina for a period of 1 year (weight 0.5); surgical intervention (i.e., coronary artery angioplasty) gives him almost complete relief for the next 4 years (weight 0.9); then he has a bad setback and is bedridden (weight 0.1); finally he dies after 2 years (weight 0.0). His QALY is the cumulative sum of the products of the weights and the durations, or 7.3 years [$= (3^*1.0) + (1^*0.5) + (4^*0.9) + 2^*0.1)$]. Geometrically, it is the total area under the bars in Figure 8.3.

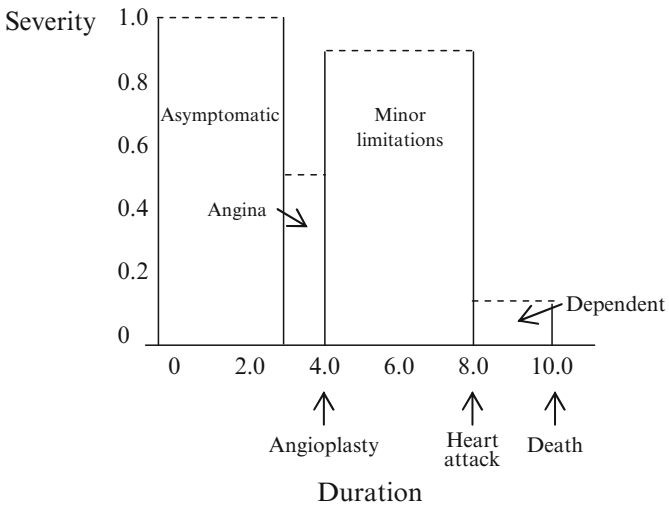


Fig. 8.3 Illustrative diagram for the measurement of quality-adjusted life years, given a major health condition, taking treatment, duration, and severity into account (Note: The total area under the bars gives the quality-adjusted life years. See text for further explanation; Source: Reprinted with permission of the Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University, from Weinstein (2005), Fig. 1)

Disability-Adjusted Life Years

Another version of QALY, DALY (disability-adjusted life years) is the combination of the years of life lost to premature death (YLL) and years lived with a disability (YLD):

$$\text{DALY} = \text{YLL} + \text{YLD} \tag{8.17}$$

One DALY is one lost year of healthy life. A premature death is one that occurs before the age to which the decedent could have been expected to survive if he or she had the life expectancy of some standard or model population. To measure years lost to premature death (YLL), the analyst has to select an achievable standard of life expectancy. The years lost is then the difference between the decedent's age at death and life expectancy at that age in the standard population. For the standard population,the Global Burden of Disease study[8] (Murray and Lopez 1996a, b)

[8] *The Global Burden of Disease and Injury Series* (Murray and Lopez, editors, 1996) is a landmark publication on the world's health condition, produced by researchers from the Harvard School of Public Health and the World Health Organization and several dozen collaborators from around the world. It is a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 for the regions and countries of the world, with projections to 2020. The study provides an internally consistent and comparable set of estimates of current patterns of mortality and disability, with detailed information on the epidemiology of 240 conditions. The result of the research is the 10-volume, *The Global Burden of Disease and Injury Series* published in 1996 with the support of the World Bank.

selected the population of Japan, the country with the highest life expectancy at birth in the world. Japan's life expectancy at birth was 82.5 years for women and 80.0 years for men at the time of the study. Assume a man in country C acquired a disability at age 63 and lived with it until his death at age 70. Since life expectancy for men at age 70 in the Japanese life table was 15 years, the first element in the DALY is 15. The second element in the DALY is 7 (= 70 − 63), the period of his disability. Hence, the DALY is 22 years (= 15 + 7). The Global Burden of Disease (GBD) study employed this type of measure, but with refinements, for international comparisons of the burden of disease.

Years of Life Lost

To calculate the DALY for a given population under the GBD protocol, years of life lost (YLL) and years lived with a disability (YLD) must be determined for each person and then cumulated to secure population totals. To use the authors' example, to calculate the DALY incurred from road traffic accidents in India in 1990, one must add the total years of life lost (YLL) in fatal road accidents by victims in 1990 and the total years of life lived with disabilities (YLD) by survivors of such accidents.

The measure developed in the GBD study for use in describing and analyzing the world health situation is more refined than the DALY measure so far described. Refinements were introduced to derive an internationally standardized measure that combines both health conditions and mortality in the assessment of health status and that measures the burden of disease and injury in a common unit – time. As I suggested, the results are also useful for measuring the cost-effectiveness of interventions, i.e., the cost per unit of disease-burden averted, as in the Weinstein study.

Selecting the standard. In selecting the target figures for life expectancy in the event of premature death, the analyst has to decide whether to take account of variations in life expectancy for the two sexes within a country, variations among the races, variations among socioeconomic groups, and variations among countries. The authors of the GBD study decided to disregard local and national-area variations in life expectancy in selecting the target, with one exception. The target expectations are the same for persons in Africa as for persons in Sweden, the same for white males in England and aboriginal males in Australia, and the same for an upper class Hindu in India and a poor Muslim in Bangladesh. The target for life expectation is assumed to differ for men and women, however. Accordingly, the GBD has assigned men a lower target for life expectancy than women. This is a reasonable choice since, in all MDC and nearly all LDC, women's life expectancy exceeds that of men and a substantial part of the male/female difference in life expectancy may have a biological basis. Hence, the GBD study assumes two target (Japan) life tables, one for males and one for females.

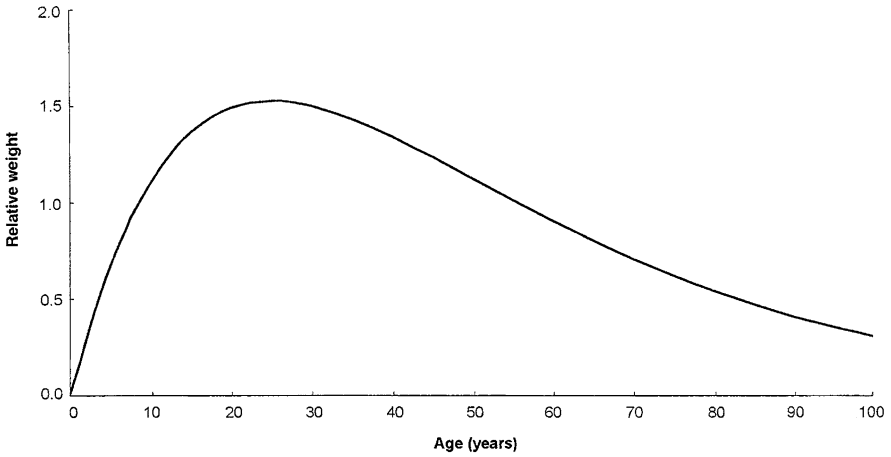


Fig. 8.4 Relative value of a year of life lived at different ages, as incorporated into DALYs (Source: Murray and Lopez (1996a), Figure 5, p. 9. Copyright © 1996 World Health Organization. Reprinted with permission

Relative value of a year of life at different ages. Different relative values of a year of life lived at different ages are incorporated into the DALY. A number of studies confirms broad social preferences for weighting the value of a year lived by a young adult more heavily than one lived by a very young child or an older adult. Therefore, the GBD researchers incorporated age-weighting into the DALY. They assumed that the relative value of a year of life rises rapidly from zero at birth to a peak in the early twenties and then declines steadily as age increases (Fig. 8.4).

Years Lived with a Disability

In order to measure years lived with a disability for the GBD study, several assumptions were made and several adjustments were carried out in the basic morbidity data. Health conditions were ranked in severity and data were secured on the duration of the conditions in individual lives. A decision had to be made whether society gains more value from experiencing a year of healthy life in the present as compared with experiencing it at some time in the future. Still another issue that had to be decided is the selection of weights for combining years of premature death and years of disability.

Quantifying years of life lived with a disability. For validly comparing the various disabilities, incorporating them in the DALY score, and implementing public policy, disability must be defined, measured, and assigned a numerical value. In the GBD study, YLD was defined as years lived with a disability of specified severity and duration, and different weights were given to disabilities based on these two attributes. Severe and mild disabilities were distinguished and delineated. For example, a year with blindness was considered as more severe than a year lived

with diarrhea, and quadriplegia was regarded as more severe than blindness. These judgments were made formal and explicit so that they could be incorporated into measurements of the disease burden. The number of years lived with a disability was calculated from information on the incidence of the disability, its age of onset, its estimated duration, and its severity. The years lived with a disability in each incident, adjusted for severity, were multiplied by the number of cases in the population.

One must make a trade-off in time units between the years lived with a particular disability and years of perfect life. The trade-off is between saving one life-year for 1,000 healthy individuals and, for example, saving one life-year for 2,000 individuals in a worse health state.[9] Health care systems often make such trade-offs implicitly. The GBD system measures and sets forth these choices explicitly. With the aid of advice from a board of 8 to 12 advisors from around the world, GDB established weights from 0 (perfect health) to 1 (equivalent to death) for the severity of a set of 22 indicators – disabling conditions, such as blindness, depression, and conditions that cause pain. These weights were then grouped into seven classes according to severity, with class I having weights between 0.00 and 0.02 and class seven having weights between 0.7 and 1.0 (Table 8.10). The weights for the 22 indicator conditions were determined by asking members of the advisory board to make the kinds of trade-offs noted above. Once weights had been assigned to the 22 indicator conditions, the board members assigned the remaining health conditions to the seven classes.

Other issues to be resolved. Another issue to be resolved is the relative worth of a year of healthy life to society now or at some future year, says in 30 years' time. Arguments have been given on both sides. Some would argue in favor of a year of healthy life earned now. In this interpretation future years are discounted as compared with current years. Others would argue the opposite position. The GBD researchers decided to discount future years of life by three percent per year. This means that a year of healthy life lived 10 years from the present is worth about 26% less than one lived now (computed as an exponential decay function). Discounting future health for a child reduces the relative impact of a child's death compared with an adult's death. For example, if age-weighting and discounting are both incorporated into the calculations, a one-year-old girl's death causes a loss of 34 years of life while a 25-year-old woman's death results in a loss of 33 years of life. Discounting reduces the value of interventions that pay off largely in the future; for example, vaccination against hepatitis B may prevent thousands of cases of liver cancer, but some decades later.

The question remains as to how to combine time lost due to premature deaths and time lived with disability. Should they be combined with different weights or

[9] In essence, the weight is set in terms of the number of people with a given condition whose claim on a fixed healthcare budget is equal, in the judgment of a member of the advisory board, to that of 1,000 healthy people. For example, if the member judges that 1,000 healthy people have an equal claim on the fixed resources as 8,000 people with some severe disability, the weight assigned to that particular disability is equal to $[1 - (1,000 \div 8,000)]$, or 0.875.

Table 8.10 Disability classes and severity weights set by the GBD study for 22 indicator health conditions

Disability class	Severity weights	Health condition
1	0.00–0.02	Vitiligo on face, weight-for-height less than two sigma
2	0.02–0.12	Watery diarrhea, severe sore throat, severe anemia
3	0.12–0.24	Radius fracture in a stiff cast, infertility, erectile dysfunction, rheumatoid arthritis, angina
4	0.24–0.36	Below-the-knee amputation, deafness
5	0.36–0.50	Rectovaginal fistula, mild mental retardation, Down syndrome
6	0.50–0.70	Unipolar major depression, blindness, paraplegia
7	0.70–1.00	Active psychosis, dementia, severe migraine, and quadriplegia

Source: Murray and Lopez with and the World Bank (1996), Table 1, p. 11. Copyright © 1996 World Health Organization. Reprinted with permission
Note: These weights were established, using the person trade-off method, with the advice of an international group of health workers. Each condition named represents a defined group of symptoms; for example, angina here is defined as reproducible chest pain, when walking 50 meters or more, that the individual would rate as a 5 on a subjective pain scale from 0 to 10

with the same weights, that is, by simple addition? The GBD study simply added the two values. The question may be asked whether the results would vary significantly by weighting the components, the ages of premature deaths and time lived with disability, differently. Changes could involve varying the method of weighting the various disabilities or changing the discount rate. These weighting schemes could possibly have a great effect on the number of years of healthy life calculated for people with different disabilities. Tests have indicated that the final measures of disease burden in terms of broad-cause groups are largely unaffected by a different pattern of age-weighting and are only slightly affected by changing the method for weighting disability on the basis of severity. Changes in the discount rate, on the other hand, can have a more substantial effect on the overall results. The most significant effect of changing the discount rate and age weights is to reduce the importance of several psychiatric conditions. The GBD researchers concluded that the accuracy of the underlying basic epidemiological data from which the disease burden is calculated would influence the final results much more than the discount rate, the age-weighting method, or the disability-weighting method. Improvements in the basic data would be more fruitful than small adjustments to the measure itself.

Concluding Note on QALY

Quality-adjusted life years is a widely used measure of both quantity and quality of life. The GBD's findings show the importance of disability in evaluating the health

status of the world's population and, in particular, the importance of encompassing mental health conditions in the evaluation since the incidence of such conditions has been shown to be both very great and grossly underestimated. Health providers and researchers who evaluate the effectiveness and cost of various health procedures or use the QALY in clinical and patient decision making still do not agree on how best to compute QALY, some seeking an alternative and others merely seeking to refine it. Questions include, what is being evaluated, whether preference weights required for estimating the measure should come from patients or the larger health care community, consumers, or policymakers, whether assigning more weight to youth than older persons is ethical, and what the uses of QUALY are.

Other Measures of Health-Related Quality of Life

Other measures of the health-related quality of life, using more inclusive definitions of health, have been constructed or proposed. Some measures include components on access to and use of health services in addition to morbidity and mortality components. Others focus more on the quality of life in a broad sense – i.e., well-being – than on formal indications of physical ill health. These measures may reflect such psychological and social aspects of health as mental health, cognitive functioning, social functioning, intimacy, happiness, and productivity, in addition to physical health and access to and use of health services. They may incorporate combinations of health measures, such as self-perceived health and a measure of cognitive functioning, or self-perceived health, a mortality measure, and some marginally pathological conditions. Among the latter may be such activity-limiting (even if not severely limiting), time-consuming, discomforting, and health-compromising conditions as poor sleep, backache, frequent fatigue, headaches, allergies, dental problems, and elimination problems.

As an example, consider the set of measures developed by the U.S. Centers for Disease Control and Prevention (CDC) to track population health-related quality of life (U.S. CDC 2000). The assessment consists of four core questions about "healthy days" that include self-assessed health, and days spent with physical illness, mental illness, and the inability to pursue one's usual activities within the past thirty days. These four items are used to create a summary index of unhealthy days. An additional ten questions can be included with the four core questions to secure more detailed information about healthy days. These additional ten questions ask about any activity limitations and days with pain, depression, anxiety, sleeplessness, or lack of vitality. Another effort to measure health-related quality of life is the Medical Outcomes Study Short Forms (SF-12 and SF-36). This study contains questions on self-assessment of general health and questions about activities and functional limitations, physical and emotional problems, and pain (Ware et al. 2000). Additional measures have been proposed. The definition and measurement of the health quality of life continue to be areas of research and development among the specialists.

Calculation of the Health Life Cycle

At the beginning of this chapter a chart that depicts the relation of survival chances of persons in various health conditions was presented. This is a stylized depiction of the actual sequence of events, since the experience of individuals can vary greatly from such generalized representations. Another generalized construct of interest is the health life cycle, a quantitative record of the sequence of critical health stages that an individual experiences between birth and death. The paradigm can be used as a framework for the study of variations in the health of the members of a population. It is useful for comparing the health history of different groups or the same group over time.

Life cycle changes are usually measured in terms of the mean or median ages at which the critical events occur. Such life cycles have been mapped for families (Spanier et al. 1985; Schoen et al. 1985) and described for the education/work/retirement phases of the life cycle (Siegel 1993). To construct such a life cycle for health, we need information on the median or mean ages of the critical health events in life, for example, the first experience of a major chronic disease, first IADL, first ADL, and death. The available data are still very limited for calculating these measures. Experts disagree on the definition of the critical events in the health cycle that lead from a state of health to a state of death, let alone the median ages of the onset of the various critical health events. Nevertheless, it is possible to begin developing such a health cycle with the existing data. Such a schema would show the median or mean age, for males and females, for a succession of health stages such as noted above.

References and Suggested Readings

General Theoretical Background

Alter, G. C., & Riley, J. C. (1989). Frailty, sickness, and death: Models of morbidity and mortality in historical populations. *Milbank Quarterly*, 67(2), 208–232.

Fries J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medecine*, 303(3), 130–135.

Gruenberg, E. M. (1977). The failure of success. *Milbank Memorial Fund Quarterly: Health and Society*, 55(1), 3–24.

Lamb, V. L., & Siegel, J. S. (2004). Health demography. In J. S. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 341–370). San Diego, CA: Elsevier/Academic.

Manton, K. G. (1982). Changing concepts of morbidity and mortality in the elderly population. *Milbank Memorial Fund Quarterly: Health and Society*, 60(2), 183–244.

Riley, J. C., & Alter, G.C. (1996). The sick and the well: Adult health in Great Britain during the health transition. *Health Transition Review*, 6, 19–44.

U.S. National Center for Health Statistics. (2001). *Healthy people 2000: Final review*. Hyattsville, MD: National Center for Health Statistics.
U.S. National Center for Health Statistics. (2001). *Healthy people* 2010 (2nd ed., 2 Vols.). Hyattsville, MD: National Center for Health Statistics.

Measurement of Active Life Expectancy

Brouard, N., Lièvre, A., & Heathcote, C. R. (2007). *Maximum likelihood computer program using interpolation of Markov chains*. Paris: Institut National d'Etudes Démographiqes. Available from http://euroreves.ined.fr/imach. Accessed 2007.
Cambois, E., Robine, J. -M., & Brouard, N. (1999). Life expectancies applied to specific statuses: a history of the indicators and the methods of calculation. *Population, 3, 1998. Population: An English Selection*, 11, 1999. Paris: INED.
Crimmins, E. M., Saito, Y. & Ingegneri, D. (1989). Changes in life expectancy and disability-free life expectancy in the United States. *Population and Development Review*, 15(2), 235–267.
Crimmins, E. M., Saito, Y. & Hayward, M. D. (1993). Sullivan and multistate methods of estimating life expectancy: Two methods, two answers. In J. M. Robine, C. D. Mathers, M. R. Bone, & I. Romieu (Eds.), *Calculation of health expectancies: Harmonization, consensus achieved and future perspectives* (pp. 155–160). Montrouge, France: John Libbey Eurotext.
Crimmins, E. M., Hayward, M. D., & Saito, Y. (1994). Changing mortality and morbidity rates and the health status and life expectancy of the older population. *Demography*, 31, 150–175.
Crimmins, E. M., Hayward, M. D., & Saito, Y. (1996). Differentials in active life expectancy in the older population of the United States. *Journal of Gerontology: Social Sciences*, 51(B3), S111–S120.
Crimmins, E. M., Saito, Y. & Ingegneri, D. (1997). Trends in disability-free life expectancy in the United States, 1970–1990. *Population and Development Review*, 23(3), 555–572.
Geronomus, A. T., Bound, J., Waldmann, T. A. Colen, C. G., & Steffick, D. (2001). Inequalities in life expectancy, functional status, and active life expectancy across selected black and white populations in the United States. *Demography*, 38, 227–251.
Guralnik J. M., Land K. C., Blazer D, Fillenbaum G. G., & Branch L. G. (1993). Educational status and life expectancy among older blacks and whites. *NewEngland Journal of Medicine*, 329 (2), 110–6
Hayward, M. D., & Grady, W. R. (1990).Work and retirement among a cohort of older men in the United States, 1966–1983. *Demography*, 27(3), 337–356.
Hayward, M. D., & Heron, M. (1999). Racial inequality in active life among adult Americans. *Demography*, 36(1), 77–91.
Jagger, C. (1997). Health expectancy calculation by the Sullivan method: A practical guide. Montpelier, VT: Euro-REVES.
Katz, S., Branch, L. G. Branson, M. H. Papsidero, J. A., et al. (1983). Active life expectancy. *New England Journal of Medicine*, 309(20), 1218–1223.
Laditka, S. B., & Hayward, M. D. (2003). The evaluation of demographic methods to calculate life expectancy. In J.- M. Robine, C. Jagger, C. D. Mathers, E. Crimmins, & R. Suzman (Eds.), *Determinants of health expectancy* (pp. 221–234). Chichester, UK: Wiley.
Laditka, S. B., & Wolf, D. A. (1998). New methods for analyzing active life expectancy. *Journal of Aging and Health*, 10(2), 214–241.
Land, K. C., Guralnik, J. M. & Blazer, D. G. (1994). Estimating increment-decrement life tables with multiple covariates from panel data: The case of active life expectancy. *Demography*, 31, 297–319.
Lynch, S. M, & Brown, J. S. (2005).A new approach to estimating life tables with covariates and constructing interval estimates of life table quantities. *Sociological Methodology*, 35(1), 189–238.

Manton, K. G., & Land, K. C. (2000). Active life expectancy estimates for the U.S. elderly population: A multidimensional continuous-mixture model of functional change applied to completed cohorts, 1982–1996. *Demography*, 37(3), 253–266.

Manton, K. G., & Stallard, E. (1991). Cross-sectional estimates of active life expectancy for the U.S. elderly and the oldest-old populations. *Journal of Gerontology: Social Sciences*, 46, 170–182.

Manton, K. G., Stallard, E., & Corder, L. (1995). Changes in morbidity and chronic disability in the U.S. elderly population: Evidence from the 1982, 1984, and 1989 National Long-Term Survey. *Journal of Gerontology: Social Sciences*, 50B(4), S194–S204.

Manton, K. G., Stallard, E., & Corder, L. (1997). Changes in the age dependence of mortality and disability: Cohort and other determinants. *Demography*, 34(1), 135–157.

Manton, K. G., Gu, X. & Lamb, V. L. (2005). Long-term trends in life expectancy and active life expectancy in the United States. *Population and Development Review*, 32(1), 81–105.

Manton, K. G., Gu, X., & Lamb, V. L. (2006). Changes in chronic disability from 1982 to 2004/2005 as measured by long-term changes in function and health in the U.S. elderly population. *Proceedings National Academy of Sciences*, 103(48), 18374–18379.

Mathers, C. D., Sadana, R., Salomon, J. A., Murray, C. J. L., & Lopez, A. D. (2001). *Healthy life expectancy in 191 countries*, 1999. Lancet, 357, 1685–1691.

Palloni, A., Ayuso, M., Guillén, M., & Monteverde, M. (2005, March 31–April 2). *A microsimulation model to estimate errors in cross-sectional estimates of disability-adjusted life expectancy*. Paper presented at the annual meeting of the Population Association of America, Philadelphia.

Robine, J- M., & Ritchie, K. (1991). Health life expectancy: Evaluation of a new global indicator for change in population health. *British Medical Journal*, 302, 457–60.

Robine, J.- M., & Romieu, I. (1998). *Health expectancies in the European Union, progress achieved* (REVES paper 319). Montpelier, France: INSERM.

Robine, J.- M., Romieu, I., & Cambois, E. (1999). Health expectancy indicators. *Demography*, 13(2), 287–290.

Robine, J- M., Jagger, C., & Romieu, I. (2001). Disability-free life expectancies in the European Union countries: Calculation and comparisons. *Genus*, LVII(2), 89–101.

Robine, J- M., Jagger, C., Mathers, C.D., Crimmins, E.M., & Suzman, R.M. (Eds.). (2003). *Determining health expectancies*. Chichester, UK: Wiley.

Rogers, A, & Ledent J. (1976). Increment-decrement life tables: A comment. *Demography* 13, 287–290.

Rogers, A., Rogers, R.G., & Belanger, A. (1990). Longer life but worsening health? Measurement and dynamics. *Gerontologist*, 30, 640–649.

Rogers, R. G., Rogers, A., & Belanger, A. (1989). Active life among the elderly in the United States: Multistate life table estimates and population projections. *Milbank Memorial Fund Quarterly*, 67, 370–411.

Sanders, B. S. (1964). Measuring community health levels. *American Journal of Public Health*, 54(7), 1063–1070.

Schoen, R. (1988). The multistate life table. In R. Schoen (Ed.), *Modeling multigroup populations* (Chap. 4). New York: Plenum Press.

Siegel, J. S. (2002), Applied demography: *Applications to business, government, law, and public policy, San Diego*, CA: Academic.

Sullivan, D. F. (1971). A single index of mortality and morbidity. *HSMHA Health Report*, 86(4), 347–354.

U.S. Department of Health and Human Services (DHHS). (2003). *Healthy people*, 2010. Table 3.2.

U.S. National Center for Health Statistics. (2001a). Summary measures of population health: Methods for calculating healthy life expectancy. In M. T. Molla, D. K. Wagener, & J. H. Madans (Eds.), *Healthy people 2010: Statistical notes*, No. 21. Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2001b). Summary measures of population health: Addressing the first goal of healthy people 2010, Improving health expectancy. By D. K. Wagener,

M. T. Molla, E. M. Crimmins, E. Pamuk, & J. H. Madans. *Healthy people 2010: Statistical notes*, No. 22. Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2008). Estimating healthy life expectancies using longitudinal data: Methods and techniques in population health measures. By M. T. Molla, & J. H. Madans. *Vital and Health Statistics*, 2(146).

U.S. Office of the Chief Actuary, Social Security Administration. (2006). A death and disability life table for insured workers born in 1985. By R. Baldwin & S Chu. *Actuarial Note*, Number 2005.6. Baltimore: Social Security Administration.

Willekens, F. J., Shah, I., Shah, M., & Ramachandran, P. (1982). Multistate analysis of marital status life table: Theory and applications, *Population Studies*, 36, 129–144.

World Health Organization. (1985). Targets for health for all: *Targets in support of the European Regional Strategy for health for all*. Copenhagen, Denmark: WHO Regional Office for Europe.

World Health Organization. (2004). *World health report*. Statistical Annex Table 8.4. Healthy life expectancy (HALE) in all WHO Member States, 2002. Geneva, Switzerland: World Health Organization.

Measures of Health-Related Quality of Life

Fryback, D. (1998). Methodological issues in measuring health status and health-related quality of life for population health measures: A brief overview of the HALY family of measures. In M. J. Field & M. R. Gold (Eds.), *Summarizing population health: Directions for the development and application of population metrics*. Washington, DC: National Academy Press/Institute of Medicine.

Hyder, A. A., & Morrow, R.H.. (2002). Healthy life years (HeaLYs). In C. L. J. Murray, J. A. Salomon, C. D. Mathers, & A. D. Lopez (Eds.), *Summary measures of population health: Concepts, ethics, measurement and applications* (pp. 245–257). Geneva, Switzerland: World Health Organization.

Institute of Medicine. (1998). *Summarizing population health: Directions for the development and application of population metrics.* (M. J. Field, & M. R. Gold, Eds.). Washington, DC: National Academy Press.

Murray, C. J. L., & Lopez, A. D. (Eds.). (1996a). The *global burden of disease: Summary*. (Global burden of disease and injury series.) Geneva, Switzerland: World Health Organization, World Bank, and Cambridge, MA: Harvard School of Public Health.

Murray, C. J. L., & Lopez, A. D. (Eds.) (1996b). *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors, and projected to 2020.* Cambridge, MA: Harvard University Press.

Murray, C. J. L., Salomon, J., & Mathers, C. (1999). A critical examination of summary measures of population health. In Global programme on evidence for health policy, Discussion paper No.2. Geneva, Switzerland: World Health Organization.

U.S. Centers for Disease Control and Prevention. (2000). *Measuring healthy days.* Atlanta, GA: Centers for Disease Control and Prevention.

U.S. National Center for Health Statistics. (1995). Years of healthy *life. Healthy People Statistical Notes,* No. 7. By P.W. Erickson, R. Wilson, and I. Shannon. Hyattsville, MD: National Center for Health Statistics.

U.S. National Center for Health Statistics. (2001). *Healthy people 2000: Final review.* Hyattsville, MD: National Center for Health Statistics.

U.S. National Center for Health Statistics. (2003). Summary measures of population health: Report of findings on methodologic and data issues. By M. T. Molla, J. H. Madans, D. K. Wagener, & E. M. Crimmins. *Healthy people 2010.* Hyattsville, MD: National Center for Health Statistics.

Ware, J. E., Kosinski, M., & Dewey, J. E. (2000). *How to score Version 2 of the SF-36^R health survey.* Lincoln, RI: Quality Metric.

Weinstein, M. (2005). Spending health care dollars wisely: Can cost-effectiveness analysis help? *Policy Brief 30.* New York: Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University.
World Health Organization. (2004). *World health report.* Statistical Annex Table 8.3, Burden of Disease in DALYs in WHO Regions, 2002. Geneva, Switzerland: World Health Organization.

Lifetime Risk of Nursing Home Residency

Foley, D. J., Ostfeld, A. M., Branch, L. G., & Wallace, R.B. (1992). The risk of nursing home admission in three communities. *Journal of Aging and Health,* 4(2), 155–173.
Kemper, P., & Murtaugh, C. (1991). Lifetime use of nursing-home care. *New England Journal of Medicine,* 324(9), 595–600.
Liang, J., & Tu, E. J-C. (1986). Estimated lifetime risk of nursing home residency: A further note. *Gerontologist,* 26(5), 560–563.
McConnel, C. E. (1984). A note on the lifetime risk of nursing home residency. *Gerontologist,* 24(2), 193–198.
Palmore, E. (1976). Total chance of institutionalization among the aged. *Gerontologist,* 16, 504–507.

Life and Health Cycles

Schoen, R., Urton, W., Woodward, K., & Baj, J. (1985). Marriage and divorce in twentieth century American cohorts. *Demography,* 22(1), 101–114.
Siegel, J. S. (1993). *A generation of change: A profile of America's older population.* New York: Russell Sage Foundation (Chap. 7).
Spanier, G. B., Roos, P. A., & Shockey, J. (1985). Marital trajectories of American women: Variations in the life course. *Marriage and the Family,* 47, 993–1003.

Chapter 9
Reproductive Health

Definition, Sources, and Quality of Data

Demographers' interest in reproductive health stems largely from their interest in fertility. Various female health conditions, e.g., reproductive tract infections, tend to reduce fertility or cause infertility. Family planning, an intrinsic component of reproductive health, has had an important impact on child and maternal mortality, as well as on the level of fertility and child spacing. The emergence of HIV/AIDS has heightened the interest of demographers in reproductive health. It is a leading cause of mortality, but it has also reduced fertility, diminished population growth, and modified age structures. It has restrained sexual and marital behavior, been responsible for transmission of disease from mothers to newborn children, played a possible role in international migration flows, and increased orphanhood and family destabilization.

Definition

In the United Nations 1994 World Plan of Action, reproductive health is defined as a state of complete physical, mental, and social well-being… in all matters relating to the reproductive system and to its functions and processes (United Nations 1994). Accordingly, reproductive health is concerned with the health correlates of reproductive events (*i.e.*, sexual activity, conception, pregnancy, fetal losses, childbirth, and the postpartum experience), including the ability to bear healthy children, avoid pregnancy losses, regulate fertility, and engage in satisfying sexual behavior without fear of disease or unwanted pregnancy. This definition goes well beyond the provision of family planning services. Reproductive health includes not only the issues relating to family planning but also those relating to diseases associated with reproductive events, such as HIV/AIDS, reproductive tracts disorders, and other sexually transmitted diseases, and to the measurement of

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_9,

sexuality, including the patterns of sexual behavior. These patterns are important in affecting the exposure to sexually transmitted diseases and reproductive tract infections, the choice and usage of contraceptive methods, the risk of becoming pregnant, and other health aspects of reproductive events.

Sources and Limitations of Data

The organizations that have provided information about reproductive health, with examples of their data products, include: The World Health Organization (*e.g.*, maternal mortality, factors in contraceptive use); Family Health International (*e.g.*, maternal mortality, reproductive morbidity, and sexually transmitted diseases); the International Women's Health Coalition (*e.g.*, reproductive tract infections); Macro International Inc. (*e.g.*, Demographic and Health Surveys); the London Maternal and Child Epidemiology Unit (*e.g.*, various health projects in Less Developed Countries); the Alan Guttacher Institute (*e.g.*, abortions); the U.S. National Center for Health Statistics (*e.g.*, maternal mortality, fetal losses, family planning practices); and the Population Council (*e.g.*, numerous community studies relating to reproductive health). The U.S. Agency for International Development and the National Institutes of Health fund some of these organizations and their programs.

Reproductive health is measured by a combination of self-reports obtained in sample surveys, clinical examinations, and laboratory analyses. Measuring reproductive health presents special problems because of the likely inconsistency between the results of different measures, the trivialized view of reproductive health conditions by the community and physicians, the fact that many conditions are asymptomatic, and the fact that some result from clandestine behavior (Obermeyer 1996; WHO 1989; Stewart et al. 1996). The discrepancy between objective and subjective information is especially problematic for the study of reproductive health. The validation of health indicators, such as by clinical examination, is rare although some studies do compare self-reports of health conditions and actual medical diagnoses (Zurayk et al. 1993).

Studies of reproductive health involving interviews pose problems that are common to all surveys but are especially difficult for this area of study: (1) how to design the studies, (2) how to access the women respondents, (3) how to ask the appropriate questions, and (4) how to interpret the responses. The role of the interviewer is critical. Whether the interviewer is a male or female, of the same or different social class as the respondent, a physician or not, or a stranger to the respondent or not, can affect the quality of the data collected. Even when the respondent wants to be fully cooperative, recall bias is a serious problem in surveys of reproductive health.

Respondents' perceptions of what is healthy and unhealthy affect the level and comparability of the measures based on the health conditions reported. Results for many less developed countries are biased by the fact that a culture of silence

heavily influences the reporting of reproductive morbidity because there are taboos regarding public discussion of certain private matters. Women tend to tolerate reproductive morbidity as part of their reproductive experience. Furthermore, women's problems are not viewed with much seriousness in these countries. This is so even from a medical point of view. Yet these conditions do affect a woman's general functioning. Because access to medical services is limited in many less developed countries by culture and geographic conditions, many women are unaware that they have a definite illness. In many Islamic and Hindu communities, the practice of Purdah forbids women from being seen in a hospital or clinic unless the husband's permission is obtained.

Collecting Data on Sensitive Subjects

In research on reproductive health, data on some sensitive subjects, such as use of condoms and number of sexual partners, are obtained. In Chap. 2, I briefly described two special methods of securing data on sensitive subjects, the randomized-response technique and the 3-card method. Here I add a few other less exotic methods that are applicable to the Less Developed Countries: direct interview, indirect interview, and use of local reporters.

In spite of the sensitivity of some subjects, direct interview of respondents has proved satisfactory under some conditions in less developed settings. The same approach in the more developed nations might have failed, and complex devices like the 3-card method or the randomized-response technique have had to be employed there. Two other techniques for securing information from respondents on sensitive subjects have been employed. In the indirect interview, computers rather than interviewers alone have been used in the questioning of some adolescents on the expectation that they might tell the truth about sensitive subjects if they have a degree of anonymity. The questions are asked through earphones and the respondent types out the responses on a keypad. In another device, the empirical data are supplemented through diaries kept by local villagers. Researchers learn what people are saying to their friends, neighbors, and relatives by using local journalists, who keep diaries of conversations they overhear. More specifically, the selected villagers have been asked to record, in hand-written journals, conversations related to the sensitive subject that they may have overheard or participated in. Accessing these journals, the researchers learn what members of the population are saying to each other about such health events as the HIV/AIDS epidemic, for example.[1]

[1]These two techniques were employed in a University of Pennsylvania/University of Malawi joint study, initiated in 1998, covering the responses of people in small rural villages of Malawi to the AIDS epidemic (Watkins 2004). As part of this study, HIV tests were also performed and, where there were no logistical problems, this device for securing sensitive data was deemed successful.

General Characteristics of Reproductive Health Studies

Studies of reproductive morbidity can be hospital-based or community-based. Formerly, studies of reproductive morbidity in the community were almost rare in the LDC (Zurayk et al. 1993) and most of the available studies were hospital-based, not community-based, even though hospital-based studies do not give a representative description of the extent of maternal mortality or reproductive morbidity in the community. The geographic focus was uneven (*e.g.*, Nigeria, Ghana, Gambia, India, Singapore, and Bangladesh), as was the focus on particular morbid conditions. Reproductive morbidity studies can deal with specific morbid conditions or a wide range of such conditions. In recent years the list of countries surveyed has expanded greatly. Numerous community reproductive health surveys have been taken, from those in the western industrialized nations (*e.g.*, U.S. Survey of Family Growth) to the many Demographic and Health Surveys taken in less developed nations.

The risk factors for reproductive health include both proximate and contextual variables (Obermeyer 1996). Some variables, such as age, parity, pregnancy history, use of health services, and medical history can be measured quantitatively with some precision. Other variables, such as the use of contraceptive methods, sexual practices, and practices relating to personal hygiene, are also quantifiable but are difficult to measure. Still others deal with perceptions, motivations, attitudes, and psychological contextual factors, such as religiosity, feelings of control, and interpretations of morbid conditions, and are hard to quantify and extremely difficult to measure. Inclusion of all three types of variables in research studies on reproductive health is necessary for the proper interpretation of trends and for use in designing programs for improving reproductive health conditions.

Biodemographic Concepts

Before proceeding to a discussion of the methods of analyzing reproductive health, I want to describe various biodemographic concepts related to reproductive health and their fundamental characteristics and relationships. I consider a wide range of concepts: Fetal sex ratio, sex ratio at birth, age at menarche, age at menopause, length of the reproductive period, initiation of sexual activity, timing of marriage, fecundity, fertility, fecundability, the age pattern of natural and observed fertility, and reproductive aging.

Fetal Sex Ratio and the Sex Ratio at Birth

Fetal Sex Ratio

The fetal sex ratio is defined as the ratio of male fetuses to female fetuses that have completed a uterogestation period of 20 or 28 weeks of more in a given year

(expressed per 100). It is not feasible to include fetuses with shorter periods of uterogestation in statistical studies because of the incompleteness of the reporting of early-stage fetuses and the indeterminacy of the sex of the fetus at this stage. However, attempts have been made to reconstruct the sex ratio back to conception in order to ascertain its level at that time and the amount of change in it from conception (designated the primary sex ratio) to birth (designated the secondary sex ratio). Again, the efforts to determine the relative loss of male and female fetuses over the entire period of uterogestation are frustrated by the great incompleteness of the data on early fetal losses and the lack of information on their sex identification.

Probability theory suggests that the sex ratio at conception is 100. However, the empirical evidence argues differently. First, the sex ratio at birth has commonly registered a value above 100. For example, the sex ratio at birth in the United States was 105.5 in 1970, 105.0 in 1993, and 104.8 in 2004 (Table 9.1). Next, the sex ratio of fetal losses (20 weeks of uterogestation or more) is much higher (*e.g.*, 118 in 1970, 110 in 1993, and 114 in 2004). (See U.S. NCHS/MacDorman et al. 2007.) Still higher fetal sex ratios are found for Japan – for example, 127.3 in 1960 and 209.6 in 1999; these were associated with sex ratios at birth of 107.1 in 1970 and 105.5 in 1999 (Davis et al. 2007). It is evident then that the fetal sex ratio at the beginning of intermediate fetal life in the United States and Japan is higher than the sex ratio of births. Information on the sex ratio of (induced) abortions is not available but there is no reason to believe that the balance of the sexes among them is affected by the mild sex preference for boys that prevails in these countries. There is also suggestive evidence that male fetuses are disproportionately represented among early fetal losses (under 20 weeks of uterogestation). Given the lack of dependable data for that period, however, we are unable to measure accurately the sex ratio at conception. Estimates of the sex ratio of fetuses at the beginning of the intermediate fetal period can be made by reconstructing pregnancies of males and females separately, from data on births, fetal losses, and (induced) abortions. The general method of doing this is described later in this chapter.

Sex Ratio at Birth

The sex ratio at birth is conventionally defined as the number of male babies per 100 female babies, although some countries formerly used the reciprocal ratio, the ratio of female babies to male babies. The balance point of the sexes is 100 but the sex ratio at birth tends to exceed 100, usually falling in the range 103–106 for national populations (Table 9.1). Demographers and social biologists have long been interested in accounting for variations in the sex ratio at birth, taking account of the characteristics of both the mother and father, the characteristics of the child, the culture, and the environment. These characteristics include principally the age, race, ethnicity, and socioeconomic status of the mother and father, the order of birth of the child, any cultural preference for boys, the practices of sex-selective abortion and female infanticide, and environmental toxins (e.g., pesticides and cigarette smoking). In addition, such factors as paternal nutrition and obesity, assisted

Table 9.1 Sex ratios at birth for selected countries: Mainly 1983–2004 (Male births per 100 female births)

Country	Year(s)	Sex ratio
Africa		
Egypt	1999	105.8
Kenya	1940–1998	101.3
Morocco	1999	104.7
South Africa	1961–1998	104.9
Tanzania	1953–1996	102.9
Tunisia	1985–1989	106.8
Uganda	1952–1995	98.8
Zambia	1954–1996	99.7
Zimbabwe	1950–1994	102.6
North America		
Canada	2000	105.6
United States	2004	104.8
Non-Hispanic whites	2004	105.3
Non-Hispanic blacks	2004	103.8
Hispanics	2004	104.2
Latin America and Caribbean		
Chile	1983–1991	104.7
Colombia	2000	105.8
Cuba	2000	108.3
El Salvador	2000	106.3
Guatemala	1983–1988	103.8
Panama	1983–1990	105.4
Uruguay	1983–1988	105.5
Venezuela	1983–1991	105.1
Asia		
Japan	2000	105.8
Malaysia	1983–1992	107.4
Pakistan	1997	107.7
Thailand	2000	105.9
Singapore	2000	109.2
Sri Lanka	1983–1987	104.4
Oceania		
Australia	2000	105.6
New Zealand	2000	106.2
Europe		
Belgium	2001	104.0
France	1983–1990	105.1
Germany	2001	105.1
Hungary	1983–1991	105.0
Italy	2001	106.1

(continued)

Table 9.1 (continued)

Country	Year(s)	Sex ratio
Netherlands	2001	104.5
Poland	1983–1991	105.8
Romania	1986–1991	105.0
Russian Fed.	2001	106.3
Spain	2001	105.7
United Kingdom	2001	105.0

Source: United Nations, *Demographic Yearbook*, 2002 and 1994; Council of Europe, *Recent Demographic Developments in Europe*, 2002; Garenne (2004) (World Fertility Survey, Demographic and Health Surveys), and U.S. NCHS (2006)

reproduction, maternal illness, and stress may play a role in sex determination of newborn children (Davis et al. 2007). The relative role of genetic and environmental factors in affecting the sex ratio at birth remains unresolved.

Race clearly matters. The sex ratio of births tends to be significantly higher for white populations than for black populations. In the United States, the figure for whites falls mostly within the range 104–106 compared to 102–104 for blacks. In 2004 the sex ratio of births was 105.3 for whites and 103.8 for blacks (Table 9.1). The long-term averages (1940–2002) were slightly higher for whites, and slightly lower for blacks, than the figures for 2004. This difference appears not only for these races in the United States but also for national populations whose residents are predominantly white or black (Table 9.1). Garenne (2002) found an average sex ratio of births from 56 surveys in Eastern and Southern Africa of 103.3, a figure significantly below the world average of 105.5.

In addition, the sex ratio of births tends to fall slightly according to the order of birth of the child and the age of the mother. While the sex ratio at birth indicates an excess of boys for all orders combined and for all ages of mothers combined, the sex ratios for first births and for young mothers tend to indicate somewhat more boy births than the ratios at the higher orders and for older mothers. With respect to age of mother, the long-term (1940–2002) U.S. averages are 105.3 for births to mothers aged 15–19 and 103.8 for mothers aged 40–44. With respect to order of birth, the long-term (1943–2002) averages are 105.7 for first-order births and 103.1 for seventh-and-higher-order births (U.S. NCHS/Mathews and Hamilton 2005). These two factors may not have an independent effect, but the order of birth appears to be the more dominant of the two factors.

The widespread preference of parents for at least one boy baby and, in some areas, only baby boys, particularly in the less developed countries of Asia and North Africa, can be a factor affecting the observed sex ratio of births in these areas. In the countries of sub-Saharan Africa, where this preference is not a cultural norm, the sex ratios at birth are relatively very low. The United Nations attributes the relative imbalance of the sexes among births and young children, in favor

of males, in mainland China, India, Taiwan, and South Korea, to sex-selective abortion, infanticide, and food favoritism. Sex-selective abortion is now generally possible with the development of ultrasound technology. Other factors explaining the high sex ratio in these countries are abandonment of female infants and, possibly, infection of the woman by the hepatitis B virus, which renders her more likely to give birth to a son. Where sex-selective abortion has been declared illegal, infanticide, abandonment, and food favoritism are more likely to be employed.

The recent trend of the South Korean sex ratio at birth is illustrative of the effect of son preference on the sex ratio in countries with falling fertility. While the sex ratio at birth is now approaching typical international levels, during the 1980s, when ultrasound technology was introduced, and later, the ratio for third and higher births reached strikingly high levels, peaking in 1993. The rapid decline of the birth rate was associated with sex-selective abortions, given the long-standing traditional preference for sons. After 1993 the sex ratio fell, as a result of greater independence of women, a policy prohibiting the determination of the sex of the fetus for purposes of abortion, and the reluctance of older women to have abortions. Sex ratios for the third and fourth child declined sharply from 202.7 and 223.6 in 1994 to 115.3 and 119.1 in 2008, respectively. Apparently the preference for sons is less intensive currently even though it persists.

Age at Menarche, Age at Menopause, and Length of the Reproductive Period

Six dimensions of reproductive life are examined in the following sections: Age of menarche, age of menopause, length of the reproductive period, age patterns of fecundity and fertility over the reproductive period, and reproductive aging. The first three of these topics are discussed in this section and the last three are discussed in the following section. Most of the available information pertains to females, and the following discussion refers principally to women.

Age at Menarche

Both the age at menarche and the age at menopause are defined in terms of the average ages at which these events begin in large populations. These ages fall within the range of ages at which children have actually been born to women; some women reach menarche and menopause outside the average ages. There are records showing the birth of children to mothers as young as age 10 and to mothers as old as age 63. However, very few girls or women give birth at such extreme ages. Reports of births to girls below age 12 and to women above age 55 are more likely to represent age-reporting errors than correctly reported events.

The mean age at menarche varies greatly among the countries of the world but varies little within countries. In the United States the figure was 12.6 years in 2002

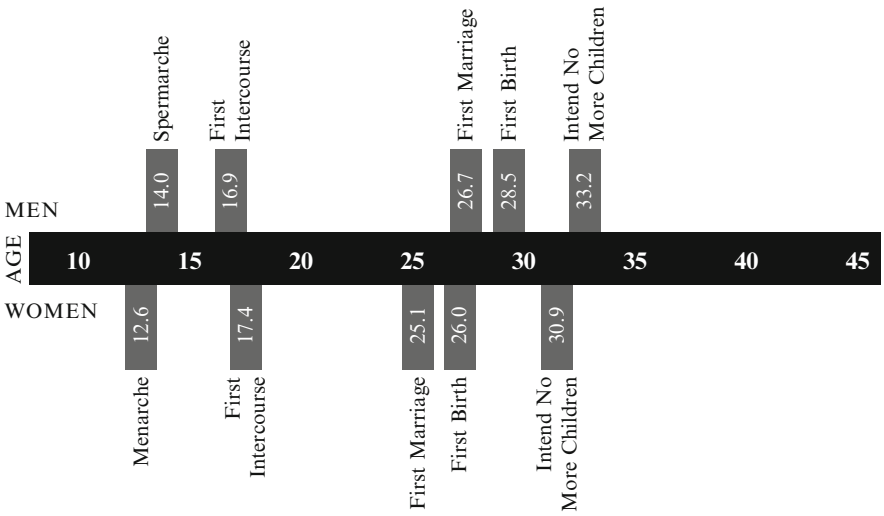


Exhibit 9.1 Sexual and reproductive timelime for males and females in the United States: Around 2000 (Source: Alan Guttmacher Institute 2002. Reprinted with permission of the Guttmacher Institute. <http://www.Guttmacher.org/pubs/fb-10-02.pdf>, accessed June 5, 2008. Based on Alan Guttmacher Institute, *In Their Own Right*, p. 8)

(U.S. NCHS/Chandra et al. 2005) (See also Exhibit 9.1). It has long been recognized that age at menarche is lower in the More Developed Countries (MDC) than in the Less Developed Countries (LDC) (See Exhibit 9.2). Currently in the MDC the mean age when menarche begins is about 13.1 years (Thomas et al. 2001). In the LDC it is a year or so higher – about 14.1 – averaging 13.8 in Asia and 14.1 in Africa and ranging up to 15.4 in Haiti. In the past, many more countries had higher menarcheal ages, but there has been a secular decline of about 3 years in the age of menarche over the last century in the MDC (Wyshak and Frisch 1982) and probably a similar decline in the LDC. While the occurrence and general timing of this event are clearly biological phenomena, the expression of its timing is affected by the environmental and socioeconomic conditions in which girls are reared. The striking progress in the health of young women in the industrialized countries in the last century, achieved mainly by public health improvements and changes in diet and lifestyle, appears to have been the main factor that contributed to the considerable decline in the age at menarche in these countries (Frisch 2002). Another principal factor accounting for the change, in addition to improved nutrition and health, is the reduced physical labor of the inhabitants of the MDC as compared with the LDC. This progress resulted at first in a wide gap in the age of menarche between the MDC and the LDC, but the gap is now diminishing.

Boys enter puberty about 2 years later than girls, and the puberty process for both sexes may spread over several years. Historical and international studies are lacking, however. Presumably the age of puberty for boys has been trending younger as it has for girls, and this means an increase in the potential gap between sexual maturity and

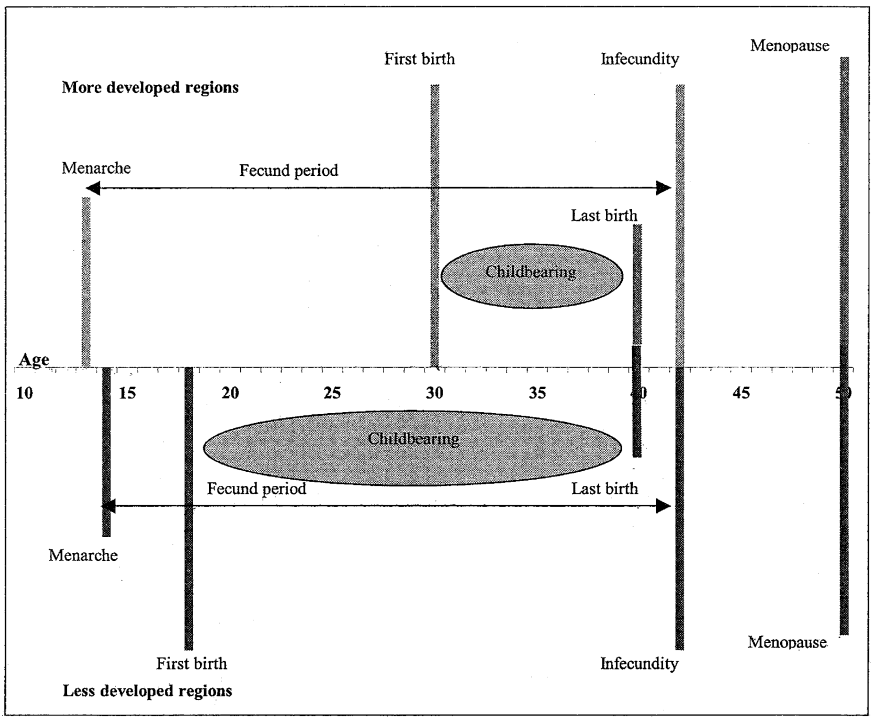


Exhibit 9.2 Generalized representation of the reproductive life span for women in the more and less developed regions: Around 1990 (Source: Alan Guttmacher Institute 2002. Based on Frank et al. (1994). The end of fertility: Age, fecundity, and fecundability in women, *Journal of Biosocial Sciences*, 26, 349–368). Copyright © 1994 Cambridge University Press. Reprinted with permission

the initiation into sexual activity and marriage. Hence, there is a widening period of exposure to unwanted pregnancy and sexually transmitted diseases, as explained further below.

Initiation of Sexual Activity

We should be concerned as much with the period of sexual activity as with the childbearing period. Historically marriage was assumed to be a satisfactory indicator of the initiation of sexual activity, which was assumed to initiate exposure to pregnancy. These assumptions are no longer valid for many countries, where the initiation of sexual activity has been largely disassociated from marriage and pregnancy. Analysts have, therefore, turned their attention to the measurement and characteristics of the initiation of sexual activity for its implications for reproductive health. Individuals initiating sexual activity are typically adolescents who are poorly informed with respect to the health risks involved and do not have

full access to reproductive health information and services. They are particularly susceptible, therefore, to unwanted pregnancies, unsafe abortions, and sexually transmitted diseases, especially HIV/AIDS. Both adolescents and youth suffer disproportionately from HIV infections and other sexually transmitted diseases.

Young women are particularly vulnerable to HIV infections for both biological and cultural reasons. According to UNAIDS (2007) infection rates in sub-Saharan Africa from HIV are five times higher among adolescent females than among adolescent males. The health risks of women are exacerbated by the traditional practice of female genital mutilation, which is still widely prevalent in sub-Saharan Africa and Northeastern Africa (Population Reference Bureau 2001).

Nearly four-fifths of the women aged 20–24 in Africa and three-fifths of the women aged 20–24 in Latin America are sexually active by age 20, most of them before age 18 and half of them before marriage (United Nations 2003). Premarital sexual initiation is common in many countries of Africa but not in all of them. The percentages of early sexual initiation are similarly high in the more developed regions. For example, in the United States premarital sexual initiation is virtually the norm. In the United States 63% of the women 20–24 years of age are sexually active before age 18, and an additional 18% at ages 18–19, most of them before marriage. In all the more developed countries except Poland, the proportion of women sexually active by age 18 exceeds one half.

Such figures are not available for young men in the United States, but sexual initiation before age 18 is also very common among adolescent males in the United States. Although menarche in girls begins earlier than puberty in boys, in the United States males begin sexual activity on average slightly earlier than females – at 16.9 years of age vs. 17.4 years of age (Alan Guttmacher Institute 2002). The figures on sexual initiation for young men in Africa and Latin America are also quite high, like those for young women. In sum, in all regions sexual initiation for both women and men occurs commonly during adolescence, with wide variations among countries.

Several factors have been identified to account for variations in the age of initiation of sexual activity and adolescent risk-taking of women. At the national level countries differ in marriage patterns, cultural norms, and socioeconomic characteristics. At the subnational level, it is informative to focus on the differences in the educational level of the woman, the influence of the family environment, and the power dynamics between the sexes. More highly educated women tend to initiate sexual activity later than those with less education. This pattern is not true for men. A favorable family environment, including the role of parents in guiding children, family stability, father's presence in the household, and parent-teen communication, also contributes to raising the age of sexual initiation. Unequal power relations between the sexes, where women are subordinated, can result in a lower age of sexual initiation for women. Such conditions may be associated with gender-based violence and sexual coercion, and result in greater vulnerability to HIV/AIDS and other sexually transmitted diseases.

The Timing of Marriage

On average, around the world the married state is still the principal setting in which childbearing occurs even though the timing of entrance into marriage does not commonly correspond to the timing of initiation of sexual activity in most societies today and the extent of unmarried parenthood is substantial in many countries. Accordingly, the timing of marriage remains quite important in its effect on reproductive health and family health. There is a notable variation in the average age (*i.e.*, singulate mean age) of marriage among the regions of the world, from 26.1 years in Europe and Northern America to 21.9 years in Africa in 1998 (United Nations 2003).[2] The mean age was 26.1 years in the United States and 30.0 years or over in several northern European countries.[3] Similarly the percent of ever-married women 15–19years of age varied from 4 in Europe and North America to 25 in Africa. In the United States and Europe today, the predominant pattern is one of late marriage. (Note, however, that, while the age at first marriage in the United States is now at an historically high level, the age at first cohabitation is much lower.) On the other hand, in the less developed regions, the age of marriage is generally much lower even though it varies over a wide range internationally, from 17 in the Niger and 18 in Bangladesh to over 26 in Namibia and Myanmar.

The fact that marriage occurs at extremely young ages in many less developed countries complicates the social situation in these countries. Women often enter into marriage at too young an age to make informed decisions pertaining to reproductive and sexual health and frequently cannot act with free and full consent. Early marriage deprives some women of decision-making power in such matters. It is associated with premature childbearing and excessive poverty. In the countries with arranged marriages, many girls get married before menarche begins. In fact, a leading purpose of early arranged marriages in the LDC is to reduce the risk of premarital pregnancy. On the other hand, in the MDC the gap between age at menarche and age at marriage has been widening as a result of the decline in the former age and the rise in the latter age. The difference has become considerable. The average interval between sexual initiation and entry into marriage exceeds 10 years in several developed countries. The marked postponement of first marriage has contributed greatly to the increase in premarital sexual activity of teenagers and youths. The widening gap has been filled in part by cohabitation as well as by promiscuity and nonmarital parenthood. The extended period between sexual maturity and marriage has special implications for the health of adolescents, who

[2]The singulate mean age at marriage is a mean based on census or survey data rather than marriage registration. It is the mean age at first marriage of those in the age range 15–49 years of age who have ever been married.

[3]Age at first marriage has fluctuated greatly in the MDC, especially in the United States. For example, the marriage age in the MDC was much lower during the Baby Boom era, as was that in the United States. The United States has always tended to have a lower age at first marriage than the MDC as a group.

are thus exposed to a long period of risk of unwanted pregnancies and sexually transmitted diseases, as noted earlier.

Consensual unions (*i.e.*, stable *de facto* unions) are recognized in many countries as an acceptable context for bearing children, and they are hardly different from marital unions with regard to living arrangements and reproductive behavior. This situation applies to the Scandinavian countries, France, and most countries of Latin America and sub-Saharan Africa. The legal status of these unions still commonly differs from that of marital unions, however. For example, in the case of dissolution of consensual unions, the women and children have less legal protection and financial support than in the case of marriage dissolution. Consensual unions tend to be less stable than marital unions and they are generally viewed as characteristic of the lower socioeconomic class.[4]

Age at Menopause

There is less information on age at menopause than on age at menarche. It is difficult to determine a woman's age at menopause in a national health survey without a probing battery of questions. The determination is difficult because menopause does not have a clearly defined beginning or ending. The biomarkers reflecting its onset and completion may appear very gradually and independently. These changes define a premenopausal period, which may last from 1 to 8 years. During this segment of the lifespan a woman's hormone (*i.e.*, estrogen) levels are more variable and are decreasing, her menstrual cycles are more irregular and longer, and there is elevated production of FSH (*i.e.*, follicular stimulating hormone) and LH (*i.e.*, luteinizing hormone). Recalling that the age of menopause is defined in terms of the mean starting age in any large population, it is estimated at about 50–51 years. This age varies only slightly from population to population, but varies considerably from woman to woman within the same population.

According to Austad (1997), as with the age at menarche, better nutrition and less physical labor are factors in raising the age at menopause and account for international differences in this age. He notes that in the nonindustrialized countries women undergo menopause a few years earlier than in the industrialized countries and that similar differences appear among the socioeconomic classes in the same country. Austad interprets this as evidence that age of menopause has increased slightly over time in the industrial countries. However, there is no empirical evidence that this part of the life cycle of females has ever changed.

[4]In the United States, consensual unions are commonly a childless state that serves as a trial period prior to marriage and has only a modest degree of legal recognition. Cohabiting couples, cohabiting relationships, or unmarried-partner households in the United States generally have no legal recognition but are accepted by much of the public with little stigma as a temporary household arrangement. A few states sanction such unions as common-law marriages.

Length of the Reproductive Period

The length of the female reproductive period is the difference between the age of menarche and the age of menopause. For populations it may be computed in two ways. The first is the mean (or median) in years of the reproductive periods of the each of the members of the survey population. The second is the difference between the mean (or median) age at menarche of the members of the survey population and the mean (or median) age at menopause of the members of that population. These measures have different values. Although they approximate one another, the first method is more precise but the second method is easier to compute. In the industrialized countries the length of the reproductive period must have increased substantially over the last century since there was a pronounced decline in the age at menarche and, probably, a small increase in the age of menopause.

The average length of life of females in the United States increased sharply between 1900 and 2000 – about 31 years – but the increase in the reproductive period amounted to only a few years. Hence, the reproductive period constitutes a much smaller proportion of the lifetime of women now than a century ago. Specifically, in the United States the reproductive period made up over two-thirds of the length of life in 1900 and less than one-half in 2000. Public health and medical developments have tended to increase the effectiveness of reproduction and the reproductive period via the increased probability of surviving childbirth, the greater survival of the woman to the end of the childbearing period, reduced fetal and infant mortality, and modern fertility technology, but later marriage and delayed childbearing have tended to reduce it.

Fecundity, Fertility, and Reproductive Aging

It is of interest to examine how fertility varies through the reproductive period, particularly in relation to the biological maximum for this variation, for the light it may throw on the optimum ages for childbearing, its effect on the total number of children born, and its effect on the health of mothers and children. Age-specific fertility schedules vary in both level and pattern. Fertility levels vary from so-called natural fertility of a few religious sects and isolated indigenous tribes to the low fertility of the western industrialized countries today, as indicated by the variation in the total fertility rate for these populations. The total fertility rate of a given year is a measure of the total number of births a woman would have in her lifetime assuming that she survived to the end of the childbearing period and that she experienced the observed age-specific fertility rates of that year. As we shall see, in spite of the tremendous variation in the level of the total fertility rate, the patterns are rather similar. Before proceeding, it will be useful to explain some biodemographic terms commonly used in fertility analysis, in particular natural fertility, fertility, fecundity, and reproductive aging.

Fecundity, Fertility, and Natural Fertility

Fecundity is the physiological capacity to bear children while fertility refers to a woman's actual birth performance. Women may be fecund or partially fecund. A woman who has borne a first child but cannot bear additional children is considered partially fecund. Fecundity changes through the childbearing period and so it is a factor in the variation in fertility age patterns. The proportion of women who remain fecund from ages 20–24 through the reproductive period has been estimated by Hyrenius (1965) as follows:

Age	Percent fecund
20–24	96
25–29	86
30–34	71
35–39	56
40–44	42
45–49	20

Source: Hyrenius (1965)

Fecundity increases from menarche to the late teens, reaches a maximum in the early twenties, and then falls to the end of the reproductive period (Kline et al. 1989). That is, an initial increase in the level of fecundity is followed by a steady decrease, and even an acceleration in the decrease at the later childbearing ages.

Natural fertility refers to the set of age-specific fertility rates that stable *de facto* unions (that is, married couples plus consensual unions) would have in the absence of birth control. It is the fertility schedule that such couples would have if the women did not try to restrict their fertility by contraception. We can still expect natural fertility levels and patterns to differ from one population to another because of differences in age at marriage, frequency of sexual intercourse, fecundity levels, post-partum infecundity, breastfeeding and weaning patterns, (spontaneous) fetal losses, the state of health and nutrition of couples, and even attitudes about family size. Although some of these factors are biological constraints others, such as age at marriage and breastfeeding and weaning practices, vary according to social customs in each society. Coale and Demeny (1966) developed the following standard age pattern of natural fertility:

Age	Pattern
15–19	1.084
20–24	1.000
25–29	0.935
30–34	0.853
35–39	0.685
40–44	0.349
45–49	0.051

Table 9.2 Various age patterns of fertility for total fertility rates of 1.2–15.0

		Pattern					
Age/year	U.S. rate	U.S. 2004	Italy 1996	Ethiopia 1990	Angola 2000	Natural fertility[a]	Maximum theoretical
Total fertility rate		2.0	1.2	7.7	7.2	11.0	15.0
Ages, 15–44	66.3	100.0	100.0	100.0	100.0	100.0	100.0
10–14	0.7	0.2	NA	NA	NA	NA[a]	NA
15–19	41.1	10.0	2.9	9.3	15.9	18.6	0.3
20–24	101.7	24.9	14.6	22.7	24.4	20.8	21.0
25–29	115.5	28.2	33.1	21.7	22.1	19.5	20.0
30–34	95.3	23.3	32.2	20.4	17.6	17.9	17.4
35–39	45.4	11.1	14.2	15.2	12.1	14.5	16.6
40–44	8.9	2.2	2.5	7.0	6.6	7.6	13.3
45–49	0.5	0.1	0.4	3.8	1.3	1.1	3.3

Source: United States: NCHS (2006);
Hutterite population and maximum theoretical: Robinson (1986); Italy, Ethiopia, and Angola: U.N. *World Population Prospects: The 2000 Revision*, Vol. I, *Comprehensive Tables.* Primary sources: U.S. and Italy, civil registration; Ethiopia, sample survey; Angola, estimated
NA not available
[a]For the Hutterite population. In the Hutterite population young births either do not occur or are not reported because marriage is not permitted until age 18

These ratios represent the relative magnitudes of the age-specific birth rates of a natural-fertility population. Another pattern of natural fertility, that of the Hutterites, is shown in Table 9.2.

Natural fertility is rare among actual populations because of the intentional control of fertility by some or most couples in nearly all modern populations. The Hutterites of the northern Plains states and the Old Order Amish are two exceptions. The Hutterites had a total fertility rate of 11–12 children per woman (Robinson 1986) and the Amish a total fertility rate of about 9 children per woman (Espenshade 1971). Even though the Hutterites marry young, are encouraged to "be fruitful and multiply," and receive good basic health care, their observed fertility levels are far below the theoretical biological maximum, which is about 15–16 children per woman. The biological maximum cannot be achieved among natural fertility populations because some couples have impaired fecundity and experience fetal losses, and women with early-life pregnancies, close spacing of pregnancies, and high parities experience excessive mortality.

Fecundity has been commonly estimated from survey data on levels of fertility (e.g., childlessness of ever-married women at ages 45–49) or survey responses to questions on fecundity status. In the MDC the extensive use of methods of fertility control limits use of percentages of childlessness of married women at ages 45–49 to inform us about the level of infecundity in these populations, as was done at an earlier era. More recently data on infecundity and fecundity impairments are directly obtained from surveys. Infecundity is measured as the failure of a couple (not surgically sterile) to conceive if they are not using contraception and have tried

to conceive for at least 12 months. About 7% of U.S. couples (the wife being 15–44 years of age) are infecund following this definition, according to the 2002 Survey of Family Growth (U.S. NCHS/Chandra et al. 2005). Impaired fecundity is also measured by specific criteria on the basis of a woman's response to survey questions relating to her fecundity status. These criteria are that the woman (not surgically sterile) believes that it is not possible for her to become a mother, that a physician has told her not to become pregnant because the pregnancy would be harmful to her or her baby, or if she has been continuously married for at least 36 months, has not used contraception, and has not become pregnant. In the United States the proportion of women 15–44 years of age reporting some form of fecundity impairment was 12% in 2002 according to the National Survey of Family Growth (U.S. NCHS/Chandra et al. 2005). This high figure may be the result of postponing childbearing.

As noted above, the biological chances of reproductive success decreases with increasing age. There is a roughly 4% failure rate at age 20 that progresses to about 20% at age 35. In some less developed regions, such as sub-Saharan Africa and Southeast Asia, there is little voluntary control of fertility, and therefore the proportion of childless women is approximately equal to the proportion of infecund women. The major cause of infecundity in these countries is reproductive tract infections, which result commonly from untreated sexually transmitted diseases. In the countries with economies in transition (*i.e.*, former socialist economies of Europe and former Soviet republics in Asia), where partial infecundity is common, however, infecundity is largely due to the high incidence and frequency of unsafe abortions.

Either member may be responsible for the infecundity of a couple. In all populations only a few percent of the couples are infecund because of purely biological reasons – that is, genetic, anatomical, or physiological reasons. Infecundity may also be acquired through lifestyle and environmental sources as well. Infecundity may result from poor nutrition and poor health, including diseases such as tuberculosis and venereal disease. Infecundity or partial infecundity may be acquired through risk-prone behavior, such as very early sexual activity or sexual activity with multiple partners, and the consequent exposure to sexually transmitted diseases.

Comparison of Observed Fertility Patterns with Natural Fertility Patterns

In western societies, intentional childlessness is more common than physically and biologically caused childlessness. In general, childlessness or low fertility results from psychological responses to social, economic, and cultural factors, leading couples to decide not to have a child or to have only one or two. Hence the observed level of fertility is considerably below the level of natural fertility. Even in the LDC, family planning is now practiced widely enough and the incidence of disease is still high enough for the highest total fertility rates to be only about 7 children per woman, as in Ethiopia today. In Italy it is only 1.2 children per woman.

The age pattern of fertility, however, shows a remarkable general similarity regardless of the level of the total fertility rate. Fertility age patterns appear to be almost entirely uninfluenced by differences in socioeconomic status, fertility and

mortality levels, and the availability and use of contraceptives. Carnes et al. 2003) found that the fertility patterns from a highly developed country (*i.e.*, a low fertility/low mortality) country, such as the United States in 1997, a natural fertility population (*i.e.*, Hutterites), and a high fertility/high mortality population (*i.e.*, a composite of pattern from Mali, 1987, Niger, 1998, and Uganda, 1987) are rather similar. Approximately 80–85% of the cumulative fertility of the women in these populations occurs by age 35. Table 9.2 displays several fertility patterns, including that of the United States in 2004, the Hutterite natural fertility pattern, a biologically maximum fertility pattern, and the fertility patterns of the lowest-fertility nation (*i.e.*, Italy) and highest-fertility nations (*i.e.*, Ethiopia and Angola). The relative similarity among the fertility patterns of these diverse populations is notable when we consider that the total fertility rate ranges from 1.2 children per woman to 15 children per woman.

Nevertheless, we can note some variation in these fertility patterns across the reproductive life span. The differences appear mainly at the youngest and oldest ages of childbearing. Compare the percentages of total fertility at ages 15–19 and ages 40–44 among the various patterns displayed in Table 9.2. Among the several factors that contribute to the differences in fertility patterns among these populations are differences in marriage patterns. For example, in areas of low fertility, women now tend to marry later and have their children later, putting their education, their careers, and their desire for additional years of independence first. While the United States falls in this group, the extremely high level of teenage fertility, most of it nonmarital fertility, in the United States is striking. By a very wide margin, the United States has the highest rate of teenage fertility of all the MDC. As shown in Table 9.2, the rate of teenage fertility in the United States is three times that of Italy and on a par with that of Ethiopia in 1990. In 2004 10% of all births in the United States were born to teenagers, and four-fifths of these occurred to unmarried women (U.S. NCHS/Martin et al. 2006). The U.S. situation may be attributed to limited sex education and the relative difficulty of obtaining contraceptives.

Patterns of Childbearing and Health Risks

The level and pattern of childbearing are associated with various health conditions of the mother and child. Early childbearing entails a heightened risk of sickness and death for young mothers (*i.e.*, mothers younger than 18 years) and their children. Similarly, "old" mothers (*i.e.*, mothers older than 34 years) and their children have disproportionately high rates of mortality and morbidity. Closely spaced births (*i.e.*, less than 2 years apart) pose similar health problems for mother and child. These health problems are due both to biodemographic factors and to socioeconomic conditions that encourage early and frequent childbearing as well as late childbearing. In the less developed countries for which data are available, including much of Asia, Latin America, the Caribbean, and sub-Saharan Africa, early childbearing is common, and a considerable proportion of births occur within 2 years of the previous birth.

In the MDC, health problems among mothers and children are becoming common in the more affluent segments of the population as well as among the lower socioeconomic segments. The childbearing choices being made put in jeopardy the intentions of many women to bear healthy children or even to bear children at all. For example, while in 2004 in the United States the maternal birth rate at age 40–44 is only about one-quarter the rate at ages 20–24 and the fecundity ratio is only two-fifths as great, the risk of bearing a child with Down's syndrome is 13 times as great for the older women as for the younger women (U.S. NCHS/Martin et al. 2006). Having allowed their most fecund years to pass by, some women become involuntarily childless, and they and others, still wishing to become mothers, are obliged to resort to artificial methods or adopt children. Although earlier menarche lengthens the fecund period for women, postponement of marriage and first births until age 30 and after reduces the effective childbearing period. Present trends in fertility practices in the industrialized countries suggest that artificial methods of fertility will be sought much more frequently in future years than at present.

There is evidence that being able to produce offspring late in the reproductive age span may contribute to the longevity of the mother. The Chinese Longitudinal Survey of Healthy Longevity is the basis of a study that concluded that late childbearing is associated with healthy longevity among the oldest-old (Yi and Vaupel 2004). Late childbearing was defined as having three or more births after age 35, or two or more births after age 40, and healthy survival was defined as survival from ages 80–84 to 90–94 and from ages 95–99 to 100–104. The survey provided evidence that early menopause (*i.e.*, under age 40) is associated with shorter longevity as well as that late menopause (*i.e.*, over age 50) is associated with greater longevity. Other researchers have previously suggested a relation between age at menopause and human longevity. Perls and Fretts (2001) also reported that women who reached menopause at the older childbearing ages tended to live longer than women who reached menopause at the younger ages. This relation can be explained by the fact that the female sex hormones (*i.e.*, estrogen and progesterone), produced prior to menopause and hence up through the older childbearing ages, are protective against the health risks of aging.

Reproductive Aging in Women

Reproductive aging refers to the process by which reproduction become less efficient and less successful as the woman gets older during the reproductive period. It starts to occur long before menopause, which, as mentioned, is preceded by a premenopausal period of a varying number of years. Reproductive aging for a population is indicated by the falling percentages of fecund women and the declining rates of natural fertility. In the later phases, it is also indicated by the various signs of the premenopause listed earlier, such as irregular and longer menstrual cycles, increased frequency of anovulation, more variable (with lower) estrogen levels, and elevated FSH (follicular stimulating hormone) and LH (luteinizing hormone). (See Ferrell et al. 2007; O'Connor et al. 1998, 2001)

There is also increased production of defective eggs. Even near the end of the menopause, conception is still possible, but with conception there is an increased risk of miscarriages and fetal losses, and chromosomal defects in any children born (Vandresse et al. 2008). According to Morris (2008), late maternal age is the only identified risk factor for Down's syndrome.

The indications of reproductive aging are now more common in the Western countries than in the LDC because of the dramatic rise in the age of parents in recent decades. The number of first-time mothers 35 years of age and over in the United States increased by 116% between 1970 and 1999 while the number of births to mothers under 30 years of age decreased about one quarter. Since the incidence of birth defects increases with a rise in the number of older mothers and the age of mothers has risen in the United States, an increase in the number of children with birth defects can be anticipated in the United States in the coming years, barring a fall in the age of motherhood or an increase in the abortions of affected fetuses.

Causes of reproductive aging. Why is there an increasing tendency for reproduction to be less successful as age increases over ages 20–24? To find the causes of reproductive aging in women, attention has been given by researchers to the aging of the hypothalamus and pituitary glands, the follicles and eggs in the ovaries, and the ovaries themselves (Wood et al. 2001). The follicles are the cavities in the ovaries containing the egg cells. Hypotheses about reproductive aging attribute it to good eggs going bad and to the depletion of follicles and eggs. With increasing age, inactive phases of follicular production tend to increase, menstrual cycles lengthen, and steroid production drops. By the late 30s and early 40s, 70–80% of the follicles and eggs have been depleted or have gone bad. Menopause occurs when less than 1,000 follicles are active out of the 300,000 or so that remain at puberty from the more than 1 million present at birth. Because egg cells cannot repair themselves, they are more subject to damage as the woman ages. As the production of follicles diminishes, there are periods of time with no follicular activity although occasionally a postmenopausal woman will have active follicles.

The BIMORA (Biodemographic Models of Reproductive Aging) research project now under way is designed to test whether follicular depletion accounts for reproductive aging (Ferrell et al. 2005). It is a 5-year longitudinal study of the menstrual cycles of 130 women, involving the collection and analysis of thousands of urine specimens for their hormonal content. The BIOMORA research study does not try to explain elevated FSH, decreased gonadotropins, the rise in fetal losses, or increase in anovulation.

Health consequences for women and offspring. One of the principal consequences of the monthly surges of estrogen during the course of the reproductive period is the increased risk of gynecological cancers (i.e., breast, endometrial, and ovarian cancers). Women who begin their menstrual cycles earlier or take longer to transition to menopause appear to be at higher risk of these cancers. Many studies show that women who become pregnant, especially at the earlier ages, are at reduced risk of breast cancer. Following menopause, when progesterone production has been eliminated and estrogen production has been vastly reduced, the rate of increase in these gynecological cancers slows down. Because the female reproductive

hormones, estrogen and progesterone, are responsible for causing these cancers, postmenopausal women as a group are less likely to get gynecological cancers than premenopausal women.

Female reproductive hormones increase the risk of breast cancer because they cause the cells lining the milk ducts in the breast to divide profusely at the end of each menstrual cycle, and DNA mutations (*i.e.*, errors in cell copying) are more likely to occur with the greater frequency of these cycles. Women who become pregnant no longer experience the repetitive division of these cells and hence are at reduced risk of DNA mutations that could cause breast cancer. Depending on the circumstances, several DNA mutations may be required to produce cancerous cells.

In the United States the recorded (age-adjusted) breast cancer rate has not changed substantially in the last several decades, although some decline has occurred since 1990. The very factors that have been extending the length of the female reproductive period–improved nutrition and reduced heavy labor – account in part for the rise in the incidence in breast-cancer or its failure to decline. As mentioned, early pregnancy reduces the risk of breast cancer and, in effect, delayed childbearing contributes to it. Increased public awareness of the risk of breast cancer has stimulated interest in more widespread testing of women, which may have restrained the rise in the rate but may also have contributed to the statistical uncertainty of the trend.

Evidently, late menopause and extended fecundity have both positive and negative consequences. They are associated with the greater longevity of the mothers and reduced levels of several chronic degenerative diseases, but they are also associated with greater risks of gynecological cancers and with a number of disorders in the offspring. I have already mentioned the increased risk, with advanced maternal age, of chromosomal defects and various genetic defects in the newborn, including Down's syndrome.

Postmenopausal women are at increased risk of a host of other conditions. Among them are cardiovascular diseases, including heart attacks and stroke, osteoporosis (loss of bone mass), osteoarthritis (loss of cartilage leading to joint damage), sarcopenia (loss of muscle mass), prolapse of abdominal organs, and various vaginal conditions.

Reproductive Span and Reproductive Aging in Men

The male reproductive period is quite different from the female reproductive period. It cannot be quantified very well. Puberty in males begins about a year or so after it does in females, but the timing of the termination of the male reproductive period cannot be specified except very broadly. The onset of male puberty ranges narrowly around age 14 in the developed countries. Puberty in males is hastened by the same factors as for women, namely, good nutrition and a sedentary life. Hence, as is true for females, it is likely that the onset of puberty for males occurred at a higher age a century ago in the industrialized countries than it does now, but there is no empirical evidence for such an inference. It is probable also that puberty among males in the less developed areas today occurs at about 15.

The male reproductive period stretches out over several decades beyond the age of the female menopause. Some boys become fathers at ages 10–14 each year and many men become fathers over age 50. Men have occasionally been reported to become fathers in their seventies and even eighties. Evolution has not directed its efforts so sharply at the timing of the male reproductive period as at the timing of the female reproductive period. Apparently such efforts are deemed unnecessary by evolution since control of female development assures control over human reproduction rather well.

Male fecundity and fertility patterns. It follows from the above analysis that the fecundity and fertility patterns of men are somewhat different from those of women. Little is known formally of the male fecundity curve. Fecundity of men does not drop off abruptly as it does for women, but falls off gradually. Yet, fecundity rates for men do substantially decline with age. The fecundity of men declines steadily from the late 30s to the 80s. Men 35–39 years old are about twice as likely to be infecund as men 20–24 years old. The male equivalent of menopause is characterized by declines in the production of testosterone and other male hormones and declines in the sperm count, the quality of the sperm, libido, and sexual performance (Fisch 2005). The speed of the fecundity decline is determined in general by a combination of genes, environment, lifestyle, and chance. The aging of a man's reproductive system depends in part on his health history and lifestyle (with a healthful diet, exercise, and avoidance of drugs, smoking, and excessive use of alcohol being favorable) as well as his genes. Poor general health may result in poor sexual health. Individuals vary greatly in their sexual health, and some young men have testosterone levels as low as those of much older men. Poor sexual health is reflected in problems of libido, sexual performance, and infecundity.

Male fertility rates start to become substantial in the early 1920s and remain so through ages 50–54. In the United States, men, like women, have been delaying the siring of children in the last few decades. The proportion of births to fathers 35 years of age and over increased from 14% in 1970 to 21% in 1999, and the number of first-time fathers 35 years and over increased by 50% (U.S. NCHS/Chandra et al. 2005).

Health consequences for offspring of late fatherhood. The trend of increased births at the later reproductive ages, combined with the biological decline in reproductive effectiveness with increasing age, contributes to an increased amount and chance of the occurrence of fertility problems among sexually active couples at the later ages. These problems include both infertility of the parents and severe genetic defects in the children. For fathers the risk of severe genetic defects is over three times as great at age 40 as at age 25 and nearly six times as great at age 45 as at age 25 (Fisch 2005).

Advanced paternal age is associated with a number of complex disorders in the children, including Down's syndrome, congenital heart defects, neural tube defects (*i.e.*, defects in the dorsal tubular structure in the embryo that differentiates into the brain and spinal cord), childhood brain cancers, mental retardation of unknown etiology, and, later in life, Alzheimer's disease and prostate cancer (Vandresse et al. 2008; Fisch 2005). Down's syndrome is especially likely to appear in children

of older fathers if both the father and mother are older than age 35 or 40 (when both sperm and ova are of poorer quality). Advanced age of the father also adds to the risk of miscarriage when the mother is older than 35 years of age. Accordingly, the ages of both partners are important in assessing the risk of having children with birth defects. Additional details on the disorders associated with advanced age of fathers were given in Chap. 6 in the section on genetic influences in mortality and morbidity.

Evolutionary Theory, Reproductive Success, and Longevity

Evolution establishes the time frame within which the biological processes of growth, development, and reproduction have to occur. Inasmuch as evolution is "committed to" the reproduction of individuals and the environmental hazards (*i.e.*, starvation, drought, disease, accidents, and predation) are considerable, reproductive success depends on assigning reproduction to an early period in the life of the woman. It is then that she is most likely to be fit for and successful in this mission. Accordingly, sexual maturation for humans occurs around 13–15 years of age and parents can become great grandparents by the age of 40. Carnes et al. (2003) maintain that this is the historical time frame that is necessary for achieving maturation and reproduction and that defines the duration of human life (the biological "warranty period"). Evolutionary theory assigns a major influence to the risks of extrinsic mortality (*i.e.*, principally, deaths from infectious diseases and violence) in molding the timing and biology of the reproduction of humans as well as the timing and biology of the reproduction of other species.

Evolution plays another role in reproduction. The disposable soma theory of aging, derived from evolutionary theory, maintains that all animals have limited energy resources and these limited energy resources are used either for (1) performance of general metabolic activities, (2) reproduction, including the birth of offspring and rearing them to reproductive age, and (3) maintenance of the developing and adult organism (Holliday 2004,1995). The resources not used for (1) must be divided between (2) and (3), that is, the more that is invested in reproduction, the less is available for maintenance, and vice versa. This means that there is a trade-off between reproduction and longevity. This theory clearly applies in reference to interspecies comparisons. Humans, like other higher primates, survive a long period, produce few offspring, have them at widely spaced intervals, and develop slowly. Those animals, such as rodents, that have high mortality rates have high rates of reproduction and develop into adults rapidly. Holliday maintains that the forces of natural selection can result in either a reduction of longevity or an increase in longevity, and that a reduction in mortality in a natural environment will lead to a selection for longer-lived species and a lower rate of reproduction. In short, there is a clear inverse relationship between these two life course parameters – life span and fecundity.[5]

[5]In a later chapter I discuss the research that links caloric restriction to an increase in the life span of various subhuman species, including mice and rats. Caloric restriction reduces fertility in these

I had earlier pointed out evidence of the association of the greater longevity of women who have a late menopause and an extended reproductive period. I have also remarked that there is an inverse relation between longevity and reproductive success demonstrated by intraspecies comparisons. These two findings would appear to need to be reconciled.

Maternal and Perinatal Health

Maternal Mortality

Definitions

According to the *International Classification of Diseases, Tenth Revision*, maternal death is the death of a woman while pregnant, or within 42 days of the termination of pregnancy, irrespective of the duration and the site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes (ICD-10, Class XV, codes O00-O99). According to the Ninth Revision of the International List, maternal deaths are the sum of deaths due to abortion (code AM 42), direct obstetric causes (code AM 43), and indirect obstetric causes (code AM 44). Direct obstetric deaths result from obstetric complications of the pregnant state (encompassing pregnancy, labor, and the puerperium), *i.e.*, from interventions, omissions, incorrect treatment, or from a chain of events resulting from any of these actions. Indirect obstetric deaths are those resulting from previously existing disease or disease that develops during pregnancy; they are not due to direct obstetric causes but are aggravated by the physiological effect of pregnancy.

WHO provides definitions of two other concepts relating to maternal deaths, namely, late maternal deaths and pregnancy-related deaths. A late maternal death is a death more than 42 days and less than 1 year after termination of pregnancy from any cause related to or aggravated by the pregnancy. A pregnancy-related death is a death of a woman while pregnant or within 42 days of the termination of pregnancy irrespective of the cause of death.

Measures: Maternal Mortality Ratio

Maternal mortality is conventionally measured by the maternal mortality ratio, representing the risk of a woman's dying from complications of pregnancy, childbirth,

---

rodents and it appears that the animals' response to this shortage of energy is to channel more of it into their own maintenance and less into reproduction (Holliday 2004).

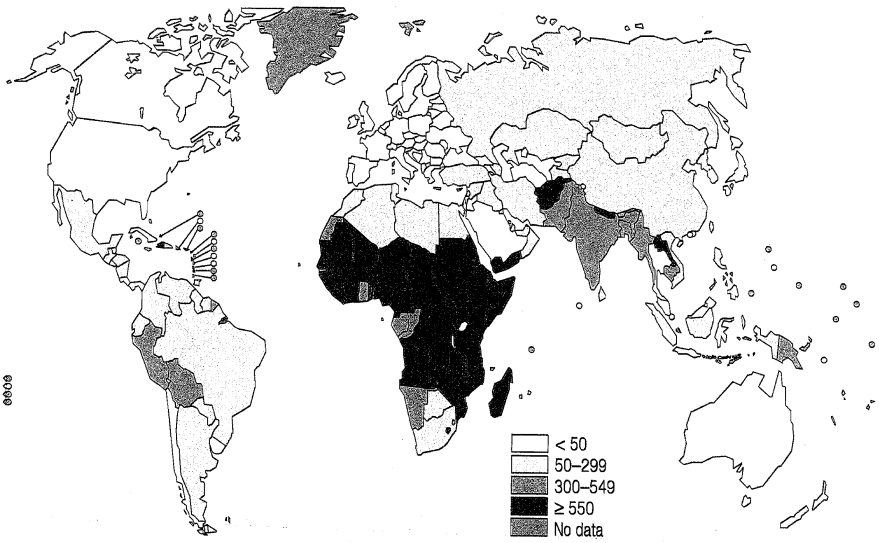


Fig. 9.1 World map of the maternal mortality ratio: 2000 (Maternal mortality ratio is the number of maternal deaths in a year per 100,000 births during the year; Source: World Health Organization 2005, Figure 1.5, page 15. Reprinted with permission of the World Health Organization)

or the puerperium (*i.e.*, during childbirth or immediately thereafter). The maternal mortality ratio is defined as the number of deaths due to puerperal causes per 100,000 births. The formula is:

$$\mathrm{MMR}_1 = (\mathrm{D_p} \div \mathrm{B}) * 100{,}000 \tag{9.1a}$$

where D_p represents deaths due to puerperal causes. Formerly, when maternal deaths were more numerous, a constant multiplier of 10,000 was commonly used, and it still may be reasonably used for countries with very high maternal mortality.

In the United States the maternal mortality ratio for 2004 was 13.1 deaths per 100,000 live births. As defined above, the number of maternal deaths does not include all deaths occurring to pregnant women but only those deaths reported on the death certificate that were assigned to causes related to or aggravated by pregnancy (ICD Codes A34, O00-O95, and O98-O99). Following the international definition, the number excludes deaths occurring more than 42 days after the termination of pregnancy and deaths of pregnant women due to external causes (e.g., injuries).

The maternal mortality ratio varies widely among the regions and countries of the world (Fig. 9.1). WHO, UNICEF, UNFPA (2003) jointly developed country-by-country estimates of maternal mortality ratios for 2000. The global estimate was 400 deaths per 100,000 (live) births (Table 9.3). The highest rates occurred in the less developed regions; the overall figure for these areas was 440 deaths per 100,000 (live) births. In the more developed regions the overall figure was 20 deaths per

Table 9.3 Maternal mortality and lifetime risk of maternal death for world regions: 2000

Region	Maternal mortality ratio[a] (1)	Number of maternal deaths (2)	Lifetime risk of maternal death, 1 in[b]: (3)
World total	400	529,000	74
More developed regions	20	2,500	2,800
United States[c]	13	540	3,300
Europe	24	1,700	2,400
Less developed regions	440	527,000	61
Africa	830	251,000	20
Northern Africa	130	4,600	210
Sub-Saharan Africa	920	247,000	16
Asia	330	253,000	94
Eastern Asia	55	11,000	840
South Central Asia	520	207,000	46
Southeastern Asia	210	25,000	140
Western Asia	190	9,800	120
Latin America and the Caribbean	190	22,000	160
Oceania	240	530	83

Source: WHO, UNICEF, UNFPA 2003

[a]Maternal deaths per 100,000 live births

[b]$1 \div [1 - (1 - MMR_1)^{1.2*TFR}]$, where MMR1 is the maternal mortality ratio per birth and TFR is the total fertility rate per woman

[c]For the year 2004. Computed by author

100,000 (live) births. In terms of specific world regions, maternal deaths ranged from 24 per 100,000 births for Europe to 920 per 100,000 births in sub-Saharan Africa. The ratios for Latvia and Romania were quite high at 41–42, but in Italy and Germany they were only 1.3 and 4.8, respectively. In Asia the ratios for Azerbaijan and Kyrgyzstan were in the 40s, and Eastern Asia (mainly China) had a ratio of 55. The high maternal mortality ratios in Eastern Europe and Asia apparently result from the high percentages of women not using contraception or using traditional methods, and the resort to abortion to avoid unwanted or ill-timed births.

Measures: Maternal Mortality Rate

In the formula for calculating the maternal mortality ratio, the number of births is employed as an approximation to the number of women exposed to the risk of dying from puerperal causes. A refinement of the maternal mortality ratio, the maternal mortality rate, broadens the denominator to include late fetal losses and induced terminations of pregnancy. Late fetal losses are fetal losses of 28 weeks

of uterogestation or more and induced terminations of pregnancy refer to (induced) abortions:

$$\mathrm{MMR_2} = \mathrm{D_p(B + L^{lf} + A)^*100{,}000} \tag{9.1b}$$

where L^{lf} represents late fetal losses and A (induced) abortions. The maternal mortality rate is just a little lower than the maternal mortality ratio.

Lifetime Risk of Maternal Death

Another measure, which reflects both the risk of death per pregnancy and the number of pregnancies, is the lifetime risk of maternal death. It is calculated by the formula,

$$\mathrm{LRMD} = 1 - (1 - \mathrm{MMR_1})^{1.2\mathrm{TFR}} \tag{9.2}$$

where MMR_1 is the maternal mortality ratio (expressed as a decimal), TFR is the total fertility rate (expressed per woman), and 1.2 is a multiplier to adjust for pregnancies not ending in live births (Tinker and Koblinsky 1993). As shown in Table 9.3, in 2000 the lifetime risk of maternal death for women in the More Developed Countries (MDC) was 1 in 2,800 whereas the risk for women in the Less Developed Countries (LDC) was 1 in 61. The regional extremes were 1 in 16 for sub-Saharan Africa and 1 in 2400 for Europe. The figure for North America (not shown) was presumably lower since the United States figure for 2004 alone was only 1 in 3,300.

Pregnancy Losses

Definition of Concepts

Much loss of potential human life occurs as a result of losses during pregnancy. As a group such losses have been called pregnancy wastage, pregnancy losses, fetal deaths, and fetal losses. There is a close relation between late fetal mortality and neonatal mortality (*i.e.*, early infant mortality), particularly with regard to causal pattern. Hence, I treat late fetal mortality, neonatal (*i.e.*, early infant) mortality, and perinatal mortality (*i.e.*, the combination of late fetal losses and early infant mortality) together in this section. According to the World Health Organization (U.S. NOVS 1954), neonatal deaths are deaths among live births during the first 28 completed days of life. Those neonatal deaths occurring during the first seven completed days of life are designated early neonatal deaths, and deaths occurring during the remainder of the period (*i.e.*, from the 8th day through the 28th day) are designated late neonatal deaths.

The WHO-recommended definition of fetal death complements WHO's definition of (live) birth and death. In 1950 WHO recommended the following definition of fetal death for international use (U.S. NOVS 1954):

> Death prior to the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of pregnancy; the death is indicated by the fact that after such separation, the fetus does not breathe or show any other evidence of life such as beating of the heart, pulsation of the umbilical cord, or definite movement of involuntary muscles.

WHO included this definition in the Ninth Revision of the International List and maintained it in the Tenth Revision of the International List in 1992.

In some countries, however, the definition of fetal death employed differs from the international recommendation: Live-born babies dying early in life (*e.g.*, before registration of the birth or within 24 h of birth) may be categorized as fetal deaths. A more important problem is the incompleteness and irregularity of reporting fetal deaths. These limitations apply to the data for most countries of the world. Reporting of early fetal deaths may be seriously incomplete even where their registration is required by law. The period of uterogestation required for registration varies widely. When registration of only late fetal deaths (rather than all fetal deaths) is mandatory, countries differ as to what has to be registered as a late fetal death, although 28 weeks of gestation is most frequently specified as the minimum period. On balance, international comparability is far greater for late fetal deaths than for all fetal deaths taken together (Casterline 1989).

WHO has recommended the term fetal deaths as the generic term for all pregnancy losses. Accordingly, fetal deaths encompass those pregnancy losses known by other names, such as miscarriages, abortions, and stillbirths. If there is need for or interest in a general term for pregnancy losses for statistical purposes, I prefer the term pregnancy losses itself rather than fetal deaths. A more restricted term is needed for pregnancy losses excluding induced terminations of pregnancy, that is, a term for spontaneous terminations of pregnancy. For this event I suggest the term fetal losses. The U.S. National Center for Health Statistics occasionally uses this term in this way. The term abortion may then be used in its restricted popular sense to represent induced terminations of pregnancy. The main reason for replacing the term fetal deaths as the general term for pregnancy losses or fetal losses by the latter terms is that, in the U.N. statistical system, death can occur only after a live birth.

The WHO also recommends that fetal deaths should be distinguished by length of gestation period as early (less than 20 completed weeks of gestation), intermediate (20 completed weeks but less than 28), and late (28 completed weeks and over). This leads to the following classification of types of pregnancy losses:

Pregnancy losses
Fetal losses (spontaneous terminations of pregnancy)

- Early fetal losses (less than 20 weeks)
- Intermediate fetal losses (20–27 completed weeks)
- Late fetal losses (28 weeks or more)
- Fetal losses of unstated duration

Abortions (induced terminations of pregnancy)

The determination of fetal age is often difficult and comparability of the tabulations is affected by the differences in the skill of the medical attendants making this determination. Even the data on late fetal losses are subject to substantial error introduced by incorrect reporting of gestational age. In view of the very poor reporting of early fetal losses and the variation from one reporting jurisdiction to another of the minimum period for which reporting of fetal losses is required, fetal loss rates in official sources are normally limited to fetal losses with 20 weeks or more, or 28 weeks or more, of uterogestation. Where the gestation period is not stated, as is often the case, the World Health Organization classifies these unstated cases in a fourth class.

In the United States, reporting is required for fetal losses of 20 weeks or more in all reporting areas (with some variation by birth weight), and the National Center for Health Statistics assigns cases of unstated gestational age to the period 20 weeks or more (U.S. NCHS 1996). In 1939 a separate *Standard Certificate of Stillbirth* was introduced in the United States. The latest *U.S. Standard Report of Fetal Death*, as the document is now called, is shown as Exhibit 9.A1.

Fetal Losses

Fetal losses may be measured by the fetal loss ratio or the fetal loss rate. The fetal loss ratio is defined as the number of fetal losses reported in a year per 1,000 live births in that year, or

$$\mathrm{FLR} = \mathrm{L}^{\mathrm{lf}} \div \mathrm{B} * 1{,}000 \tag{9.3a}$$

where L^{lf} represents late fetal losses, or both intermediate and late fetal losses. For the United States in 2004, the calculation of the fetal loss ratio is,

$$25{,}655 \div 4{,}112{,}055 * 1{,}000 = 6.24, \text{for fetal losses of 20 weeks or more}$$

The component ratios are:

$$12{,}895 \div 4{,}112{,}055 * 1{,}000 = 3.14, \text{for 20–27 weeks}$$

$$12{,}760 \div 4{,}112{,}055 * 1{,}000 = 3.10, \text{for 28 weeks or more}$$

Because of the variability of coverage of early and intermediate fetal losses mentioned earlier, it is preferable from the point of international comparability to compute the fetal loss ratio on the basis of late fetal losses only.

The same data required to compute the fetal loss ratio can be used to compute the theoretically more precise fetal loss rate. The fetal loss rate relates fetal losses more closely to the population at risk than the fetal loss ratio. The formula includes the fetal losses in the denominator as well as in the numerator:

$$\mathrm{FLR}^{\mathrm{p}} = \mathrm{L}^{\mathrm{lf}} \div (\mathrm{B} + \mathrm{L}^{\mathrm{lf}}) * 1{,}000 \tag{9.3b}$$

Exhibit 9.A1.--**U.S. Standard Report of Fetal Death: 2003 Revision**

LOCAL FILE NO. STATE FILE NUMBER:

MOTHER

1. NAME OF FETUS (optional-at the discretion of the parents) | 2. TIME OF DELIVERY (24hr) | 3. SEX (M/F/Unk) | 4. DATE OF DELIVERY (Mo/Day/Yr)

5a. CITY, TOWN, OR LOCATION OF DELIVERY | 7. PLACE WHERE DELIVERY OCCURRED (Check one) ☐ Hospital ☐ Freestanding birthing center ☐ Home Delivery: Planned to deliver at home? ☐ Yes ☐ No ☐ Clinic/Doctor's office ☐ Other (Specify)_____ | 8. FACILITY NAME (If not institution, give street and number)

5b. ZIP CODE OF DELIVERY

6. COUNTY OF DELIVERY | 9. FACILITY ID. (NPI)

10a. MOTHER'S CURRENT LEGAL NAME (First, Middle, Last, Suffix) | 10b. DATE OF BIRTH (Mo/Day/Yr)

10c. MOTHER'S NAME PRIOR TO FIRST MARRIAGE (First, Middle, Last, Suffix) | 10d. BIRTHPLACE (State, Territory, or Foreign Country)

11a. RESIDENCE OF MOTHER-STATE | 11b. COUNTY | 11c. CITY, TOWN, OR LOCATION

11d. STREET AND NUMBER | 11e. APT. NO. | 11f. ZIP CODE | 11g. INSIDE CITY LIMITS? ☐ Yes ☐ No

FATHER

12a. FATHER'S CURRENT LEGAL NAME (First, Middle, Last, Suffix) | 12b. DATE OF BIRTH (Mo/Day/Yr) | 12c. BIRTHPLACE (State, Territory, or Foreign Country)

DISPOSITION

13. METHOD OF DISPOSITION: ☐ Burial ☐ Cremation ☐ Hospital Disposition ☐ Donation ☐ Removal from State ☐ Other (Specify)_____

ATTENDANT AND REGISTRATION INFORMATION

14. ATTENDANT'S NAME, TITLE, AND NPI
NAME: _____
NPI:_____
TITLE: ☐ MD ☐ DO ☐ CNM/CM ☐ OTHER MIDWIFE ☐ OTHER (Specify)_____

15. NAME AND TITLE OF PERSON COMPLETING REPORT
Name _____
Title _____

16. DATE REPORT COMPLETED ____/____/____ MM DD YYYY

17. DATE RECEIVED BY REGISTRAR ____/____/____ MM DD YYYY

18. CAUSE/CONDITIONS CONTRIBUTING TO FETAL DEATH

CAUSE OF FETAL DEATH

18a. INITIATING CAUSE/CONDITION
(AMONG THE CHOICES BELOW, PLEASE SELECT THE ONE WHICH MOST LIKELY BEGAN THE SEQUENCE OF EVENTS RESULTING IN THE DEATH OF THE FETUS)
Maternal Conditions/Diseases (Specify) _____
Complications of Placenta, Cord, or Membranes
☐ Rupture of membranes prior to onset of labor
☐ Abruptio placenta
☐ Placental insufficiency
☐ Prolapsed cord
☐ Chorioamnionitis
☐ Other Specify)_____
Other Obstetrical or Pregnancy Complications (Specify) _____
Fetal Anomaly (Specify) _____
Fetal Injury (Specify) _____
Fetal Infection (Specify) _____
Other Fetal Conditions/Disorders (Specify) _____
☐ Unknown

18b. OTHER SIGNIFICANT CAUSES OR CONDITIONS
(SELECT OR SPECIFY ALL OTHER CONDITIONS CONTRIBUTING TO DEATH IN ITEM 18b)
Maternal Conditions/Diseases (Specify) _____
Complications of Placenta, Cord, or Membranes
☐ Rupture of membranes prior to onset of labor
☐ Abruptio placenta
☐ Placental insufficiency
☐ Prolapsed cord
☐ Chorioamnionitis
☐ Other Specify)_____
Other Obstetrical or Pregnancy Complications (Specify) _____
Fetal Anomaly (Specify) _____
Fetal Injury (Specify) _____
Fetal Infection (Specify) _____
Other Fetal Conditions/Disorders (Specify) _____
☐ Unknown

Mother's Name _____ Mother's Medical Record No. _____

18c. WEIGHT OF FETUS (grams preferred, specify unit)
_____
☐ grams ☐ lb/oz

18e. ESTIMATED TIME OF FETAL DEATH
☐ Dead at time of first assessment, no labor ongoing
☐ Dead at time of first assessment, labor ongoing
☐ Died during labor, after first assessment
☐ Unknown time of fetal death

18f. WAS AN AUTOPSY PERFORMED? ☐ Yes ☐ No ☐ Planned

18g. WAS A HISTOLOGICAL PLACENTAL EXAMINATION PERFORMED? ☐ Yes ☐ No ☐ Planned

18d. OBSTETRIC ESTIMATE OF GESTATION AT DELIVERY
_____ (completed weeks)

18h. WERE AUTOPSY OR HISTOLOGICAL PLACENTAL EXAMINATION RESULTS USED IN DETERMINING THE CAUSE OF FETAL DEATH? ☐ Yes ☐ No

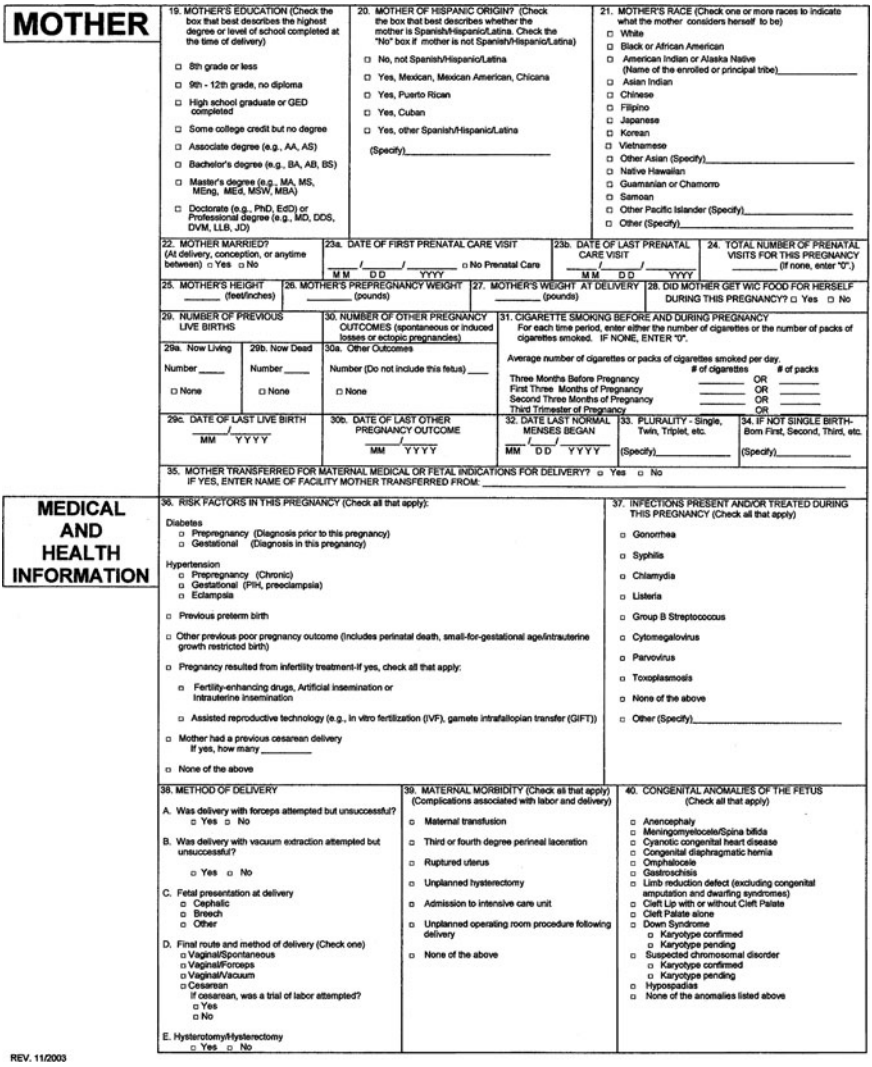
REV. 11/2003

Exhibit 9.A1 U.S. Standard report of fetal death: 2003 revision (Source: U.S. National Center for health statistics, www.cdc.gov/nchs)

For U.S. in 2004, $25,655 \div (4,112,055 + 25,655) = 6.20$, for 20 weeks or more

$12,895 \div (4,112,055 + 12,895) = 3.13$, for 20–27 weeks

$12,760 \div (4,112,055 + 12,760) = 3.09$, for 28 weeks or more



MOTHER

19. MOTHER'S EDUCATION (Check the box that best describes the highest degree or level of school completed at the time of delivery)
- ☐ 8th grade or less
- ☐ 9th - 12th grade, no diploma
- ☐ High school graduate or GED completed
- ☐ Some college credit but no degree
- ☐ Associate degree (e.g., AA, AS)
- ☐ Bachelor's degree (e.g., BA, AB, BS)
- ☐ Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)
- ☐ Doctorate (e.g., PhD, EdD) or Professional degree (e.g., MD, DDS, DVM, LLB, JD)

20. MOTHER OF HISPANIC ORIGIN? (Check the box that best describes whether the mother is Spanish/Hispanic/Latina. Check the "No" box if mother is not Spanish/Hispanic/Latina)
- ☐ No, not Spanish/Hispanic/Latina
- ☐ Yes, Mexican, Mexican American, Chicana
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, other Spanish/Hispanic/Latina

(Specify)_____

21. MOTHER'S RACE (Check one or more races to indicate what the mother considers herself to be)
- ☐ White
- ☐ Black or African American
- ☐ American Indian or Alaska Native (Name of the enrolled or principal tribe)_____
- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Other Asian (Specify)_____
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander (Specify)_____
- ☐ Other (Specify)_____

22. MOTHER MARRIED? (At delivery, conception, or anytime between) ☐ Yes ☐ No

23a. DATE OF FIRST PRENATAL CARE VISIT ____/____/____ (MM DD YYYY) ☐ No Prenatal Care

23b. DATE OF LAST PRENATAL CARE VISIT ____/____/____ (MM DD YYYY)

24. TOTAL NUMBER OF PRENATAL VISITS FOR THIS PREGNANCY _____ (If none, enter "0".)

25. MOTHER'S HEIGHT _____ (feet/inches)

26. MOTHER'S PREPREGNANCY WEIGHT _____ (pounds)

27. MOTHER'S WEIGHT AT DELIVERY _____ (pounds)

28. DID MOTHER GET WIC FOOD FOR HERSELF DURING THIS PREGNANCY? ☐ Yes ☐ No

29. NUMBER OF PREVIOUS LIVE BIRTHS

29a. Now Living Number _____ ☐ None

29b. Now Dead Number _____ ☐ None

30. NUMBER OF OTHER PREGNANCY OUTCOMES (spontaneous or induced losses or ectopic pregnancies)

30a. Other Outcomes Number (Do not include this fetus) ____ ☐ None

31. CIGARETTE SMOKING BEFORE AND DURING PREGNANCY
For each time period, enter either the number of cigarettes or the number of packs of cigarettes smoked. IF NONE, ENTER "0".

Average number of cigarettes or packs of cigarettes smoked per day.

	# of cigarettes		# of packs
Three Months Before Pregnancy	_____	OR	_____
First Three Months of Pregnancy	_____	OR	_____
Second Three Months of Pregnancy	_____	OR	_____
Third Trimester of Pregnancy	_____	OR	_____

29c. DATE OF LAST LIVE BIRTH ____/____ (MM YYYY)

30b. DATE OF LAST OTHER PREGNANCY OUTCOME ____/____ (MM YYYY)

32. DATE LAST NORMAL MENSES BEGAN ____/____/____ (MM DD YYYY)

33. PLURALITY - Single, Twin, Triplet, etc. (Specify)_____

34. IF NOT SINGLE BIRTH- Born First, Second, Third, etc. (Specify)_____

35. MOTHER TRANSFERRED FOR MATERNAL MEDICAL OR FETAL INDICATIONS FOR DELIVERY? ☐ Yes ☐ No
IF YES, ENTER NAME OF FACILITY MOTHER TRANSFERRED FROM: _____

MEDICAL AND HEALTH INFORMATION

36. RISK FACTORS IN THIS PREGNANCY (Check all that apply):

Diabetes
- ☐ Prepregnancy (Diagnosis prior to this pregnancy)
- ☐ Gestational (Diagnosis in this pregnancy)

Hypertension
- ☐ Prepregnancy (Chronic)
- ☐ Gestational (PIH, preeclampsia)
- ☐ Eclampsia
- ☐ Previous preterm birth
- ☐ Other previous poor pregnancy outcome (includes perinatal death, small-for-gestational age/intrauterine growth restricted birth)
- ☐ Pregnancy resulted from infertility treatment-If yes, check all that apply:
  - ☐ Fertility-enhancing drugs, Artificial insemination or Intrauterine insemination
  - ☐ Assisted reproductive technology (e.g., In vitro fertilization (IVF), gamete intrafallopian transfer (GIFT))
- ☐ Mother had a previous cesarean delivery If yes, how many _____
- ☐ None of the above

37. INFECTIONS PRESENT AND/OR TREATED DURING THIS PREGNANCY (Check all that apply)
- ☐ Gonorrhea
- ☐ Syphilis
- ☐ Chlamydia
- ☐ Listeria
- ☐ Group B Streptococcus
- ☐ Cytomegalovirus
- ☐ Parvovirus
- ☐ Toxoplasmosis
- ☐ None of the above
- ☐ Other (Specify)_____

38. METHOD OF DELIVERY

A. Was delivery with forceps attempted but unsuccessful? ☐ Yes ☐ No

B. Was delivery with vacuum extraction attempted but unsuccessful? ☐ Yes ☐ No

C. Fetal presentation at delivery
- ☐ Cephalic
- ☐ Breech
- ☐ Other

D. Final route and method of delivery (Check one)
- ☐ Vaginal/Spontaneous
- ☐ Vaginal/Forceps
- ☐ Vaginal/Vacuum
- ☐ Cesarean
  - If cesarean, was a trial of labor attempted?
  - ☐ Yes
  - ☐ No

E. Hysterotomy/Hysterectomy ☐ Yes ☐ No

39. MATERNAL MORBIDITY (Check all that apply) (Complications associated with labor and delivery)
- ☐ Maternal transfusion
- ☐ Third or fourth degree perineal laceration
- ☐ Ruptured uterus
- ☐ Unplanned hysterectomy
- ☐ Admission to intensive care unit
- ☐ Unplanned operating room procedure following delivery
- ☐ None of the above

40. CONGENITAL ANOMALIES OF THE FETUS (Check all that apply)
- ☐ Anencephaly
- ☐ Meningomyelocele/Spina bifida
- ☐ Cyanotic congenital heart disease
- ☐ Congenital diaphragmatic hernia
- ☐ Omphalocele
- ☐ Gastroschisis
- ☐ Limb reduction defect (excluding congenital amputation and dwarfing syndromes)
- ☐ Cleft Lip with or without Cleft Palate
- ☐ Cleft Palate alone
- ☐ Down Syndrome
  - ☐ Karyotype confirmed
  - ☐ Karyotype pending
- ☐ Suspected chromosomal disorder
  - ☐ Karyotype confirmed
  - ☐ Karyotype pending
- ☐ Hypospadias
- ☐ None of the anomalies listed above

REV. 11/2003

Exhibit 9.A1 (continued)

In international comparisons, this measure also should best be calculated with late fetal losses only. That is, the denominator should represent pregnancies carried to term or nearly carried to term. Late fetal loss ratios and late fetal loss rates for several countries are presented in Table 9.4.[6]

[6]In spite of the theoretical advantage of the fetal loss rate, the fetal loss ratio may still be considered preferable for international comparisons. The registration of fetal losses is irregular and the effect of this irregularity is compounded when fetal losses are included with the births in the base of

Table 9.4 Late fetal loss ratios and late fetal loss rates, for selected countries: 2007

			Ratio	Rate
	Late fetal losses	Births	[(1) ÷ (2)] ∗1, 000 =	(1) ÷ [(2) + (1)] ∗1, 000 =
Country	(1)	(2)	(3)	(4)
Canada[a]	1,012	342,176	2.96	2.95
Costa Rica	345	73,144	4.72	4.69
Cuba	1,017	112,472	9.04	8.96
Chile[b]	2,124	231.383	9.18	9.08
United States[a]	12.567	4,138,349	3.04	3.03
Italy	1,570	563,933	2.78	2.78
Japan	2,254	1,089,818	2.07	2.06
Kazakhstan	2,112	321,963	6.56	6.52
Australia[a]	757	259,177	2.92	2.91
Romania	1,009	214,728	4.70	4.68
Russian Federation	8,612	1,610,122	5.35	5.32

Source: Based on United Nations, Demographic Yearbook, 2007, Tables 9 and 12
[a]For 2005
[b]For 2006

Specific fetal loss ratios and fetal loss rates may be calculated in terms of the period of gestation of the fetus or in terms of the age of mother (requiring data on the date of termination of the pregnancy for fetal losses and on the age of mother for births, respectively). Other characteristics of importance in the analysis of fetal losses are the marital status of the mother (distinguishing marital from nonmarital pregnancies), sex of the fetus, total pregnancy order (counting live births plus fetal losses and induced abortions), weight of the fetus, and plurality of fetus (1, 2, 3+). Other factors include the cause of the fetal loss, hospitalization, age of father, date of marriage (for marital pregnancies), level of education of parents, and the occupational characteristics of the parents.

Induced Terminations of Pregnancy

Technically the term abortion encompasses both spontaneous and induced terminations of pregnancy but, for statistical purposes, it refers only to those abortions that are induced. Induced abortions may be legal or illegal as determined by the laws of each country (or state/province in some countries). A copy of the model

---

the fetal loss rate. Because of the likelihood that poor registration of fetal losses will occur in association with poor registration of births and, hence, that the errors in each component will offset one another to some extent, fetal loss ratios may sometimes be of satisfactory quality even where the basic data are questionable.

TYPE/PRINT IN PERMANENT BLACK INK FOR INSTRUCTIONS SEE HANDBOOK

U.S. STANDARD

REPORT OF INDUCED TERMINATION OF PREGNANCY

STATE FILE NUMBER

1. FACILITY NAME (If not clinic or hospital, give address)	2. CITY, TOWN, OR LOCATION OF PREGNANCY TERMINATION	3. COUNTY OF PREGNANCY TERMINATION
Merrywood Clinic	Louisville	Jefferson

4. PATIENT'S IDENTIFICATION	5. AGE LAST BIRTHDAY	6. MARRIED?	7. DATE OF PREGNANCY TERMINATION (Month, Day, Year)
25466	23	☐ YES ☒ NO	November 20, 1997

8a. RESIDENCE-STATE	8b. COUNTY	8c. CITY, TOWN, OR LOCATION	8d. INSIDE CITY LIMITS? (Yes or No)	8e. ZIP CODE
Ohio	Hamilton	Cincinnati	☒ YES ☐ NO	45202

9. OF HISPANIC ORIGIN? (Specify No or Yes – If yes, specify Cuban, Mexican, Puerto Rican, etc.)	10. RACE	11. EDUCATION (Specify only highest grade completed)	
☐ No ☒ Yes Specify: Puerto Rican	☐ American Indian ☐ Black ☒ White ☐ Other (Specify) _____	Elementary/Secondary (0-12)	College (1-4 or 5+)
		12	

12. DATE LAST NORMAL MENSES BEGAN (Month, Day, Year)	13. CLINICAL ESTIMATE OF GESTATION (Weeks)	14. PREVIOUS PREGNANCIES (Complete each section)			
		LIVE BIRTHS		OTHER TERMINATIONS	
		14a. Now Living	14b. Now Dead	14c. Spontaneous	14d. Induced (Do not include this termination)
September 5, 1997	10 weeks	Number _____ ☒ None	Number _____ ☒ None	Number _____ ☒ None	Number _____ ☒ None

16. TYPE OF TERMINATION PROCEDURE (Check only one)

☒ Suction Curettage

☐ Medical (Nonsurgical), Specify Medication(s) _____

☐ Dilation and Evacuation (D&E)

☐ Intra-Uterine Instillation (Saline or Prostaglandin)

☐ Sharp Curettage (D&C)

☐ Hysterotomy/Hysterectomy

☐ Other (Specify) _____

16. NAME OF ATTENDING PHYSICIAN (Type/Print)	17. NAME OF PERSON COMPLETING REPORT (Type/Print)
Edmund Matthew Stone, M.D.	Julia Lynn Koval

7-97/R

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES — CENTERS FOR DISEASE CONTROL AND PREVENTION — NATIONAL CENTER FOR HEALTH STATISTICS — 1997 REVISION

Exhibit 9.A2 Standard report of induced terminations of pregnancy: 1997 revision (Source: U.S. National Center for Health Statistics, www.cdc.gov/nchs)

U.S. Standard Report of Induced Terminations of Pregnancy is shown here as Exhibit 9.A2. It is a short form that contains some of the same questions as the much longer *U.S. Standard Report of Fetal Death*.

Definitions, sources, and limitations. Spontaneous abortions, or spontaneous terminations of pregnancy, correspond roughly to what in popular usage is designated as miscarriages and stillbirths. Miscarriages usually refer to spontaneous terminations of pregnancies of short duration and stillbirths to those with longer periods of gestation, but the dividing line varies from country to country (Loghi et al. 2008). In the United States, the term stillbirth has a legal meaning that depends on the period of uterogestation and that varies according to political jurisdiction. Spontaneous abortions are far more common than induced abortions (James 1970). It has been

Table 9.5 Global and regional estimates of induced abortions: 1995 and 2003 (Includes both legal and illegal abortions)

Region	No. of abortions (millions)		Abortion rate[a]	
	1995	2003	1995	2003
World	45.6	41.6	35	29
More developed countries	10.0	6.6	39	26
Excluding Eastern Europe	3.8	3.5	20	19
Less developed countries[b]	35.5	35.0	34	29
Excluding China	24.9	26.4	33	30
Estimates by region				
Africa	5.0	5.6	33	29
Asia	26.8	25.9	33	29
Europe	7.7	4.3	48	28
Latin America	4.2	4.1	37	31
Northern America	1.5	1.5	22	21
Oceania	0.1	0.1	21	17

Source: Alan Guttmacher Institute, (2009). Facts on induced abortions worldwilde. In Brief, New York: Guttmacher Institute. <http://www.guttmacher.org/pubs/fb-IAW.pdf>, accessed June 5, 2008. Reprinted with the permission of the Alan Guttmacher Institute. Data complied by the World Health Organization and the Alan Guttmacher Institute.

[a] Abortions per 1000 women 15–44 years of age

[b] Abortions within Africa, the Americas (excluding Canada and the United States), Asia (excluding Japan), and Oceania (excluding Australia and New Zealand)

estimated that between 10% and 40% of pregnancies result in spontaneous abortions in the first trimester. The percentage drops sharply in the second trimester, when only about 1% or 2% of pregnancies abort spontaneously.

As stated, in statistical usage, the term abortion is limited to induced abortions. It may be defined more precisely as induced terminations of pregnancy before the fetus has became capable of surviving to the neonatal period and maintaining an independent life outside the uterus of the mother. Formerly the term abortion was not used in international statistical compilations because it was believed to have legal, programmatic, and ethical implications that are not desirable in a statistical concept. Currently, however, worldwide estimates of (induced) abortions are being made on the basis of (1) official statistics or other national data on legal abortions in about 60 countries (available in the public policy databank maintained by the Population Division, United Nations); (2) population surveys for a few countries without official statistics; (3) special studies for about 10 countries where abortion is highly restricted; and (4) estimates of illegal abortions from abortion-related hospitalization data and other sources developed by the World Health Organization and the Guttmacher Institute. Summary estimates on abortions, both legal and illegal, for the world have been generated by the Guttmacher Institute in collaboration with the World Health Organization for both 1995 and 2003 (Table 9.5).

Those countries where (induced) abortion is legal generally require that all abortions be reported to health authorities; these countries provide official information on abortions. Such data may be compiled by state governments in some countries as well, as in the United States. In 1970 the U.S. National Center for Health Statistics began to separate reports of abortions, received from and compiled by the states, from other fetal losses, in order to be able to provide separate data on (induced) abortions and to improve the comparability of data on both fetal losses and abortions. Reporting of legal abortions is commonly incomplete. Its completeness varies according to several factors, including the narrowness or breadth of the grounds for obtaining a legal abortion, the existence of strong independent health jurisdictions, the availability of medical health insurance coverage, the degree to which agency health reports are consolidated within a government, and the extent to which abortions are carried out in private facilities outside hospitals or public facilities (United Nations 2003).

Abortion is illegal in most parts of Africa, Latin America, and Asia, except China. Stringent legal restrictions do not assure a low abortion rate. Where abortion is illegal, the frequency of abortions has to be ascertained indirectly. The data may come from hospital records of women admitted for complications of abortions, from survey responses to questions on women's pregnancy histories, or even from surveys of abortion-service providers. Estimates based on such sources are rough approximations at best.

Measures of legal abortions. Several measures of abortions (*i.e.*, legal induced terminations of pregnancy) are in common use. Defined in this narrow sense, they parallel those for fetal losses. The general abortion rate is defined as the number of (legal) induced pregnancy terminations in a calendar year (A) per 1,000 women 15–44 years of age (or 15–49 years) at midyear (P^{f}_{15-44}):

$$\text{General abortion rate} = A \div P^{f}_{15-44} {}^{*}1{,}000 \tag{9.4}$$

The abortion ratio is the number of number of (legal) induced pregnancy terminations in a calendar year per 1,000 births:

$$\text{Abortion rate} = (A \div B)^{*}1{,}000 \tag{9.5a}$$

or, alternatively, the number of abortions in a calendar year per 1,000 births in the 12-month period from July 1 of the year to June 30 of the following year, to match the time of abortion with the time of birth more closely. The abortion rate (approximating a probability of abortion) is defined as the number of abortions during a calendar year per 1,000 births, late fetal losses, and abortions in the year:

$$\text{Abortion rate} = [A \div (B + L^{If} + A)]^{*}1{,}000 \tag{9.5b}$$

Alternatively, the births may be taken from July 1 of the year to June 30 of the following year. The restriction to legal, induced abortions and late fetal losses is

intended, as explained above, to lend stability and comparability to the measures by excluding the much more uncertain numbers of illegal abortions and early and intermediate fetal losses. Further disaggregation and refinement of these measures employ such variables as the woman's age, the total number of prior births, and the marital status of the woman.

Finally, I list (a) the percent of pregnancies that are terminated by abortion, or the number of legal abortions per 100 pregnancies, and (b) the average number of legal abortions per woman during her reproductive life.

$$\text{Abortion incidence rate} = (\text{A} \div \text{Pr})^{*}100 \tag{9.6}$$

Pregnancies can be estimated as the sum of births, late fetal deaths, and legal abortions, in a given calendar year.

$$\text{Average number of abortions per woman} = \text{A}_\text{T} \div \text{P}^{\text{f}}_{45-49} \tag{9.7}$$

where A_T represents the lifetime number of abortions reported by all women of reproductive age in a population.

Incidence and trends. The estimated world total number of abortions was 42 million for the year 2003 (Table 9.5), of which only about 52% occurred in countries where abortion is legal and safe and about 48% occurred in countries where abortion is illegal or restricted and unsafe. According to the United Nations (2003), in 1995 eighty percent of reported legal abortions were performed in four countries: China (7.4 million), the Russian Federation (2 million), the United States 1.2 million, and Viet Nam (1.2 million). General abortion rates, *i.e.*, abortions per 1,000 women 15–44 years of age, varied widely among the countries where abortion is legal. The United States had an intermediate figure of 20. The large majority of countries, including China, most countries of Europe, and a few in Western Asia, had rates of 10–25. Some countries of Europe, such as Germany and Spain, have very low rates (*i.e.*, less than 10). However, most of the successor states of the former Soviet Union, including the Russian Federation, Ukraine, Latvia, and Turkmenistan, as well as former socialist countries in Eastern Europe such as Bulgaria, Hungary, and Romania, have high abortion rates, that is, 25 or more abortions per 1,000 women 15–44 years of age.

It should be evident that abortion is used as a common form of family planning in the formerly socialist countries of Eastern Europe and the successor states of the former Soviet Union. In several such countries (*e.g.*, Russian Federation, Ukraine, Belarus, Estonia, Romania) at least one out of every two pregnancies is terminated by abortion (*i.e.*, 50 or more legal abortions per 100 known pregnancies). In the Russian Federation, women have an average of over two abortions (2.1) in their reproductive lifetime, whereas the average is 0.5 in the United States and 0.3 in Finland (United Nations 2003).

The high abortion numbers in Eastern Europe, the Russian Federation, and western Asia are a legacy of their lack of contraceptives during the Cold War

era. The general abortion rate reached unprecedented peaks in these countries just after the collapse of the former Soviet Union. Since that time the rate has declined in all less developed regions of the world and, especially, in the former member countries of the Soviet Union (Table 9.5). The new trend in Eastern Europe is due to several factors. New family planning programs were established, increased supplies of contraceptives became available, and religious groups became active in attacking the practice of abortion.

Regional variations in the characteristics of women having abortions are evident in the data. In the more developed countries and in Africa most abortions are secured by unmarried women (United Nations 2003). For example, in the United States the percentage of such abortions is 81 and in Mozambique it is 66. On the other hand, in Eastern Europe, the countries of the former Soviet Union, Asia, and Latin America, most abortions are secured by married women. For Albania the percentage of such abortions is 97, the Dominican Republic 88, and the Philippines 91. Age is another variable distinguishing the pattern of abortions. The proportion of pregnancies that are terminated is typically highest among adolescents and women aged 40 and older. Abortion among adolescents accounts for 3–20% of the total reported legal abortions in 39 out of 40 countries for which data are available. For the United States 20% of all abortions performed are secured by youths under age 20; that is, the percentage of abortions secured by U.S. youth ranks at the highest level among the countries for which data are available. Fifteen of the countries have percentages under 10.

Illegal and unsafe abortions. The principal concern with illegal abortions is the health risks associated with them. In fact, WHO's (1992) definition of an unsafe abortion includes all (induced) abortions performed in countries where abortion is illegal. Abortion is illegal is most less developed countries but for these countries few data on illegal abortions are available. Singh and Wulf (1994) made estimates of abortion rates for six Latin American countries on the basis of abortion-related hospitalization data. Abortion rates were quite high in five of the countries, namely Brazil, Chile, Colombia, Dominican Republic, and Peru; the estimates were 36 abortions or more per 1,000 women 15–44 years of age in these countries.

According to the World Health Organization (1992) an unsafe abortion is defined as a procedure for terminating an unwanted pregnancy either by persons lacking the necessary skills, or in an environment lacking the minimal medical standards, or both. It is assumed that induced abortions performed in countries where abortion is illegal, or legal but highly restricted, or performed by procedures that violate existing regulations (*e.g.*, where the fetus is over the allowable gestational limit) or by procedures not consistent with the state of the art, all meet the definition of unsafe abortion. Estimates of unsafe abortions per 1,000 women 15–44 years of age for the world's less developed regions are highest for Latin America and East Africa (over 30) (Fig. 9.2). Then in order come two other parts of sub-Saharan Africa, West Africa and Middle Africa, and finally comes South Central Asia (all over 20).

WHO estimates that unsafe abortions caused an annual average 78,000 maternal deaths in 1995–2000, approximately 68,000 deaths in 2004, and millions of injuries

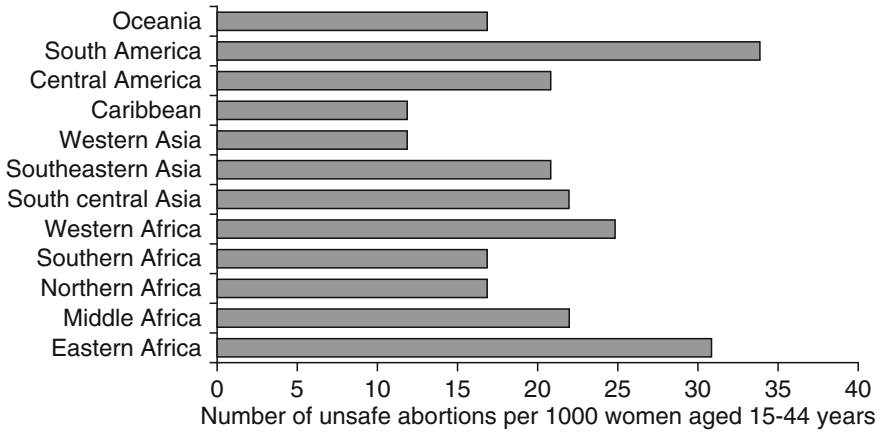


Fig. 9.2 Unsafe abortions per 1,000 women 15–44 years of age: World regions, Around 2003 (Source: WHO, UNICEF, and UNFPA 2003, Figure 1. Reprinted with permission of the World Health Organization. Based on Lancet series on Sexual and Reproductive Health, publication 2004, Table 3, page 13. www.un.org/apps/news/story.asp?NewsID=20445&Cr=sex& CR1=*2003 UNFPA global survey)

in each of these years (WHO 1998, 2006). The risk of death from an unsafe abortion is highest in Africa among the regions and lowest in Europe. In Africa the case-fatality rate reaches 7 deaths per 1,000 unsafe abortions and in Europe it is only about 0.5 per 1,000 unsafe abortions. Depending on the area, from 10% to 50% of all women who undergo unsafe abortions have complications requiring medical care. The most frequent complications are incomplete abortion, sepsis, hemorrhage, and intra-abdominal injury (United Nations 2003). A study conducted in Belize in 1990 illustrates the cost of unsafe abortions in terms of sickness and resources: Abortion was the third leading cause of hospitalization there (United Nations 1996).

The abortion problem is greatest where family planning services, information, and supplies are unavailable or inadequate. This situation is typical of several countries with Economies in Transition and many less developed countries. When abortion is made legal, safe, and accessible, women's health status improves rapidly, as in South Africa, where abortion was legalized in 1996. The incidence of illegal abortion may fall in the countries with high rates as they move on the path to lower fertility and seek to employ more humane family planning methods.

Estimating the Number and Rate of Pregnancies

Given sufficient data on births, fetal losses, and abortions, we can develop estimates of the number of women who are pregnant at various dates, pregnancy outcomes during various periods, and various types of pregnancy rates for such periods. In its simplest expression, total pregnancies at any date may be reconstructed from

subsequent births, (induced) abortions, and (spontaneous) fetal losses, and the total number of pregnancy terminations or outcomes in any year may be derived by combining births, fetal losses, and abortions during the year:

$$\text{Pregnancies} = \text{B} + \text{L}^{\text{f}} + \text{A} \tag{9.8}$$

Similarly, a general pregnancy (outcome) rate may be calculated as the ratio of the sum of the three components of pregnancy outcomes in the year per 1,000 women 15–44 years of age at the middle of the year. A set of consistent rates for the components of the pregnancy (outcome) rate may be derived in this process:

$$\text{GPR} = \text{GFR} + \text{GLR} + \text{GAR} \tag{9.9}$$

In this equation, the terms are the general fertility rate, the general fetal loss rate, and the general abortion rate, the rates are computed as the number of pregnancy outcomes per 1,000 women 15–44 years of age in a year, G equals general and FR, LR, and AR refer to the rates for the three components. In words,

General pregnancy rate = general fertility rate + general fetal loss rate + general abortion rate

For example, in the United States in 2005, the general pregnancy (outcome) rate was 103.2, derived as the sum of the general fertility rate (66.7), the general fetal loss rate (17.1), and the general (induced) abortion rate (19.4). (See U.S. NCHS/Ventura et al. 2009). The bulk of the pregnancies as estimated by NCHS ended in births (67%), 17% ended in fetal losses, and 19% ended in abortions. NCHS includes in the fetal loss total all reported fetal losses, that is, all (spontaneous) fetal losses from recognized pregnancies of all gestational periods. While this practice may yield a more complete estimate of pregnancy outcomes during a year than use of later fetal losses only, it may contribute to a greater lack of comparability of the estimates from year to year, given the evidence that most "miscarriages" are unknown to the mothers and many fetuses are "aborted" naturally. I prefer to include only intermediate and late fetal losses in this calculation. In any case, the true distribution of pregnancy outcomes includes a much greater proportion of fetal losses than even NCHS shows.

As suggested above, we can distinguish a least three types of estimates of the number of pregnancies, namely, an estimate of initial pregnancies, or conceptions, during a year, an estimate of pregnancies at a given date, and an estimate of pregnancy outcomes during a year. The number of conceptions during a year and the number of pregnancies at specific dates can be calculated by reverse estimation from the later component outcomes. Again, pregnancy outcomes for a particular year may be derived from the component data for births, fetal losses, and abortions for that year. These calculations differ in their complexity and precision. I describe some of them in Appendix 9.2.

Perinatal Mortality

The causes of death in early infancy are so akin to those accounting for fetal losses that various measures combining fetal losses and deaths of early infancy have been devised. The combination of these events, called perinatal mortality, is also intended to eliminate the errors resulting from deliberate and inadvertent misclassification among fetal losses, births, and neonatal deaths. The combined risk of dying during the period near parturition (*i.e.*, just before, during, and after birth) is measured by various perinatal mortality ratios and perinatal mortality rates. The formulas differ with respect to the age limits of the fetal losses and the infant deaths to be included, and with respect to whether fetal losses are included in or excluded from the base of the measures. Neonatal deaths, deaths under 1 week, and deaths under 3 days, in combination with intermediate and late fetal losses, or late fetal losses only, are possible ways of operationally defining perinatal deaths. In view of the general lack of tabulated data on infant deaths under 3 days, the choice of this period is not very useful for international comparisons.

The World Health Organization defines the perinatal period as the period of prenatal existence after viability of the fetus is reached, the duration of labor, and the early part of extra-uterine life. WHO recommends in ICD-9 that countries should present, solely for international comparisons, standard perinatal statistics in which the numerator and denominator of all rates are restricted to fetuses and infants weighing 1,000 g or more or, where birth weight is unavailable, the corresponding gestational age (28 weeks) or body length (35 cm. crown-heel) (United Nations 1994; World Health Organization 1992). According to the Tenth Revision, the perinatal period commences at 22 completed weeks (154 days) of gestation (the time when birth weight is normally 500 g) and ends at 7 completed days after birth.

Because birth weight and gestational age are not recorded on the death certificate in the United States, the U.S. National Center for Health Statistics (NCHS) was unable to adopt these definitions. NCHS now uses two definitions of perinatal mortality, designated I and II. In definition I, the one it recommends for international comparisons, the lower age limit of fetal loss is taken as 28 completed weeks of gestation and the early part of extra-uterine life is taken to be the first 7 days of life. The perinatal mortality rate is:

$$\mathrm{PMR_I} = (\mathrm{D^w} + \mathrm{L^{If}}) \div (\mathrm{B} + \mathrm{L^{If}} + \mathrm{D^w}) {}^{*} 1{,}000 \tag{9.10a}$$

where D^w represents infant deaths under 1 week, $\mathrm{L^{lf}}$ represents late fetal losses, and B represents births. The corresponding formula for the perinatal mortality ratio is, then, the number of deaths under 1 week of age plus late fetal losses (28 weeks of uterogestation) per 1,000 live births in a year:

$$\mathrm{PMR_a} = (\mathrm{D^w} + \mathrm{L^{If}}) \div \mathrm{B}^{*} 1{,}000 \tag{9.10b}$$

Table 9.6 Perinatal mortality ratios and perinatal mortality rates, for selected countries: Various years, 2003–2007

				Perinatal ratio[b]	Perinatal rate[b]
	Late fetal losses[a]	Deaths <1week	Births	[(1) + (2)] ÷(3) =	[(1) + (2)] ÷[(3) + (1) + (2)] =
Country, year	(1)	(2)	(3)	(4)	(5)
Canada, 2005	1,012	1,158	342,176	6.34	6.30
Cuba, 2007	1,017	220	112,472	11.00	10.88
Chile, 2006	2,124	995	231,383	13.48	13.30
Italy, 2005	1,776	1,073	554,022	5.15	5.11
Japan, 2007	2,254	1,052	1,089,818	3.03	3.02
Russian Fed., 2007	8,612	6,060	1,610,122	9.11	9.03
Kazakhstan, 2007	2,112	2,161	321,963	13.27	13.10
United States, 2003	12,485	15,133	4,089,950	6.75	6.71

Source: Based on United Nations, *Demographic Yearbook*, 2007, Tables 9 and 12
[a]Fetal losses of 28 weeks or more of gestational age
[b]Ratios or rates per 1,000

The definition of the perinatal mortality rate under definition II is the more inclusive one and is based on the number of infant deaths under 28 days and fetal losses of 20 weeks gestation or more:

$$PMR_{II} = (D^m + L^{if}) \div (B + L^{if} + D^m)^*1{,}000 \qquad (9.11a)$$

where L^{if} represents fetal losses of 20 weeks gestation or more and D^m represents infant deaths under 28 days. The corresponding perinatal mortality ratio differs by including only births in the base of the ratio. It is defined as:

$$PMR_b = (D^m + L^{if}) \div B^*1{,}000 \qquad (9.11b)$$

where the symbols have the same meaning as in the formulas above. These two measures give essentially the same indications of international differences, as can be noted in Table 9.6. The perinatal mortality rate has a small theoretical advantage over the perinatal death ratio, but the perinatal death ratio is probably more stable for international comparisons.

Factors Important in Maternal and Perinatal Mortality

The factors contributing to maternal mortality and those contributing to perinatal mortality are essentially the same. The direct causes of maternal and perinatal mortality include obstetric complications and unsafe abortions. Among the important indirect factors affecting the risk of maternal and perinatal mortality are the woman's

age and parity (*i.e.*, number of previous births), and the length of her birth intervals. Maternal and perinatal mortality rates are higher for very young women, higher-parity women, older women, and women with short birth intervals. Associated with these demographic characteristics are various health and socioeconomic conditions, such as a chronic degenerative disease (particularly diabetes), a sexually transmitted disease, malnutrition, poverty, unwanted pregnancies, prior difficult abortions, inadequate prenatal and obstetric care, and/or lack of access to hospital maternity services.

Birth Spacing

Because of its importance for the health of mothers and children, especially in the less developed countries, birth spacing merits special attention. Much research has consistently demonstrated that, when mothers space their births at least 2 years apart, they and their children are more likely to survive and be healthy than with shorter birth intervals. The infants are less likely to be premature, to suffer from low birth weight, and to be malnourished. Many reproductive health programs have recommended 2-year intervals and in surveys most women say that a 2-year birth interval is best.

Newerstudies show, however, that longer intervals are even better for infant and maternal survival and health (Setty-Venugopal and Upadhyay 2002). The new evidence, given in a Johns Hopkins University report authored by these researchers, indicates that couples who space their births three to 5 years apart increase their children's chances of survival as well as their own over those for births occurring between 2 and 3 years apart (Fig. 9.3). Children born three to 5 years after a previous birth are about 2 1/2 times more likely to survive through age 5 than children born before 2 years.[7] These findings emerge from a study conducted as part of the Demographic and Health Surveys (DHS) Program using data from 18 countries in four regions and assessing outcomes of more than 430,000 pregnancies. The DHS study statistically controlled for differences in demographic and socioeconomic variables, differences in prenatal care, the sex and survival of the previous child, and other factors that are believed to affect maternal and perinatal survival and health.

A study in 2000 by the Latin American Center for Perinatality and Human Development supplements the DHS study on childspacing, providing further evidence from over 450,000 women that spacing births farther apart improves mothers' health. This study pooled and analyzed data collected from hospital records between 1985 and 1997 in 19 countries of Latin America and the Caribbean. The data cover a variety of indicators, including mothers' sociodemographic characteristics, their

[7]Other more detailed findings are that, compared with children born less than 2 years after a previous birth, children born three to 4 years after a previous birth are 1.5 times more likely to survive the first week of life, 2.2 times more likely to survive the first 28 days of life, 2.3 times more likely to survive the first year of life, and 2.4 times more likely to survive to age 5.

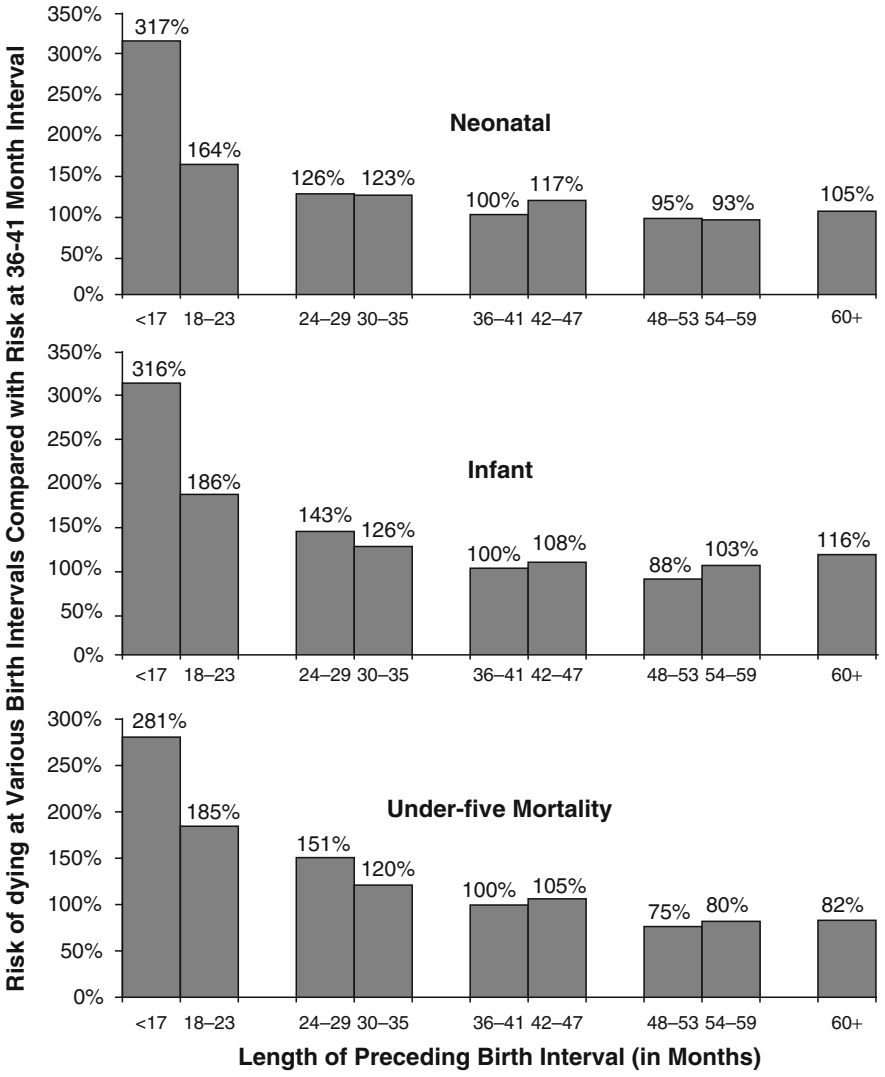


Fig. 9.3 Relative risk of dying during the neonatal period, infancy, and childhood through age five, by length of the preceding birth interval: 18 countries in 4 regions, 1992–1997 (Source: Setty-Venugopal and Upadhyay 2002. Based on Rutstein 2002a).

reproductive history, the health care they received during pregnancy and delivery, and their health and survival after delivery. Specific findings are that, compared with women who have births at 9-to-14-month intervals, women who have their babies at 27-to-32-month birth intervals are 1.3 times more likely to avoid anemia, 1.7 times more likely to avoid third-trimester bleeding, and 2.5 times more likely to survive childbirth (Setty-Venugopal and Upadhyay 2002).

Cultural factors affecting birth spacing. In the less developed regions women tend to have their births at shorter intervals than in the more developed regions. This is true in general of women who begin motherhood at young ages and have less education. Almost everywhere, women's birth intervals are shorter than they would prefer. In most countries a substantial unmet need for spacing births remains. Half of the total potential demand for contraception is for birth spacing. If women could achieve their preferred intervals, child mortality would fall. For example, Setty-Venugopal and Upadhyay (2002) calculate that in Kenya under-5 mortality would drop 17% if all women spaced their births at least 3 years apart. In Nigeria, under-5 mortality would fall 23%. Addressing the unmet need for birth spacing in the world would contribute to the survival of possibly an additional few million children to age 5.

Birth spacing is affected by various cultural norms. These include the pressure to prove one's fertility (and virility) by having the first child early, having many children, and having them in rapid succession. They include breastfeeding practices, postpartum abstinence, and preference for sons. In nearly all less developed countries women breastfeed their newborn children but breastfeeding differs among cultures both in duration and frequency. The DHS study, described by Setty-Venugopal and Upadhyay (2002), found that among the less developed regions the duration of breastfeeding ranges from an average of 14 months in Latin America and the Caribbean to 21 months in sub-Saharan Africa. Whether and how long a mother breastfeeds her child influences her childspacing practices and hence the child's survival chances. Breastfeeding practices help determine how long women will remain amenorrheic and thus less likely to get pregnant after giving birth. If the period of postpartum abstinence as traditionally practiced in a society exceeds the period of breastfeeding, the childspacing period may even be lengthened. Son preference is another factor tending to shorten birth intervals. As noted previously, preference for sons is widespread in the less developed regions, as in China, India, and other parts of South and East Asia. Couples who prefer sons tend to have their next child soon after the birth of a daughter.

Biological and behavioral basis for research findings. The exact biological and behavioral mechanisms accounting for the findings on the adverse effects of short birth intervals are not understood, but various researchers have proposed such factors as maternal depletion syndrome, premature delivery, milk diminution, and sibling rivalry. For instance, shorter birth intervals may not allow mothers enough time to restore nutritional reserves that allow for adequate fetal nutrition and growth. Fetal-growth retardation can result in low birth weight, which adds to the risk of premature death. Shorter intervals tend to be associated with an increased risk of premature birth, which can also result in low birth-weight babies and hence in premature death. Children born too close together compete for resources and maternal care, including breastfeeding.

Family Planning and Sexual Health

Role of Family Planning

The decision whether to have a child or not, the number of children wanted, the timing of childbearing, the use of contraception, and the particular methods of contraception employed have a major effect on the birth rates of a country. These are all part of family planning. Family planning has been used throughout prerecorded and recorded time by couples to affect the circumstances of childbearing, whether by late marriage, abstinence, breastfeeding, or other methods, but in modern times it has been made much more efficient by the development of contraceptive methods. Family planning is intended to make all pregnancies wanted, and wanted at the time of the pregnancy. Consequently, family planning tends to reduce, if not eliminate, the incidence of fetal losses and abortions, especially unsafe abortions, and reduce maternal and neonatal mortality. Family planning has been demonstrated to be a cost-effective way of improving maternal and child health. There is evidence that the greater the availability of family planning materials, services, and programs, the lower abortion rates are. Family planning is critical in preventing unwanted pregnancies and in spacing wanted children, so as to achieve the optimum outcome for the health of the mother and her children. Indirectly family planning can empower women to have the type of family they want and contribute to their opportunity to obtain an education, avoid poverty, and become upwardly mobile. Despite declines in fertility over the past half century, 26 out of 32 countries analyzed by DHS have high rates of unmet need for family planning. As described later, this unmet need is greatest in sub-Saharan Africa but it is also great in Asia and Latin America.

Concepts of Effectiveness of Contraception

Effectiveness of contraception may be considered in four dimensions: Theoretical effectiveness, use-effectiveness, unrestricted use-effectiveness, and demographic effectiveness. Theoretical effectiveness, also known as physiological effectiveness, refers to the effectiveness of a contraceptive method under ideal conditions, *i.e.*, when the method is used consistently and as instructed. A high level of theoretical effectiveness with a method can ordinarily be achieved under test conditions. Use effectiveness, also known as clinical effectiveness, refers to the actual experience of couples using a method, incorporating the occasional failures of the equipment and errors in its application by users and covering periods during which contraception is practiced regularly or irregularly. The use-effectiveness of a method may be expected to be substantially lower than the theoretical effectiveness of the same method. However, with permanent methods, such as intra-uterine devices, use-effectiveness may approximate theoretical effectiveness rather closely.

The experience of couples who were exposed to unintended pregnancy by periods of nonuse as well as irregular use even though they had adopted contraception is described as unrestricted use-effectiveness. This type of measure may be still lower than the other two. Demographic effectiveness refers to the effect on population growth of the use of the method in terms of births averted.

Theoretical effectiveness is determined in laboratory tests. Measurement of use effectiveness depends on the findings of surveys in which respondents report on their fecundity status, childbearing plans, and use of contraceptive methods. Demographic effectiveness requires model calculations of population scenarios with alternative assumptions of fertility based on varying assumptions regarding the use of contraceptive methods. A contraceptive averts births by delaying conception when a woman is in a fecundable condition.

Fecundability refers to the ability to conceive. Some fecundable women are infecund and, therefore, infertile; they are able to conceive but physiologically cannot complete a pregnancy. The term fecundability has taken on more specific meanings. It has been defined as the probability of conception per act of coitus and also as the probability of conceiving during a month for women who are cohabiting with male partners and who are not pregnant, sterile, or infecundable.

Contraceptive Methods and Prevalence

Methods of Contraception

There are numerous methods of contraception, some to be implemented by the male partner and others to be implemented by the female partner. Most methods of contraception require primary action by the female partner, as suggested by the list of methods given below. The methods differ greatly in their clinical and use effectiveness. The so-called traditional methods as a group are less effective than the so-called modern methods as a group. The modern methods include condoms (mainly for men but also for women), injectables (*e.g.*, Depo-Provera™, Lunelle™), oral pills, patches, IUDs, other vaginal methods (*i.e.*, spermicides, diaphragms, caps), RUA 486 (*i.e.*, the morning-after pill), and voluntary sterilization (of a woman or man). Traditional methods, also known as natural methods, include abstinence, rhythm, withdrawal, douche, and folk methods.

Measures of Contraceptive Prevalence and Findings

Both measures of prevalence and measures of incidence are applied to gauge the extent of contraceptive use. The most commonly used measures are relatively simple measures of prevalence, but multiple-decrement life tables have also been employed in the measurement of contraceptive practice.

The measures of contraceptive prevalence include the percent of women of reproductive age (15–44 or 15–49) ever using a contraceptive method and the percent currently practicing contraception. These measures are often calculated for specific groups such as age groups or marital groups. According to the National Survey of Family Growth conducted in the United States in 2002, nearly all (98%) women 15–44 years of age who have ever had sexual intercourse have used a contraceptive method, but a far smaller percentage used a method at first intercourse – 63% – and 62% used a contraceptive method in the month of interview (U.S. NCHS/Chandra et al. 2005). The latter percent varied between 32% for ages 15–19 and 61–71% for the other age groups.

The major methods depended upon by U.S. women 15–44 years of age at the time of interview in 2002 were the oral contraceptive pill (19%), female sterilization (17%), and the male condom (15%). (See U.S. NCHS/Chandra et al. 2005.) The pill is the leading contraceptive method among women under 35 years of age, while female sterilization is the leading method among women 35–44 years of age. A substantial share of users (1 in 6) were seeking more certainty of protection in that they used more than one contraceptive method at the same time. The common combination was the condom and the pill.

Education was a major factor affecting contraceptive use. For example, with increasing education a smaller and smaller proportion of unmarried women 22–44 years of age who had sexual intercourse in the 3 months prior to the interview did not use a method of contraception at last intercourse. Although an increasing proportion with increasing education used some form of birth control, 8% of those who were college graduates did not use any type of birth control. More specifically, 63% of college graduates did not use a condom. Use of a condom is of special interest because of its effectiveness in the prevention of STDs.

In the Less Developed Countries prevalence ratios of contraceptive use are based on the Demographic and Health Surveys (DHS), other sample surveys, or the number of patients served annually by the national family planning program. Figures on the prevalence of current contraceptive use by married women, according to method, in selected sub-Saharan African countries, other less developed countries, and Japan for various years from 1998 to 2002, are given in Table 9.7, using data mostly from the DHS. Seven of the 16 selected countries in sub-Saharan Africa show a current use of 15.0% or less and only 4 of the countries show a figure of 30% or more. For the 16 African countries shown, the highest figure was 39% for Kenya. Contraceptive use is much higher in North Africa, Latin America, and Asia than in sub-Saharan Africa.

The popularity of the methods varies from country to country. In Bolivia traditional methods were most popular at 23% out of 48% percent of users. Traditional methods were also common in several other countries listed – Gabon, Japan, Turkey, and Jordan. In India the leading method was female sterilization at 34% out of 48%, and in Egypt over one-third of contraceptors used the IUD. As we saw, female sterilization and the oral pill are common in the United States.

Table 9.7 Prevalence ratios of contraceptive use by married women of reproductive age, by method, for selected less developed countries and Japan: Various years, 1998 to 2002

Country, year	No method	Any method	Pill	IUD	Condom	Sterilization		Other modern	Traditional
						Male	Female		
Burkina Faso, 1998–1999	88.9	11.9	1.8	0.4	1.2	(NA)	0.1	1.4	7.0
Camaroon, 1998	80.7	19.3	2.0	0.6	2.1	(NA)	1.5	0.9	12.3
Cote d'Ivoire, 1998–99	85.0	15.0	3.5	0.4	1.8	(NA)	0.1	1.4	7.8
Ethiopia, 2002	91.9	8.1	2.5	0.1	0.3	(NA)	0.3	3.1	1.8
Gabon, 2000	67.3	32.7	4.8	(NA)	5.1	(NA)	1.0	0.8	21.0
Ghana, 1998	78.0	22.0	3.9	0.7	2.7	(Z)	1.3	4.1	9.2
Guinea, 1998	93.8	6.2	2.1	0.2	0.6	(NA)	(NA)	1.3	2.0
Kenya, 1998	61.0	39.0	8.5	2.7	1.3	(NA)	6.2	12.6	7.5
Malawi, 2000	69.4	30.6	2.7	0.1	1.6	0.1	4.7	16.5	4.9
Mali, 2001	91.9	8.1	2.8	0.2	0.3	(Z)	0.3	2.2	2.3
Mauritania, 2000–01	92.0	8.0	2.6	0.8	0.8	(NA)	(NA)	1.0	2.8
Nigeria, 1999	84.7	15.3	2.4	2.0	1.2	(NA)	0.3	2.7	6.7
Senegal, 1999	89.5	10.5	3.2	0.9	0.7	(NA)	0.5	2.8	2.3
Tanzania, 1999	74.6	25.4	5.3	0.4	2.7	(Z)	2.0	6.3	8.5
Uganda, 2000–01	77.2	22.8	3.2	0.2	1.9	(Z)	2.0	6.7	8.8
Zambia, 2001–02	65.8	34.2	11.9	0.1	3.8	(NA)	2.0	4.9	11.6
Bangladesh, 1999–2000	46.2	53.8	23.0	1.2	4.3	0.5	6.7	7.7	10.3
India, 1998–99	51.8	48.2	2.1	1.6	3.1	1.9	34.2	(NA)	5.0
Japan, 1998	42.3	57.7	1.1	3.2	79.6	1.2	4.7	1.9	24.5
Philippines, 1999	51.2	48.8	15.3	3.7	1.6	0.1	11.0	3.0	14.0
Egypt, 2000	43.9	56.1	9.5	35.5	1.0	(NA)	1.4	8.5	2.2
Jordan, 2002	44.2	55.8	7.5	23.6	3.4	(NA)	2.9	1.2	17.2
Turkey, 1998	36.1	63.9	4.4	19.8	8.2	(NA)	4.2	1.1	25.5
Bolivia, 1998	51.7	48.3	3.8	11.1	2.6	(NA)	6.5	1.1	23.1
Colombia, 2000	23.1	76.9	11.8	12.4	6.1	1.0	27.1	5.0	13.7
Ecuador, 1999	34.2	65.8	11.1	10.1	2.7	(NA)	22.5	5.1	14.4
Haiti, 2000	71.9	28.1	2.3	(NA)	2.9	(NA)	2.8	14.3	5.7
Nicaragua, 1998	39.7	60.3	13.9	9.1	2.6	0.5	26.1	5.2	2.5
Paraguay, 1998	37.7	62.3	13.1	11.1	7.3	(Z)	8.0	8.0	14.6

Source: U.S. Census Bureau (2004). Primary source: DHS (MACRO) and other sample surveys
Prevalence ratios are percentages of married women 15–49 years of age using contraceptives
NA not available, Z less than 0.05

The condom is very widely employed in Japan. Many countries (*e.g.*, Kenya, Senegal, Philippines, Peru, Paraguay) show a wide range of preferences of method.[8]

A companion measure gives the percent of married women of reproductive age (15–49) using contraceptives according to age of the woman at time of interview. Table 9.8 shows this measure for nearly the same list of countries and years, from the same source, as the table on type of contraceptive method. For women aged 15–19 years, the percentages are notably low in most less developed countries. From there the percentages move up quickly to a plateau at ages 25–39. At ages 40–44 the percentages are still substantially above those at ages 15–19. This suggests that, with increasing age, women are more disposed to accept the use of contraceptives as part of marital life.

Wantedness of Births, Reasons for Nonuse of Contraceptives, and Unmet Need

In the United States in 2002, according to the National Survey of Family Growth, 35% of women 15–44 years of age reported ever having had an unintended (*i.e.*, unwanted or mistimed) birth: 12% reported an unwanted birth and 23% reported a mistimed birth (U.S. NCHS/Chandra et al. 2005). Of the mistimed recent births, 12% were reported as occurring 2 or more years too soon. Having an unintended birth is closely associated with the age, education, and income of the mother. There is a pronounced inverse relation between these characteristics and the experience of having an unintended birth. Another view of these events is given by recent mistimed pregnancies, which includes births and fetal losses. Mistimed pregnancies are far greater among women under 20 years of age than among women 30–44 years of age. Among mistimed pregnancies those that are more seriously mistimed are at greater health risk than pregnancies that come at longer intervals, say at intervals of 2 years or more.

---

[8]The condom was originally designed to prevent infection but in the middle of the nineteenth century it began to assume a major role as a contraceptive. The condom continued in this use for many decades and by the mid-1930s it became the most common contraceptive method. (During World Wars I and II, when condoms were distributed to military personnel, its use as a protection against STDs was given new impetus but this practice also stigmatized its use as encouraging illicit sex.) By the 1960s it was replaced by the female methods of contraception, namely the pill and the IUD, in the industrialized countries (except in Japan). Use of the condom as a standard method of contraception is rare in the nonindustrialized countries as compared with the industrialized countries. According to the United Nations, 21% of married couples in the More Developed Countries and 5% of couples in the Less Developed Countries rely on the condom.

The emergence of HIV/AIDS in the 1980s, the rapid spread of the disease, and the current HIV/AIDS epidemic have brought the condom into use as the principal means of preventing the transmission of the disease. The condom is quite effective in the reduction of sexually transmitted diseases, but religious and cultural beliefs have been barriers to its use. It is greatly underused, especially in the areas where STDs are most prevalent. Sub-Saharan Africa exhibits the least condom use – several percent at most – and the worst HIV/AIDS epidemic. Strong efforts are being made to encourage greater use of the condom there as elsewhere.

Table 9.8 Prevalence ratios of contraceptive use by married women of reproductive age, by age, for selected less developed countries: Various years: 1998 to 2002

	Age group (years)						
Country, year	15–19	20–24	25–29	30–34	35–39	40–44	45–49
Burkina Faso, 1998–1999	0.9	12.9	12.6	15.4	13.6	10.0	5.1
Camaroon, 1998	15.4	16.8	20.3	19.3	23.9	23.4	15.5
Cote d'Ivoire, 1998–1999	10.7	13.9	18.4	25.6	13.1	21.5	8.3
Ethiopia, 2002	3.9	7.5	9.6	9.0	10.9	7.9	4.1
Gabon, 2000	40.0	37.4	38.2	34.6	28.6	22.7	15.9
Ghana, 1998	19.2	20.7	22.2	24.8	26.3	19.3	15.8
Guinea, 1999	3.4	6.1	5.5	6.9	7.8	7.1	6.3
Kenya, 1998	18.0	31.2	40.1	45.6	47.2	44.3	31.1
Malawi, 2000	15.2	26.3	34.6	35.8	36.7	37.7	25.7
Mali, 2001	4.9	7.1	8.6	9.8	10.5	7.9	5.6
Mauritania, 2000–2001	5.2	6.9	8.0	10.3	11.8	6.2	2.9
Nigeria, 1999	4.2	8.2	13.9	20.1	20.7	20.5	16.5
Senegal, 1999	3.1	8.5	10.1	12.0	13.0	14.3	11.0
Tanzania, 1999	10.6	28.4	24.9	29.5	28.4	30.7	16.9
Uganda, 2000–2001	12.0	21.0	24.4	26.6	25.8	26.7	18.0
Zambia, 2001–2002	25.5	31.6	36.4	44.0	36.9	34.9	18.4
Bangladesh, 1999–2000	38.1	47.1	58.1	64.2	67.7	61.9	43.1
India, 1998–1999	8.0	26.0	49.3	62.7	67.4	64.9	57.2
Philippines, 1999	21.9	41.5	52.3	56.1	57.4	51.1	32.2
Jordan, 2002	21.3	42.2	54.0	60.1	63.9	65.6	47.5
Saudi Arabia, 1996	17.0	29.4	36.3	34.9	34.5	28.5	19.9
Turkey, 1998	33.6	52.9	67.0	74.3	76.3	70.0	41.4
Bolivia, 1998	30.7	42.8	53.3	52.6	56.2	51.5	28.8
Colombia, 2000	57.2	69.1	77.4	80.3	83.7	82.6	71.4
Ecuador, 1999	37.1	57.2	67.2	74.6	76.1	69.4	55.5
Haiti, 2000	16.4	32.1	31.3	32.4	31.8	24.0	16.3
Nicaragua, 1998	39.8	53.8	64.6	69.4	70.0	63.7	48.1
Paraguay, 1998	44.4	60.9	62.0	62.7	67.5	64.0	(NA)

Source: U.S. Census Bureau 2004 Primary source: DHS (MACRO), CDC, or special survey
Prevalence ratios are percentages of married women 15–49 years of age using contraceptives

U.S. women 15–44 who had ever used a particular method of contraception and reported discontinuing it because of dissatisfaction with it gave the following reasons for their actions (U.S. NCHS/Chandra et al. 2005). Of the 29% who discontinued the pill, 65% reported that they experienced side effects, 13% were worried that they might experience side effects, and 13% did not like changes in their menstrual cycle. There was less dissatisfaction with the condom. Of the 12% who discontinued it, two reasons were commonly given: The partner did not like it and it decreased sexual pleasure. For the 42% who discontinued the injectables Depo Provera™ and Norplant™, by far the leading reason was the experience of side effects. Other reasons given for discontinuing a method included its cost,

difficulty or messiness in use, the failure of the method, failure of the method to protect against disease, and the advice of a physician.

A national sample survey conducted by the Guttmacher Institute in the United States in 2004 provided information about reasons for failure to use contraceptives by women at risk of unintended pregnancy (Frost et al. 2007). The most common reasons given were problems accessing or using methods (40%), including mainly side effects (17%). Infrequent sexual activity was another commonly cited reason for nonuse; 29% of nonusers gave this reason. In addition, 18–21% reported nonuse because of ambivalence about becoming pregnant, and 6–7% thought they could not get pregnant. Gaps in use coincided with certain life course events, such as moving to a new home, taking a new job, forming a new relationship, or experiencing a personal crisis.

Similar reasons are given by women in the Less Developed Countries, as reported by the Guttmacher Institute on the basis of data from the Demographic and Health Surveys (Sedgh et al. 2007). The women mention concerns about the health risks and side effects of various methods, inconvenience of the methods, doubts about the risk of getting pregnant, lack of availability of a method, opposition of the woman or her partner to use of contraception, and so on.

Other measures of contraceptive use distinguish the potential demand for contraception, the demand that is satisfied, and the unmet need (Sedgh et al. 2007; Casterline and Sinding 2000; Westoff and Bankole 1995; Goliber 1997). The concept of unmet need for contraception is important because addressing unmet need has become the basis of many family planning programs around the world. Unmet need may be measured simply as the proportion of women of reproductive age not using contraception but wishing either to prevent unwanted childbearing (after having achieved their desired number of children) or to postpone the next wanted birth. There is a standard definition of unmet need with more specific conditions, used in connection with DHS. The concept refers to women who are (1) married, in a consensual union, or unattached and sexually active; (2) fecund; (3) not desirous of having a child in the next 2 years; and (4) not using a contraceptive method, either modern or traditional.

The Sedge et al. (2007) study of recent DHS found that unmet need remains high in the Less Developed Countries (LDC). According to the study, one in seven married women in the LDC now has an unmet need for contraceptives. The relative need varies sharply between sub-Saharan Africa and the other Less Developed Regions. In the former region, one in four married women has an unmet need while in the latter regions one in seven to one in ten married women has an unmet need.

Use of Life Tables to Measure Family Planning Practices and Outcomes

Life table analysis may produce estimates of cumulative pregnancy rates and continuation rates for an array of exposure periods for a particular method of

contraception or for acceptors in general. Such measures may be derived from a multiple-decrement life table (Potter 1966). Cumulative pregnancy rates represent the proportion of acceptors who become pregnant during a given exposure period and hence represent the failure of a method. Continuation rates represent the proportion of acceptors who continue to use a contraceptive method for a given period after acceptance. The life-table technique gives measures of the gross and net use-effectiveness of a contraceptive method. For the construction of a multiple-decrement life table designed to measure net use-effectiveness of contraception or a particular contraceptive method, contraceptive acceptance rates and contraceptive termination rates for each period of exposure are needed (Potter 1967). A multiple-decrement table designed to measure the use-effectiveness of intrauterine devices, for example, would be constructed on the basis of acceptance rates for the devices and termination rates representing devices lost as a result of pregnancies, expulsions, and removals, by months since insertion. The life table yields cumulative pregnancy and continuation ratios as a result of the use of intrauterine devices.

Sexual Health

Although menarche in girls begins earlier than puberty in boys, in the United States males begin sexual activity on average slightly earlier than females. In most countries sexual activity begins between ages 15 and 19 years, so that the trend toward later marriage, particularly in the industrialized countries, has contributed to an increase in the prevalence of premarital sex. Sexual activity among teenage single persons is more common in the industrialized countries than in the unindustrialized countries and, as a result, adolescents in these countries have high rates of unintended pregnancies, unsafe abortions, and sexually transmitted diseases.

On the other hand, in sub-Saharan Africa, where a very high percentage of adolescents are sexually active (70% of total), most sexually active adolescents are likely to be married and a substantial share (20%) have their first child by age 18. Married women in sub-Saharan Africa often find it more difficult than single women to negotiate safe sex and few insist on condoms in their partners. Such a power relation increases the possibility of physical or sexual violence, which is quite common in many countries. The subordinate role of women forces them to tolerate multiple partners in their husbands, who may bring HIV and STDs home to their monogamous wives.

Young people aged 15-to-24 are particularly vulnerable to HIV and account for nearly half of all new infections reported in 2002 (Fig. 9.4). Family planning/contraception is an important tool in the fight against HIV. More young women would not incur HIV if contraception were more widely used. If unintended pregnancies are reduced through contraception, the transmission of HIV by women with HIV to their children would be reduced.

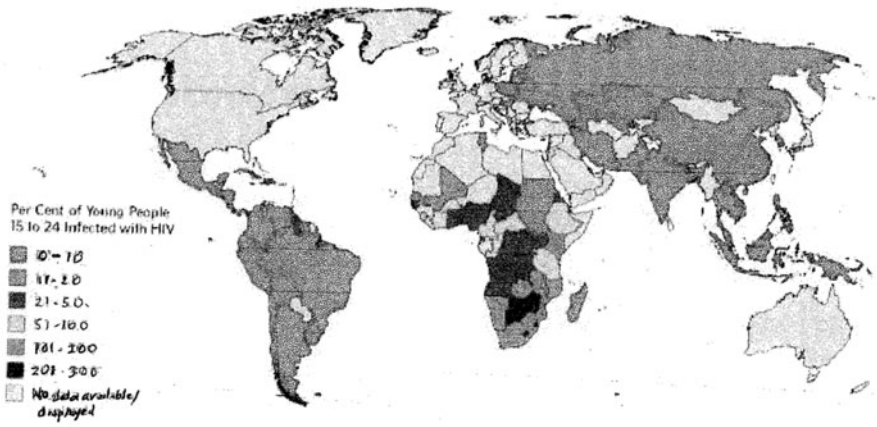


Fig. 9.4 World map showing percent of youth 15–24 years of age newly infected with HIV: 2002 (Source: UNAIDS 2002)

The 1994 UN Conference defined access to reproductive and sexual health services as a human right. The data show that this right is not being enjoyed in many parts of the world. In the world as a whole an estimated 62% of married women use contraception, but in sub-Saharan Africa the figure is only between 10% and 20%. Nevertheless, women want to control the number and spacing of their children. The current unmet need for contraception averages 10–12% in Asia and Latin America, and 24% in sub-Saharan Africa. In most countries a substantial unmet need for spacing births remains. Half of the total potential demand for contraception is for spacing. Addressing the unmet need for spacing would help millions of women achieve their family planning goals.

There is strong evidence that women want to control their lives and will seek out family planning assistance when it is made available. The national data on contraceptive use suggest that the more women have control over their own lives, the more likely they are to use family planning. On the other hand, studies carried out in connection with the Demographic and Health Surveys in the 1990s indicated that half of the women in the LDC identified as having an unmet need for contraception would not use it even if it were available (United Nations 2003). Specific education about family planning could make a difference in this number since lack of knowledge is the most frequently cited reason for not using birth control. Soap operas presenting birth control in a positive light led to increased contraceptive use and changed attitudes in India, Kenya, and Mexico. However, education does not necessarily alter cultural beliefs, and cultural beliefs play a big part in attitudes toward birth control. Religion also might be expected to have a large influence. In Kenya, where the Catholic Church is strongly opposed to use of contraceptives, the 90% access to contraceptives is matched by only a 30% use.

1. Total fertility rate
2. Contraceptive prevalence
3. Maternal mortality ratio
4. Antenatal care coverage
5. Births attended by skilled health personnel
6. Availability of basic essential obstetric care
7. Availability of comprehensive essential obstetric care
8. Perinatal mortality rate
9. Prevalence of low birth weight
10. Prevalence of positive syphilis serology in pregnant women
11. Prevalence of anemia in women
12. Percentage of obstetric and gynecological admissions due to abortion
13. Reported prevalence of women with genital mutilation
14. Prevalence of infertility in women
15. Reported prevalence of urethritis in men
16. Prevalence of HIV infection in pregnant women
17. Knowledge of HIV-related preventive practices

Exhibit 9.3 WHO's short list of indicators for global monitoring of reproductive health (Source: WHO 2006)

Monitoring and Evaluation

The Millennium Development Goals (MDG) promulgated by the United Nations in 2000 are a framework for measuring development progress in the world (Chap. 15). They aim particularly to reduce world poverty in all its forms and to establish a global partnership for development. General goals with respect to reproductive health are incorporated in the MDG. The final section of this chapter sets forth the shortlist of 17 indicators for global monitoring of reproductive health issued by the World Health Organization in the years following the earlier announcement of the MDG. In effect, this list recapitulates many of the measures and issues discussed earlier in this chapter. The indicators are shown in Exhibit 9.3 and the WHO's full statement on their generation, interpretation, and analysis is given in WHO (2006).

The shortage of reliable data represents a long-standing barrier to monitoring reproductive health. The United Nations has urged governments and nongovernmental organizations to strengthen national information systems in order to produce reliable statistics on reproductive health, including statistics on outcomes, access to services, and use of these services. The 17 indicators listed include indicators of outcome (10), access (5), and use (2). All the goals do not clearly fall within these three categories, but most of them do. For example, the maternal mortality ratio, perinatal mortality rate, and prevalence of HIV infection in pregnant women are outcome indicators; availability of basic essential obstetric care and availability of comprehensive essential obstetric care are access indicators; and contraceptive prevalence is a use indicator.

Appendix 9.1

The following formulas show the relation of various types of age-specific birth rates:

$$\frac{b_a}{P_a^f} = \text{Age-specific birth rate}$$

$$\frac{P_{am}^f}{P_a^f} = \text{Percent married women}$$

$$\frac{b_a^m}{P_a^{fm}} = \text{Age-specific birth rate for married women}$$

$$\frac{P_{anc}^{fm}}{P_a^{fm}} = \text{Percent noncontraceptors of married women}$$

$$\frac{b_{anc}^m}{P_{anc}^{fm}} = \text{Age-specific birth rate for noncontraceptor married women}$$

$$\frac{P_{ancfc}^{fm}}{P_{anc}^{fm}} = \text{Percent of noncontraceptor married women who are fecund}$$

$$\frac{b_{ancfc}^m}{P_{ancfc}^{fm}} = \text{Age-specific birth rate for noncontraceptor married women who are fecund}$$

These formulas tell us that the conventional age-specific birth rate can be converted into an age-specific marital birth rate by dividing it by the proportion of women who are married; the marital birth rate can be converted into a birth rate for noncontraceptor married women by dividing it by the proportion of married women who are noncontraceptors, and that the birth rate for noncontraceptor married women can be converted into the birth rate for married noncontraceptor women who are fecund by dividing it by the proportion of married noncontraceptor women who are fecund.

Appendix 9.2

To estimate the number of pregnancy outcomes for a given year, we can simply combine births, fetal losses, and abortions for the year, as indicated in Eq. 9.8. The fetal losses may nominally include all fetal losses reported, although this number will be grossly incomplete as an estimate of the total fetal losses because of the many unreported miscarriages. Alternatively, they may include only fetal losses of 28 weeks of uterogestation or more or fetal losses of 20 weeks of uterogestation

or more. These are more dependable figures for what they represent. The abortion figure may include illegal abortions as well as legal abortions, or only legal abortions; any estimate of pregnancy outcomes including illegal abortions will be incomplete as a total to the extent that illegal abortions are not fully reported. The estimates described below are based on fetal losses of 20 weeks of uterogestation or more and legal abortions.

To estimate the number of pregnancies at some particular date, we can employ either the conventional formula or some refinement of it. To estimate the number of pregnant women for July 1 of some year by the conventional procedure, we could simply combine births, fetal losses, and abortions in that year. Then we could use each of the components and the total of the components as numerators, and the number of women of childbearing age as denominators, in computing the general rate of pregnancy in the year, disaggregated into the general birth rate, the general fetal loss rate, and the general abortion rate.

The refined procedure is based on the fact that the pregnancy outcomes of a given year do not correspond closely to the pregnancies at any given date. This is because births, fetal losses, and abortions all have different characteristic gestation periods. The births of the current year were conceived approximately between April 1 of the previous year and March 31 of the current year. Assuming an even distribution of conceptions through this period, I assign them to the middle date, or October 1, to represent the date of the pregnancies that eventuated in births in the following year. Inasmuch as the fetal losses occurred after 20 weeks of uterogestation, I assume that conceptions during the period, August 10 of the previous year to August 10 of the current year gave rise to the fetal losses of the current year. The midperiod date for the fetal losses is February 10, therefore, assuming an even distribution of these events. I am assuming that the abortions were concentrated in the fourth to the sixth months of pregnancy even though they could have occurred over a wider period. Then, these pregnancies were mainly conceived between August 1 of the previous year and August 1 of the current year, with a mid period date of February 1 of the current year.

One purpose of this exercise is to derive a "population base" exposed to the risk of a pregnancy outcome in the form of a birth, fetal loss, or abortion in a given year. With such varied dates of reference for the components of the pregnancies that had outcomes in the year, we need to weight the dates if we wish to determine a single reference date for the initiation of the pregnancies. We weight them by the relative frequencies of the components:

Component	Distance from Jan 1	Weight[a]
Births	−3.0	.646
Fetal losses	+1.33	.166
Abortions	+1.0	.188
Weighted average	−1.94 = ca. November 2 of prior year	

[a]Based on the distribution for 2005

Hence, the approximate median date of reference – the date of conception – of the pregnancies that resulted in the outcomes of the current year is November 2 of the previous year. We also want sets of weights to derive an alternative base for calculating the risk of a conception ending in a birth, fetal loss, or abortion or any of these in the particular year. The weights give shares of births in the current year and births in the previous year, shares of fetal losses in the 2 years, and shares of abortions in the 2 years, similar to separation factors in computing infant mortality rates. On the basis of the periods of conception I have set down above for each component, the weights are:

	Current year	Prior year
Births	.250	.750
Fetal losses	.611	.389
Abortions	.583	.417

Applying these weights cumulatively to the events of each year yields the number of conceptions at risk of each or all of the events occurring in the current year.

It would be useful to display the relations between conceptions, pregnancies, fetal losses, abortions, and births in the form of a Lexis diagram. This diagram would link conceptions during a year, through fetal losses and abortions, with pregnancies at various dates in the same year and births in that year, and further, through additional fetal losses and abortions, with births in the following year. Such a Lexis diagram is complicated by the fact that births occur after only 9 months of uterogestation while birthdays come every 12 months.

References and Suggested Readings

General

Fathalia, M. (1988). Reproductive health: A global overview. *Annals of the New York Academy of Sciences*, *626*, 1–10.

Goliber, T. J. (1997). Population and reproductive health in sub-Saharan Africa. *Population Bulletin*, *52*(4), 1–44 (Washington, DC: Population Reference Bureau).

Lamb, V. L., & Siegel, J. S. (2004). Health demography. In J. S. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography*, (2nd ed.). San Diego: Elsevier/Academic.

National Academy of Sciences. (1997). *Reproductive health in developing countries*: *Expanding dimensions, building solutions*. Washington, DC: National Academy Press.

Obermeyer, C. M. (1996). A research agenda for reproductive health. *IUSSP Newsletter*, *54*, 10–19.

Rodgers, J. L., & Kohler, H. P. (2003). *The biodemography of human reproduction and fertility*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Rodgers, J. L., Rowe, D. & Miler, W. B. (2000). *Genetic Influences on human fertility and sexuality*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

United Nations (UN). (2003). *World population monitoring: Reproductive rights and reproductive health*, ST/ESA/SER.A/215. New York: United Nations.
United Nations Population Fund (UNFPA). (2004). *The state of world population: The Cairo Conference at Ten; Population, reproductive health, global efforts to end poverty*. Geneva, Switzerland: UNFPA.
World Health Organization (WHO). (2006). *Reproductive health indicators – Guidelines for their generation, interpretation, and analysis for global monitoring*. Geneva, Switzerland: World Health Organization.

Definitions, Sources, and Quality of Data

Hammond, E. I. (1965). Studies in fetal and infant mortality. I. A methodological approach to the definition of perinatal mortality. *American Journal of Public Health*, *55*(7), 1012–1023.
Stewart, M. K., Stanton, C. K., Festin, M., & Jacobson, N. (1996). Issues in measuring maternal morbidity: Lessons from the Philippines Safe Motherhood Survey Project. *Studies in Family Planning*, *27*(1), 29–36.
United Nations. (1994). *World plan of action conference document*. In International conference for population and development, Cairo, Egypt. New York: United Nations.
U.S. National Center for Health Statistics (NCHS). (1996). *Vital statistics of the United States*, *1991*. Hyattsville, MD: National Center for Health Statistics.
U.S. National Office of Vital Statistics (NOVS). (1954). *Vital statistics of the United States, 1950*. Washington, DC: U.S. Public Health Service.
Watkins, S. C. (2004). Navigating the AIDS epidemic in rural Malawi. *Population and Development Review*, *30*(4), 690–698.
World Health Organization (WHO). (1989, August 30–September 1). *Measuring reproductive morbidity*. Report of a Technical Working Group, Geneva, Switzerland: World Health Organization.
Zurayk, H., Khattab, H., Younis, N. et al. (1993, August 24–September 2). *A comparison of women's report and medical diagnosis of reproductive morbidity conditions in rural Egypt.* In Proceedings of the International Union for the scientific study of population, Montreal, Canada.

Biodemographic Concepts, Fertility Patterns, and Reproductive Aging

Alan Guttmacher Institute. (2002). Sexual and reproductive health: Women and men. In *Facts in brief*. New York City: Alan Guttmacher Institute.
Alan Guttmacher Institute. (2007). Facts on induced abortions. In *Facts in brief.* New York City: Alan Guttmacher Institute.
Austad, S. (1997). *Why we age*? New York: Wiley.
Carnes, B. A., Olshansky, S. J., & Grahn, D. (2003). Biological evidence for limits to the duration of life. *Biogerontology*, *4*(1), 31–45.
Chandra, A., & Stephen, E. H. (1998). Impaired fecundity in the United States: 1982–95. *Family Planning Perspectives*, *30*(1), 34–42.
Coale, A.J., and P. Demeny. 1966. *Regional model life tables and stable model life tables*. Princeton, NJ: Princeton University Press.
Davis, D. L., Webster, P., Stainthrope, H., et al. (2007). Declines in sex ratio at birth and fetal deaths in Japan, and in U.S. whites but not African Americans. *Environmental Health Perspectives*, *115*(6), 941–946.

Dunne, M. P., Martin, N. G., Stathman, D. J., et al. (1999). Genetic and environmental contributions to variance in age at first sexual intercourse. *Psychological Science*, *8*, 211–216.

Eaton, J. W., & Mayer, A. J. (1953). The social biology of very high fertility among the Hutterites: The demography of a unique population. *Human Biology*, *25*, 206–264.

Espenshade T. (1971). Method for estimating the level of natural fertility in populations practicing birth control. *Demography*, *8*(4), 525–536.

Ferrell, R. J., O'Connor, K. A., Rodriguez, G., Gorrindo, T., et al. (2005). Monitoring the transition to menopause in a five-year prospective study: Aggregate and individual changes in steroid hormones and menstrual cycle lengths with age. *Menopause*, *12*(5), 567–577.

Ferrell, R. J., O'Connor, K. A., Rodriguez, G., Holman, G., et al. (2007). Monitoring reproductive age in a five-year prospective study: Aggregate and individual changes in luteinizing hormone (LH) and follicule-stimulating hormone (FSH) with age. *Menopause*, *14*(1), 29–37.

Fisch, H. (2005). *The male biological clock* (p. 24). New York: Free Press.

Frisch, R. E. (2002). *Female fertility and the body fat connection*. Chicago: University of Chicago Press.

Garenne, M. (2002). Sex ratios at birth in African populations: A review of survey data. *Human Biology*, *74*(6), 889–900.

Garenne, M. (2004). Sex ratios at birth in populations of Eastern and Southern Africa. *South African Journal of Demography*, *9*(1), 91–96.

Henry, L. (1961). Some data on natural fertility. *Eugenics Quarterly*, *8*, 81–91.

Holliday, R. (1995). *Understanding aging*. Cambridge, UK: Cambridge University Press.

Holliday, R. (2004). The multiple and irreversible causes of aging. *Journal of Gerontology: Biological Sciences*, *59A*(6), 568–572.

Hyrenius, H. (1965). *New technique for studying demographic-economic-social interrelations*. Paper originally presented at the United Nations Conference on the application of science and technology for the benefit of the less developed areas, Geneva, February 4–20, 1963. University of Göteborg, Göteborg, Sweden, Reports 3.

Kline, J., Stein, Z., & Susser, M. (1989). *Conception to birth: Epidemiology of prenatal development*. New York: Oxford University Press.

Kohler, H. J., Rodgers, J. L., & Christensen, K. (1999). Is fertility behavior in our genes? *Population and Development Review*, *25*, 253–288.

Lian, Z. H., Zack, M. M., & Erickson, J. D. (1986). Paternal age and the occurrence of birth defects. *American Journal of Human Genetics*, *39*, 648–660.

Morris, J. K. (2008). Down's syndrome live births in England Wales: the effect of the National Fetal Anomaly Screening Programme *Genus*, *LXIV*(3–4), 91–98.

O'Connor, K. A., Holman, D. J., & Wood, J. W. (1998). Declining fecundity and ovarian aging in natural fertility populations. *Maturitas*, *30*, 127–136.

O'Connor, K. A., Holman, D. J., & Wood, J. W. (2001). Menstrual cycle variability and the perimenopause. *American Journal of Human Biology*, *13*(4), 465–478.

Perls, T., & Fretts, R. (2001). The evolution of menopause and human life span. *Annals of Human Biology*, *28*, 247–245.

Rinaldi, A., Bigazzi, R., de Iasio, S., & di Bacco, M. (2003). Individual variability of sex ratio: A statistical control. Genus, *49*(2), 29–36.

Robinson, W. (1986). Another look at the Hutterites and natural fertility. *Social Biology*, *33*, 65–76.

Thomas, F., Renaud, F., Benefice, E., de Meeus, T., & Guegan, J. F. (2001). International variability of ages at menarche and menopause: patterns and main determinants. *Human Biology*, *73*(2), 271–290.

United Nations (UN). (2003). *World population monitoring: Reproductive rights and reproductive health*, ST/ESA/SER.A/215. New York: United Nations.

U.S. National Center for Health Statistics (NCHS). (1997). Fertility, family planning, and woman's health: New data from the National Survey of Family Growth. By J. A. Abma, W. Chandra, W. Mosher, L. S. Peterson, & L. J. Piccinino. *Vital and Health Statistics*, *23*(19).

U.S. National Center for Health Statistics (NCHS). (2005). Trend analysis of the sex ratio at birth in the United States. By T. J. Mathews & B. E. Hamilton. *National Vital Statistics Reports*, *53*(20).

U.S. National Center for Health Statistics (NCHS). (2006). Births: Final data for 2004. By J. A. Martin, B. E. Hamilton, P. D. Sutton, S. J. Ventura, et al. *National Vital Statistics Reports*, *55*(1).

Vandresse, M., Gourbin, C., Horvath-Puho, E., Csaky-Szunyogh, M., et al. (2008). Impact of late fertility on congenital abnormalities. *Genus*, *LX–IV*(3–4), 33–62.

Vermeulen, A. (2000). Andropause. *Maturitas*, *34*, 5–15.

Wood, J. W., O'Connor, K. A., Holman, D. J., Brindle, E., et al. (2001). *The evolution of menopause by antagonistic pleiotropy* (*Working paper* 01–04. Center for Studies in Demography and Ecology). Seattle, WA: University of Washington.

Wyshak, G., & Frisch, R. E. (1982). Evidence for a secular trend in age at menarche. *New England Journal of Medicine*, *306*(17), 1033–1035.

Yi, Z, & Vaupel, J. W. (2004). Association of late childbearing with healthy longevity among the oldest-old in China. *Population Studies*, *58*(1), 37–54.

Maternal Health

AbouZahr, C., & Royston, E. (1991). *Maternal mortality: A global fact book*. Geneva, Switzerland: World Health Organization.

Fortney, J. A., Susanti, I., Gadalla, S., Saleh, S., et al. (1986). Reproductive mortality in two developing countries. *American Journal of Public Health*, *76*(2), 134–138.

Graham, W., & Campbell, O. (1990). *Measuring maternal health: Defining the issues*. London: Maternal and Child Epidemiology Unit, London School of Hygiene and Tropical Medicine.

International Union for the Scientific Study of Population (IUSSP). (1992a). *The analysis of maternity histories*. Liège, Belgium: Ordina Editions.

International Union for the Scientific Study of Population (IUSSP). (1992b). *Measurement of maternal and child mortality, morbidity and health care: Interdisciplinary approaches*. Liège, Belgium: Ordina Editions.

Philippines, *National Statistics Office, and Macro International. (1994). National safe motherhood survey*, *1993*: *Demographic and health surveys.* Calverton, MD: Macro International.

Population Reference Bureau. (2001). *Abandoning female genital cutting: Prevalence, attitudes, and efforts to end the practice*. Washington, DC: Population Reference Bureau.

Rooks, J., & Winikoff, B. A. (1990). *A reassessment of the concept of reproductive risk in maternity care and family planning services*. In Proceedings of a Seminar, 12–13 Feb 1990. A Supplement of the *Population and Development Review*. New York: Population Council.

Rutstein, S. (2002a). *Effect of birth intervals on mortality and health: Multivariate cross-country analyses*. Paper presented at the Champions Meeting on Birth Spacing, 31 Jan 2002. Washington, DC: CATALYST Consortium.

Rutstein, S. (2002b). *Relationship between pregnancy intervals and perinatal mortality*. In Proceedings of the 2nd Champions Meeting on Birth Spacing, 2 May 2002. Washington, DC: CATALYST Consortium.

Setty-Venugopal, V., & Upadhyay, U. D. (2002). Birth spacing: Three to five saves lives. *Population Reports: Issues in World Health*, *30*(3), Series L, No. 13. Baltimore, MD: Johns Hopkins University, Bloomberg School of Public Health.

Shahidullah, M. (1995). The sisterhood method of estimating maternal mortality: The Matlab experience. *Studies in Family Planning*, *26*(2), 101–108.

Stecklov, G. (1995). Maternal mortality estimation: Separating pregnancy-related and non-pregnancy-related risks. *Studies in Family Planning*, *26*(1), 33–38.
Stewart, M. K., Stanton, C. K., Festin, M., & Jacobson, N. (1996). Issues in measuring maternal morbidity: Lessons from the Philippines Safe Motherhood Survey Project. *Studies in Family Planning*, *27*(1), 29–36.
Tinker, A. & Koblinsky, M. A. (1993). *Making motherhood safe* (World Bank Discussion Paper No. 202). Washington, DC: The World Bank.
World Health Organization (WHO). (1990, August 30–September 1). *Measuring reproductive morbidity*. (Report of a Technical Working Group, WHO/MCH/90.4). Geneva, Switzerland: WHO.
World Health Organization (WHO). (1996). *Revised 1990 estimates of maternal mortality: A new approach by WHO and UNICEF*. Geneva, Switzerland: World Health Organization.
World Health Organization (WHO). (2005). *World health report*. Geneva, Switzerland: World Health Organization.
World Health Organization, UNICEF, and UNFPA.(2003). *Maternal mortality in 2000: Estimates developed by WHO, UNICEF, and UNFPA*. Geneva, Switzerland: World Health Organization.
Younis, N., Khattab, H., Zurayk, H., El-Mouelly, M., et al. (1993). A community study of gynecological and related morbidities in rural Egypt. *Studies in Family Planning*, *24*(3), 175–186.
Zurayk, H., Khattab, H., Younis, N., El-Mouelly, M., & Fadle, M. (1993). Concepts and measures of reproductive morbidity. *Health Transition Review*, *3*(1), 17–40.

Measures of Pregnancy Losses

Alan Guttmacher Institute. (2007). Facts on induced abortions worldwide. In *Brief facts*. New York: Alan Guttmacheer Institute.
Anderson, B., Katus, K., Puur, A., & Silver, B. (1994). The validity of survey responses on abortion: Evidence from Estonia. *Demography*, *31*(1), 115–132.
Casterline, J. B. (1989). Collecting data on pregnancy loss: A review of evidence from the World Fertility Survey. *Studies in Family Planning*, *20*(2), 81–95.
Coeytaux, F., Leonard, A., & Royston, E. (Eds.). (1990). *Methodological issues in abortion research*. A PDR Supplement: Proceedings of a seminar held in December 1989. New York: Population Council.
Henshaw, S. K., & Van Vort, J. (Eds.). (1992). *Abortion handbook: Readings, trends, and state and local data to 1988*. (1992 ed.). New York: Alan Guttmacher Institute.
James, W. H. (1970). The incidence of spontaneous abortion. *Population Studies*, *24*, 241–245.
Kerr, M. G. (1971). Prenatal mortality and genetic wastage in man. *Journal of Biosocial Science*, *3*(2), 223–237.
Loghi, M., Georgi, P., & Pierannunzio, D. (2008). Negative pregnancy outcomes by gestational age in Italy. Comparing different international definitions. *Genus*, *LXIV*(3–4), 83–89.
Singh, S. & Wulf, D. (1994). Estimated levels of induced abortion in six Latin American countries. *International Family Planning Perspectives*, *20*(1), 4–13.
United Nations (UN). (1996). *Consideration of reports submitted by States parties under Article 18 of the convention to eliminate discrimination against women: Initial and second periodic reports of States parties*. Belize: Committee on the Elimination of All Forms of Discrimination against Women. CEDAW/C/BLZ/1–2.
U.S. National Center for Health Statistics (NCHS). (2007). Fetal and perinatal mortality, United States, 2004. By M. F. MacDorman, M. L. Munson, & S. Kimeyer. *National Vital Statistics Reports*, *56*(3).

U.S. National Center for Health Statistics (NCHS). (2009). Estimated pregnancy rates for the United States, 1990–2005: An update. By S. J. Ventura, J. C. Abma, W. D. Mosher, & S. K. Henshaw. *National Vital Statistics Reports*, *58*(4).

World Health Organization (WHO). (1970). *Spontaneous and induced abortion* (Technical report series No. 461). Geneva, Switzerland: World Health Organization.

World Health Organization (WHO). (1992). *The prevention and management of unsafe abortion: Report of Technical Working Group* (WHO/MSM/92.5). Geneva, Switzerland: World Health Organization.

World Health Organizaton (WHO). (1998). *Unsafe abortion: Global and regional estimates of incidence of and mortality due to unsafe abortion with a listing of available country data* (3rd ed.). Geneva, Switzerland: World Health Organization.

Family Planning and Sexual Health

Casterline, J. B., & Sinding, S. W. (2007). *Unmet need for family planning in developing countries and implications for population policy* (Policy Research Division, Working Paper 135). New York: Population Council.

Degrées du Loû, A. (1999). Reproductive health and AIDS in sub-Saharan Africa: Problems and prospects. *Population*, 4, 1998. *Population*: *An English Selection 11*, 1999, 61–88. Paris, France: INED.

de Guibert-Lantoine, C., & Leridon, H. (1999). Contraception in France: An assessment after 30 years of liberalization. *Population*: *An English Selection*, *11*, 89–114. Paris, France: INED.

Frost, J. J., Singh, S. & Finer, L. B. (2007). U.S. women's one-year contraceptive use patterns, 2004. *Perspectives on Sexual and Reproductive Health*, *39*(1), 48–55.

International Union for the Scientific Study of Population (IUSSP). (1975). *Measuring the effect of family planning programs on fertility*. Liège, Belgium: Ordina Editions.

International Union for the Scientific Study of Population (IUSSP). (1984). *Survey analysis for the guidance of family planning programs*. Liège, Belgium: Ordina Editions.

National Academy of Sciences. (1998). *Adolescent reproductive behavior in the developing world*. Washington, DC: National Academy Press.

Potter, R. G. (1966). Application of life table techniques to measurement of contraceptive effectiveness. *Demography*, *3*(2), 297–304.

Potter, R. (1967). *The multiple-decrement life table as an approach to the measurement of the use-effectiveness and demographic effectiveness of contraception*. In Proceedings of the international union for the scientific study of population, 1967, Sydney, Australia.

Ross, J. A., Mauldin, W. P., & Miller, V. C. (1993). *Family planning and population: A compendium of international statistics* (A PDR Supplement). New York: Population Council.

Sedgh, G., Hussein, R., Bankole, A., & Singh, S. (2007). *Women with an unmet need for contraception in developing countries: Their reasons for not using a method*. New York: Guttmacher Institute.

U.S. Census Bureau. (2004). *Global population profile*, *2002* (*International population reports* WP/02). Washington, DC: U.S. Government Printing Office.

U.S. National Center for Health Statistics. (2005). Fertility, family planning, and reproductive health of U.S. women: Data from the 2002 National Survey of Family Growth. By A. Chandra, G. M. Martinez, W. D. Mosher, J. A. Abma, & J. Jones. *Vital and Health Statistics*, *23*(25). Hyattsville, MD: National Center for Health Statistics.

United Nations Population Fund (UNFPA). (2004). *State of the World, 2004. The Cairo Conference at Ten: Population, reproductive health, and the global effort to end poverty*. Geneva, Switzerland:UNFPA.

UNAIDS. (2002). *Report on the Global HIV/AIDS Epidemic, 2002*. Geneva, Switzerland: UNAIDS.

UNAIDS. (2007). *AIDS Epidemic update: December 2007*. Geneva, Switzerland: UN AIDS and WHO.

Westoff, C. F. & Bankole, A. (1995). *Unmet need*: 1990–1994 (*DHS comparative studies*, No. 16). Calverton, MD: Macro International.

Chapter 10
Health and Migration

Introduction[1]

There are numerous ways in which migration and health influence each other, but only modest attention was devoted to these relationships in published research until the 1980s. It is reasonable to expect linkages between migration and health for several reasons. Migrants tend to be a selected subgroup of people from their area of origin and they often carry with them unique lifestyles and health attributes. The process of migrating from one place to another often has important health consequences for both the migrants and the people to whom they are exposed in the place of destination.

Such movement may be expected to modify the health situation in both the area of origin and the area of destination. Whether the movement of people from one area to another has a positive or negative effect on the health level of the populations in the areas of origin and destination depends on the health status, lifestyles, and even genetic predispositions of the migrants and the origin and destination populations at the time the migration occurs and in the years that follow. The movement of persons may have the opposite effect on the receiving area from that on the sending area. For example, the migration of relatively healthy people from a relatively unhealthy

[1]Various concepts relating to spatial mobility will be introduced as considered useful during the course of the chapter. This footnote defines a few basic concepts following U.S. practice. At this general level, other countries use a similar typology. Movers are persons who change their usual residence. If the distance is substantial within the country or the movement is across a specified local boundary within the country (e.g., county in the United States), the movers are called internal migrants. They are designated either inmigrants or outmigrants with respect to a particular area depending on the direction of the movement. The balance of inmigration and outmigration is designated net inmigration or net outmigration, depending on the preponderant direction of the movement. Persons who change usual residence between countries are labeled international migrants, and they are called immigrants or emigrants with respect to a particular country according to the direction of the movement. The balance of immigration and emigration is called net immigration or net emigration as appropriate.

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_10,

area will worsen the overall health status of the population of origin because of the selective removal of a healthier segment of the population, and if the migrants are healthier than the people living in the area of destination, they may improve the overall health situation of that area. Migrants may be self-selected with respect to their health condition and/or their health condition may be affected by the movement itself. International migrants may be bearers of disease and, because of their often relatively low socioeconomic status and their possible lack of immunity to "foreign" diseases, they may be more vulnerable to an adverse health environment in the receiving area. The mere change of environment may modify their susceptibilities to disease. Migration may affect their access to and use of health services and of health resources, and affect their health behavior, in the new area of residence.

Trying to disentangle cause and effect, and the role of the selection factor, in the relation between migration and health is difficult and uncertain. This is all the more true if the study vehicle is a cross-sectional survey or even a series of such surveys rather than a longitudinal survey, or the study material is aggregated data for area units rather than microdata. In migration studies, two areas are involved and relations between the migrants and the resident populations in the areas of origin and destination are complex. Furthermore, health and migration may be both a cause and a consequence of one another.

Studies of the relation of migration and health deal with a variety of questions linking these two factors. Some illustrations are as follows. How do the mortality and morbidity of migrants compare with the mortality and morbidity of natives/residents of the destination country/area? How do they compare with the mortality and morbidity of the nonmigrants in the area of origin? Do the health differences between the immigrants and residents of the destination area vary with respect to age and sex? Can any differences be explained in terms of acculturation, lifestyle, and psychological factors? Can the effect of differences in socioeconomic status and neighborhood characteristics in the home residence be distinguished from socioeconomic status and neighborhood characteristics acquired in the destination area? Can this analysis be done for birth cohorts, departure cohorts, or arrival cohorts, not simply cross-sectional groups or age groups? Do the differences in the various comparisons change with the generations, that is, is there evidence of health assimilation? How do the national subgroups of the immigrants fare in these comparisons? What are the determinants and consequences of the findings and the policy implications of the findings?

Finally, much could be learned about the relation of migration and health if detailed longitudinal data about migration and health were available for individuals in a life course framework. Much of the health-induced migration occurs in later life on the occasion of a person's loss of functional capacity or death of a spouse (Litwak and Longino 1987). The U.S. Health and Retirement Study and other longitudinal surveys are providing new data on these matters and are making possible the formulation of firmer hypotheses and inferences regarding the cause-and-effect relations of migration and life course events such as health.

International Migration

Sources and Limitations of Data

Data on the health of immigrants are not usually found in the statistical, demographic, or immigration yearbooks of countries. For example, there are no data on the health of the immigrants admitted to the United States in the *Yearbook of Immigration Statistics* (or its earlier versions). A few statistics were once given on exclusions for reasons of physical or mental health and immigrants admitted under occupational preferences as health providers. Some information may be available from the health records of the country of birth or last permanent residence or from emigration documents of these countries. Information on the health status of a country's emigrants is, however, very difficult to secure.

Some information on the disability status of foreign-born persons can be obtained from census records and national sample surveys (e.g., American Community Survey), but the national sample health surveys and administrative records on health conditions and on the use of health services are more productive sources. From the health surveys we need to secure tabulations of persons according to nativity status (i.e., native, foreign-born) and, for the foreign born, according to country of last residence or country of birth, as well as data on health status. For the United States, the National Health Interview Survey, the National Longitudinal Mortality Study, and the Health and Retirement Study are sources of data of this kind and can be used in combination with the Multiple-Cause-of-Death Data Set and the National Death Index to determine whether any of the interviewees died. Administrative records useful for analysis of the mortality and health of immigrants include Medicare records, Social Security records, including the Numident files (the application for Social Security benefits), and the vital statistics data compiled by the National Center for Health Statistics.

Analysis of the interrelations of health and international migration is often made in terms of a comparison between natives and foreign-born persons, who serve as proxies for non-immigrants and immigrants, respectively. As suggested, data of this type are commonly provided by censuses, surveys, and health records, and not by immigration records. An extension on the usual survey design of interviewing foreign-born persons in a country and comparing their health status with native respondents is to trace and interview also the immigrants' close relatives (e.g., siblings) who remained in the country of origin. Such data would aid in testing the "healthy immigrant hypothesis," described below.

Multiplicity or network sampling in the country under study is another possible method of securing information about emigrants and their characteristics in a sample survey. It identifies emigrants by inquiring about emigration not only of the members of the sample households but also about the emigration of their close relatives. More specifically, information on who has gone abroad to live during a specified recent period (e.g., a year or 5 years) is obtained about close relatives of the original sample households, such as siblings, parents, and children. This survey

design, in effect, extends the size of the sample in order to secure more reliable information on a relatively uncommon event. The results may be biased in that single persons who have gone abroad or died, or households all of whose members have gone abroad or died, are less likely to be represented in the sample although these events may be reported by relatives in network sampling. Moreover, the respondents may not provide accurate information about the characteristics and movements of their emigrant relatives. The results are subject to other problems as well. (See piece on Multiplicity in Appendix A for further details.)

Data on foreign-born persons have analytical limitations in representing immigrants quantitatively during any specified period. They understate the number of immigrants during the period of study in that they exclude those immigrants who died or emigrated during the period. We can, therefore, characterize the foreign-born population at any date as surviving net immigrants.[2] In addition, data needed to study assimilation of the immigrants with respect to health and other characteristics, such as year of arrival, or period of residence, in the country of immigration, are often unavailable.

Several federal agencies (NIH, NSF, CIS) and the Pew Charitable Trust are now supporting a longitudinal survey of immigrants in the United States, called the New Immigrant Survey (NIS). It is a survey of recent legal immigrants, based on nationally representative samples of the records of immigrants newly admitted to permanent legal residence in the United States during 1996 (Pilot Study), 2003, and 2007, compiled by the U.S. Citizenship and Immigration Services (CIS). It is a multi-cohort, retrospective panel survey whereby new samples of immigrants are selected at regular intervals and each sample is reinterviewed periodically to observe changes over the life cycle of the immigrants. Information is also being obtained about and from the children of the immigrants, whether born abroad or in United States. Questions relate to the past as well as the present, and include the subjects of health and a range of demographic and socioeconomic characteristics. Research results are being made available as the study proceeds. For example, reduction in the prevalence of smoking among the immigrant population is an area of interest

[2]Foreign-born persons at a given date for a given country may be represented as the survivors of immigrants to that date, less emigrants to that date, for the country.

$$P_{fb} = I * s - E$$

where the foreign-born population (P_{fb}) is set equal to the difference between immigrants (I), reduced by a survival rate (s), and emigrants (E). Foreign-born persons may include, for example, persons who immigrated a year earlier as infants, youth, or elderly persons, or elderly persons who immigrated four decades earlier in their youth or six decades earlier as children. The complementary group is the native population, who represent the survivors to the observation date of persons born in the country in question, plus any returning citizens who were born abroad, minus any natives who left the country before the observation date. Data pertaining to a specific year-of-immigration cohort (i.e., foreign-born persons arriving in the same year or group of years) or immigrant birth cohort (i.e., foreign-born persons born in the same year or group of years) define specific current segments of the foreign-born population (year-of-immigration group or birth cohort of immigrants).

and we learn from the Pilot Study that, within this population, 25% of the men and 7% of the women smoke as compared to 26% and 21%, respectively, of the native population.

Background Data on Immigration

According to the U.N. *International Migration Report* for 2006, 62 million migrants moved from less developed countries to more developed countries in 2005. Nearly as many, 61 million, moved from more developed countries to less developed countries. In general, the migrants from less developed countries are moving to improve their work situation. This is true also for many of those moving to less developed countries. Some in the less developed countries, however, moved because of civil strife and natural disasters. Such refugee migration is most common in Africa.

Because of population aging, the more developed countries are confronting a serious labor shortage and are turning to immigrants to ease the problem. In the last decade or so, United States, Germany, and Spain have been leading destinations of migrants from the less developed countries, attracted by the greater earning opportunities in these western countries. In 2005 38 million foreign-born persons lived in the United States, a greater number of immigrants than in any other country. The Russian Federation and Germany had more than 10 million immigrants each. Even the 13% of the United States population that were immigrants, however, does not compare with the 33% in the Persian Gulf States, where migrant workers have come from South Asia in vast numbers to build new giant centers of mid-east culture.

As a result of lax enforcement of border and visa controls and lax enforcement of business hiring of illegal workers, the United States now plays host to some 12 million illegal immigrants, according to the Pew Hispanic Center/Passel (2006). An estimated 56% of these are Mexicans and another 22% come from other Latin American countries.

Health Issues as Determinants and Consequences of Migration

I had considered dividing this discussion of the relation of international migration and health under the headings, health issues as determinants of migration and health issues as consequences of migration. However, a rigid distinction is difficult, if not impossible, to make. Sometimes the directional flow of cause and effect is opposite to that anticipated, is in both directions, or cannot be specified. Since migration is defined in terms of an origin and a destination, the health determinants and health consequences of migration may both be associated with a single migratory movement. These health correlates of migration may appear in a variety of forms – health status, health behavior and lifestyle, utilization of health services, differences in the predisposition to disease, and the availability of health resources–and are often linked with one another.

I illustrate the interplay of international migration and health with the following tale, where a health situation induces emigration from one country to another and an associated health situation results from the immigration to the other country. The migration of Jews to the United States and Israel from Kiev, Ukraine, in the former Soviet Union following the accident at the Chernobyl nuclear power plant in 1986 is the background for this joint role of migration as consequence of a health situation in the area of origin and a cause of a health situation in the area of destination (Glicksman 2005). Many Jews lived in Kiev near the Chernobyl plant as a result of their evacuation from other parts of the Soviet Union when the Germans invaded Russia during World War II. The environmental disaster in Chernobyl led to a second mass migration of these Jews, many to the United States and Israel. Health problems, already realized or anticipated, forced their departure even though many of the older Jews were comfortable with the system and were reluctant to leave the Soviet Union. In relocating to the United States, the immigrants were carrying the seeds of future illness and death from cancer. The physical effects were real, even though they could not be seen or felt by many of the immigrants for a considerable number of years after their arrival in their new residences abroad. By 2005, an excessive number of cases of cancer had been reported for the community of Russian Jewish immigrants in Philadelphia, PA, where many of the former Kiev emigrants had come to reside.

Legal Restrictions and Legal Preferences

Legal restrictions may influence the interrelation of health circumstances and immigration. One such restriction relates to the exclusion of prospective immigrants on the basis of their health status or presence of specific diseases. Another is the imposition of limits or preferences on the numbers of health providers who may enter a country and the setting of conditions under which they may enter. We hear little on these matters currently in the United States although there are federal laws bearing on both of them. Historically, the United States has placed many restrictions on the eligibility for entrance of aliens into the United States on the basis of health.[3] Few persons have been excluded or deported under these laws,

[3]The U.S. Immigration Act of 1882 prohibited the entry of "idiots and lunatics," and in 1891 Congress prohibited the entry of aliens if they suffered from "a loathsome or dangerous contagious disease." The Act of 1893 called for reporting of facts regarding the physical and mental health of alien arrivals at the United States, so as to improve the determination of admissibility according to the expanding list of grounds for exclusion. The 1903 law excluded "insane" persons, and the 1907 law excluded persons with mental and physical defects that could affect their ability to earn a living, and persons afflicted with tuberculosis. The 1917 law expanded the list of mental health conditions for which an alien could be excluded. Thereafter, the law was little concerned with setting health requirements for immigrants. In 1921 and 1924, the restrictions were based on country of origin, with quotas being assigned to various countries. The law of 1961 placed the decision on health conditions meriting exclusion of immigrants from the country in the hands of the U.S. Public Health Service.

however, except perhaps in the earlier periods of heavy immigration, 1901–1910 and 1911–1920. This category has not been reported separately by the U.S. Immigration and Naturalization Service since 1980; in the decade preceding 1980 the numbers were trivial.

The terms of the laws relating to the admission of foreign health providers into the United States have shifted over time depending on domestic conditions of supply and demand. Physicians and nurses have generally enjoyed the status of occupational preference groups in the rules for being admitted to the United States and achieving permanent residence.[4] For example, the Immigrant Nurses Relief Act of 1989 (INRA) created a temporary visa sub-category (H-1A) explicitly for foreign registered nurses in response to a widely acknowledged shortage in that profession. This law facilitates and streamlines the entry of new foreign nurses. At the same time, however, in an effort to restrain U.S. dependence on foreign nurses, INRA also required employers to actively develop, recruit, and retain U.S. nurses. To this end, U.S. employers of foreign nurses are required to certify to the U.S. Department of Labor that the foreign nurses to be employed are necessary to maintain health-care services in the United States and that the petitioning employers are taking steps to train, recruit, and retain U.S. nurses. Admissions were initiated on October 1, 1990 but began to slow off after October 1, 1995. A few additional registered nurses have been admitted from 2000 on with temporary visas (H-1C) under the Nursing Relief for Disadvantaged Areas Act.

Movement in Response to Foreign Health Programs

The health systems of some countries offer more "perks" to its residents than others. It may reasonably be postulated that, if a country has a more generous health system than its neighbor, this fact may serve as a motive for migrating to the former country from the latter one. United States provides free public health care to the indigent ("Medicaid") and some poor children ("SCHIP") and public health insurance to qualified persons 65 years and over ("Medicare"), but otherwise it has the least generous and least efficient system of the industrialized countries. About 15% of the population has no health insurance at all and much of the population is underinsured. Yet, few persons migrate from the United States to France, Great Britain, the Netherlands, Denmark, Sweden, or other western or northern European country that has a more comprehensive and efficient health system. It does not appear, therefore,

[4]Even before INRA (1989), the Quota Act of 1921 placed nurses and physicians on a nonquota basis. The 1976 Act placed restrictions on foreign medical school graduates coming to the United States for practice or training in medicine, but the Act of 1977 eased some of these restrictions and exempted alien physicians already in the United States from examination requirements. The 1989 Act (INRA) removed the numerical limitation on certain nurses employed in the United States from securing permanent resident status. The Act of 1996 again tightened restrictions regarding foreign physicians' ability to work in the United States, but they continued to fall in an occupational preference group.

that this factor is an important one in persuading people to move from the United States to another industrialized country with a better system. On the other hand, there is a modest movement from the United States to Canada, given its geographic propinquity, to take advantage of the latter's program of universal health insurance, and the generous welfare system of the United States, including free health services for the poor, may be a factor in encouraging movement to the United States from Mexico and other nearby less developed countries. Because of the lower costs of medical services in many foreign countries, thousands of U.S. residents go abroad each year temporarily to seek medical attention for their illnesses.

The essentially "open door" U.S. immigration policy for Latin America (in spite of closely guarded borders and modulating socioeconomic conditions in the United States), the poor living levels in these countries, their closeness to the United States, and the U.S. demand for cheap labor have encouraged the immigration to the United States of Mexicans and Central Americans, many of whom enter illegally. In spite of the border health inspections and the more robust condition of migrant youth seeking work, many of the newer immigrants suffer from numerous health conditions and are making heavy demands on U.S. public health facilities, particularly in the border states.

Effect on Area of Origin

When people leave a country in large numbers, the demographic, socioeconomic, and health characteristics of the population of the country of origin is probably affected because the migrants are likely to have characteristics different from those of the population left behind. Furthermore, if the migrants are mainly young workers, as is likely to be the case, the wealth and labor pool of the country of origin are reduced and the financial contribution of the migrants to existing public programs, such as the health insurance system and retirement programs, is normally terminated. In the following paragraphs, I elaborate on the consequences of these movements for the country of origin.

Loss of health personnel and revenue in areas of origin; subsidizing the areas of destination. The emigration of young workers who have been contributing to the health insurance system of the country of origin tends to weaken it financially. Because of their youth, they are likely to have made relatively few demands on the health system while contributing to its support. Such a movement also tends to effect a shift in age structure that makes it more difficult to finance the systems of health insurance and old-age and disability retirement.

Many foreign students come to the United States, mainly from the LDC, to pursue advanced programs in the health sciences, such as medicine, dentistry, nursing, epidemiology, and other health specialties, and then remain in the United States to fill vacant positions in this country. A far greater number of foreign graduates in the health sciences, however, come to the United States to work in the U.S. health-care system and stay indefinitely as compared with the number

of foreigners who come to the United States to study and stay in the country on completing their studies. Many foreign physicians pursue residencies in the United States and then set up practices here.

Between a third and a half of medical graduates in South Africa emigrate to more affluent countries, such as the United States, the United Kingdom, Australia, and Canada. Many come to the United States and acquire permanent resident status (OECD Observer 2004). As a result of such transfers, foreign physicians account for 20% of the physicians in practice in the United States (Pol and Thomas 2001). A small additional number of "immigrant" physicians are American citizens who received their medical education in a Caribbean or other Latin American country and return to the United States for advanced training and employment. A similarly high percentage of foreign medical graduates practice in Canada, the United Kingdom, and Australia as in the United States (Padareth et al. 2003).

Other types of health workers emigrate in large numbers from their countries after their training has been completed and even after years of experience. The emigration of nurses poses a serious ongoing problem for the Philippines and South Africa, particularly South Africa. The departure of many nurses from these countries places an excessive burden on their health systems and the nurses who remain behind, and may be responsible in part for their high rates of maternal and infant mortality. In South Africa the exodus of healthcare workers has depressed the ratios of healthcare personnel to population, especially in the more highly skilled health areas, and contributed to the maldistributions of health personnel between the public and private sectors and urban and rural areas. The health workers move from areas of low socioeconomic development to more highly developed areas, for example from lower to higher-income countries within southern Africa and from African countries to industrialized countries (Padareth et al. 2003). They are motivated to leave their poor work conditions (low remuneration, work-associated risks including diseases like HIV-AIDS, unrealistic work loads, poor infrastructure, etc.) and to escape political insecurity, a repressive political environment, personal insecurity, and deteriorating public services (WHO/Awases et al. 2004). The emigration situation is exacerbated when civil war, famine, and natural disaster in the country of origin encourage emigration and discourage return migration, even though these events make the need for health specialists in these countries more pressing.

The voluntary expatriation of new graduates in the health professions from the LDC contributes to the "brain-drain" and to the lack of trained health professionals in the LDC. It also hinders the efforts of these countries to upgrade their public health systems. In this way, a less developed country, which may have invested large sums of money in the rearing and training of health specialists, loses not only much needed skills but also some of their healthiest residents. The process can be self-renewing and cumulatively self-defeating. The inadequacy of the health systems in the LDC and their inability to secure sufficient aid from the richer countries to improve their systems induce a continuing flow of health professionals out of these countries into the more industrialized countries.

Subsidizing the countries of origin. Each year migrant-sending countries receive millions of dollars in remittances from migrant-receiving countries. The countries that are the source of these remittances are relatively affluent. Immigrant workers in the United States, the largest contributor group, sent $42 billion overseas in 2006. The remittance flow to the less developed countries is far greater than that from the United States alone, however. Formal remittances to the less developed countries amounted to $270 billion in 2007. The amount has been rising exponentially; in the mid-1990s, the total was only $60 billion. Most of this "overseas" money was sent to Mexico; India and China are other major recipients.

While the remitted money is removed from the U.S. economy and the economy of many other affluent countries, it contributes to the economic development of the poorer migrant-sending countries and, in effect, invests in their health systems as well as their other public programs. In this way the more affluent countries support the health programs of the less developed countries. It is possible that one consequence of remittances from the United States is to contribute to the creation of more efficient national health systems abroad than prevails in the United States.

Effect on Areas of Destination

Immigration has played a role both in augmenting the resources available for supporting the health systems of several industrial countries and in complicating their problems of managing the control of infectious diseases and restricting the movement of illegal drugs.

Some industrialized countries have actively recruited health workers for their own health systems from the LDC. To these workers they offer an improved quality of life and opportunities for specialization, improved pay, and advancement. The transfer is to the advantage of the industrialized countries to which the emigrants come. The movement serves to expand the corps of health specialists in the receiving countries while depriving the countries of origin, often countries of the less developed regions, of much needed trained health personnel. The foreign health workers fill positions in hospitals and other medical facilities in venues that might go unserviced. Depending on the economic situation in the receiving country, however, it may also deprive residents of jobs they seek or diminish their opportunities for advancement.

As stated, a substantial share of the health personnel in the United States was trained abroad. The flood of graduates of universities of LDC practicing in the United States has raised concerns about the quality of medical personnel. The feeling is widespread that these transplanted physicians and nurses are not as well prepared as graduates of U.S. universities, and that more screening and training should be required of the immigrants. This situation has also raised questions about the adequacy of U.S. educational facilities to meet the demand for training of U.S. residents in the health specialties. Individual states have different statutes regarding requirements for retraining and recertification of foreign-educated health personnel, and the stringency of these requirements depends in part on the supply/demand situation for such personnel in each state.

Effect of immigration on resources of receiving country. New immigrants from the LDC are generally people of modest means who seek to improve their economic condition by settling in a more affluent country and obtaining more lucrative employment than available in their home country. They tend to be less affluent than the resident population of the receiving country. For this reason we would expect them to incur more health problems and require more health services than the resident population. In addition, new immigrants are likely to be uninsured or, at best, underinsured, and to be less able to pay for needed health services on their own or even with health insurance. Hence, they tend to use emergency units in hospitals and public health facilities relatively more often than the resident population, and they frequently rely on public support to cover the costs. Local governments often complain about the heavy financial burden that immigrants place on local budgets by their use of public health facilities.

The United States and the industrial countries of northern and western Europe have had large influxes of immigrants in the last several decades. These immigrants have come from Asia, Africa, and Latin America, as well as southern Europe, unlike the immigrants of pre-World War II years, who originated almost wholly in southern and eastern Europe. New and heavy demands have been placed on the health systems in many local areas of the United States and other industrial destination countries because of the sheer numbers of the immigrants and the costs of their care. Moreover, the immigrants are likely to have distinctive health problems characteristic of their native countries. Immigration is considered responsible for a resurgence, in the western industrial countries, of some old diseases that had largely been eradicated, such as tuberculosis and whooping cough, and for the emergence of some new diseases that these countries have not previously experienced. The special burden that the destination areas have of managing their health systems because of immigration is exacerbated by the low income of the immigrants, their linguistic and other cultural differences, their lack of health insurance, and their dependence on public health facilities or private emergency clinics.

Comparative mortality of immigrants and natives. Immigration carries some health risks for new immigrants. Immigrants may enjoy an acquired immunity to some diseases characteristic of their native country but be unable to cope with the mutations of the same pathogenic microbes in the new country of residence. As a result, they become subject to some local communicable diseases. Migration may affect other aspects of environmental exposure to health risks. Commonly, migration leads to a change from a more healthful to a less healthful diet and, as a result, migrants may develop a greater risk of acquiring various chronic diseases. For example, Chinese male immigrants moving to the United States rapidly lose their protection against prostate cancer after having adopted the American diet. Similarly, acculturation to the United States has been associated with an increase in heart disease of Japanese immigrants (Keys 1980; Marmot and Syme 1976). Finally, migration is a stressful process in itself, typically triggering various adverse physiological responses that contribute to poor health.

Migrants from the LDC to the MDC may be expected to have lower mortality rates and be healthier than their compatriots who remain behind, but they would also be expected to have higher mortality rates and be less healthy than the residents of the country of immigration. This hypothesis can be tested by comparing the mortality and morbidity levels of the nonmigrants in the country of emigration, emigrants from this country, and the native population of the country of immigration. In fact, the mortality and other health consequences of such movements between several countries have been intensively studied. Some of the studies are described below.

In spite of these negative expectations, the evidence from several more developed countries supports the conclusion that adult migrants to these countries have lower mortality than the native population among whom they settle. The reader may recall the Hispanic "mortality paradox" (Chap. 7), which is based on strong evidence that the Hispanic population of the United States has a more favorable mortality record than the white non-Hispanic population of the United States. The Hispanic paradox is relevant in this connection because such a large share of the Hispanic population, especially the segment of Mexican origin, is foreign-born. Additional comparisons suggest that not only do Hispanics in the United States have lower mortality than the white non-Hispanic population but they also have lower mortality than the native white non-Hispanic population and the native Hispanic population of the United States.

As reported by Markides and Eschbach (2005), similar results were found by several analysts for all major "ethnic" groups in the United States (whites, Hispanics, blacks, Asians). Using death registration statistics, Sevak and Schmidt (2008) obtained supporting results for whites, blacks, and Hispanics, but not Asians. The Hispanic mortality advantage is most marked among Mexican-Americans, especially those at the older ages (Cho et al. 2004). To explain this special case, at least in part, Markides and Eschbach noted the role of selective migration; that is, the more healthy persons were self-selecting themselves as migrants from Mexico to the United States. This selective process is complemented by another selective one. Some elderly foreign-born Mexican-Americans return to Mexico after they become sick and remain there until their death. This selective emigration of unhealthy oldsters has been called the "salmon bias" (Palloni and Arias 2004). Inasmuch as this factor is not a statistical artifact, I prefer the label "unhealthy return-migration effect" following Deboosere and Gladeyne (2005). Palloni and Arias found only modest support for the "healthy migrant effect," as the first selective process has been dubbed. They reason that, even if it initially existed, it would diminish with increasing duration of residence in the United States as migrants assimilate the unhealthy behaviors and lifestyles of native Americans. They concluded, however, that the "salmon bias" was consistent with the pattern of mortality advantage of older foreign-born Mexican Americans.

The analysis by Sevak and Schmidt (2008) of both death registration data and data from the Health and Retirement Study (HRS) provides strong evidence that immigrants as a group have lower mortality rates than non-immigrants in the United States. Figure 10.1a, b show the mortality advantage of white immigrants over white non-immigrants, separately for males and females, in terms of 1-year death rates

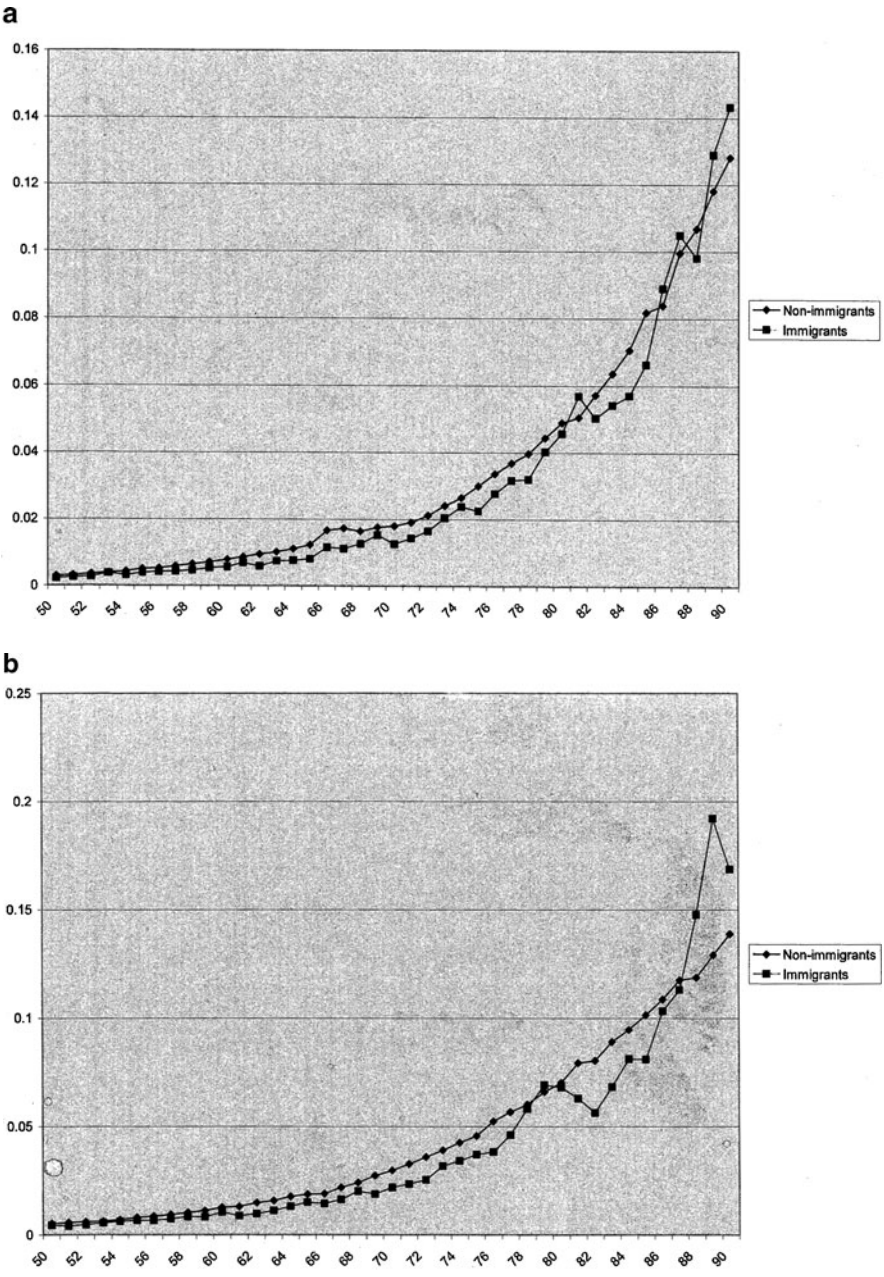


Fig. 10.1 (**a**) Death rates for white men, by nativity: United States, 2001 (Source: Sevak and Schmidt (2008)). Reprinted with permission of the University of Michigan Retirement Research Center; Primary source: U.S. NCHS Vital Statistics Detail Mortality Data and U.S. Census Bureau decennial census IPUMS. (**b**) Death rates for white women, by nativity: United States, 2001

from age 50 on, using death registration data. The analysis on the basis of HRS data showed that immigrant men have a 44% lower mortality hazard than non-immigrant men and immigrant women have a 41% lower mortality hazard. Most of this difference remains even after the analysis is controlled for differences in education, income, and health insurance coverage. Sevak and Schmidt also suggest that the finding could reflect selection effects; specifically, that the persons who leave their home country need greater resources to leave and hence may be the wealthier and healthier ones. Despite the differences in mortality, immigrants in the same HRS sample do not report higher subjective survival probabilities than the U.S. natives.

Foreign-born persons show lower death rates than their native counterparts in a number of countries. Similar results were found for Australia (Donovan et al. 1992) and Canada (Chen et al. 1996; McDonald and Kennedy 2004). Deboosere and Gladeyne (2005) obtained similar findings for immigrants and natives in Belgium, and reported the additional cases of Moroccans in France (Courbage and Khlat 1996), Turks in Germany (Razum et al. 1998), and Greeks in Australia (Kouris-Blazos 2002).

Effect of immigration and acculturation on the health situation. The evidence regarding the health effects of migration I have considered so far uses mortality as a measure of health. The question still remains whether migrants are healthier than natives as indicated by a measure of health other than death. It is reasonable to hypothesize that they would not be. Migrants from the LDC are generally of lower socioeconomic status and there is considerable evidence that lower socioeconomic status is associated with higher morbidity as well as higher mortality. It appears, however, that some migrant groups can have both relatively high life expectancy and relatively high morbidity (Uitenbroek and Verhoeff 2002). Low socioeconomic status can be associated with some health conditions that are not potentially lethal. Recall also the comparative health conditions of men and women, wherein women have higher life expectancy and higher morbidity, and men have lower life expectancy and lower morbidity (Chap. 7).

Two reports presenting selected prevalence measures of physical health status and limitations, for the United States, based on the National Health Interview Surveys for the 1980s–2003, show that immigrants have better health than the native population (U.S. NCHS/Dey and Lucas 2006; U.S. NCHS/Stephen et al. 1994). Despite limited access to health care, lack of a usual place of health care, lack of health insurance, and unfavorable socioeconomic circumstances, foreign-born adults in the ethnic/racial groups in the United States have more favorable health indicators than the natives of the same ethnic/racial group (Table 10.1). These indicators include self-assessed health, bed-disability days, and functional limitations (ADLs and IADLs). White, Hispanic, black, and Asian immigrant adults had significantly better health than their U.S.-born counterparts. These findings are consistent with other studies that have shown a considerable health advantage for Hispanic and black immigrants. Furthermore, the prevalence of diabetes, hypertension and cardiovascular diseases was higher among U.S.-born adults than the foreign-born of similar racial/ethnic groups (Table 10.2).

Table 10.1 Age-adjusted percents for selected health status measures and average number of bed days, for adults, by nativity and race: United States, 1998–2003

Health status measure											
		Self-assessed health[a]			Limitation in ADL[b] or IADL[c]			Annual bed days			Mental health
Nativity and race	Total	Excellent or very good	Good	Fair or poor	ADL	IADL	None	1–6 days	7 days or more	Average number of bed days	Serious psych. distress[d]
U.S.-born adults											
Total[e]	100.0	63.7	24.4	11.8	1.7	3.8	62.0	28.6	9.4	4.9	2.9
Non-Hispanic											
White	100.0	66.2	23.3	10.3	1.5	3.5	61.5	29.5	9.0	4.6	2.7
Black	100.0	50.6	29.3	20.0	3.0	5.9	64.4	24.3	11.3	6.7	3.3
Asian	100.0	68.9	21.6	9.1	0.7	2.9	58.3	34.3	7.3	3.4	1.4
Hispanic	100.0	54.9	26.2	18.8	2.4	4.4	63.3	26.1	10.6	5.7	4.4
Foreign-born adults[f]											
Total [e]	100.0	60.5	27.1	12.4	1.7	3.0	72.8	20.0	7.2	3.4	2.9
Non-Hispanic											
White	100.0	66.4	23.6	9.9	1.9	3.3	67.9	24.3	7.8	4.1	2.9
Black	100.0	64.4	23.4	12.0	1.5	3.0	72.7	20.3	7.0	2.7	1.9
Asian	100.0	65.0	25.0	10.0	1.6	2.6	71.4	21.2	7.4	3.4	1.7
Hispanic	100.0	54.3	29.8	15.8	1.7	3.0	75.0	17.8	7.2	3.4	3.6

Source: U.S. NCHS/Dey and Lucas (2006). Primary source: National Health Interview Survey, 1998–2003

Note: Estimates are age-adjusted to the year 2000 population as standard, using four age groups: 18–34 years, 35–44 years, 45–64 years, and 65 years and over

[a] Self-assessed health is based on the question in the survey, "Would you say [subject's] health is excellent, very good, good, fair, or poor?"

[b] Limitation in ADL is based on the question, "Because of a physical, mental, or emotional problem does [person] need the help of another person with personal care needs, such as eating, bathing, dressing, or getting around inside the house?"

[c] Limitation in IADL is based on the question, "Because of a physical, mental, or emotional problem does [person] need the help of another person with routine needs, such as doing household chores, doing necessary business, shopping, or getting around for other purposes

[d] Six questions were asked relating to the experience of psychological distress during the preceding 30 days. The response codes (0–4) of the six items for each person are summed to yield a scale with a 0–24 range. A value of 13 or more for this scale is used here to define severe psychological stress

[e] Persons of other races and unknown race and ethnicity are included in the total

[f] Foreign -born persons are defined as persons living in the United States who are not U.S. citizens by birth, including all naturalized citizens, legal permanent residents, persons on temporary long-term visas, and illegal aliens

Table 10.2 Age-adjusted percents of adults with selected risk factors or chronic diseases, by nativity and race: United States, 1998–2003

	Risk factor or chronic disease				
Nativity and race	Obesity[a]	Current smoking[b]	Diabetes[c]	Hypertension[c]	Cardiovascular diseases[c]
U.S.-born adults					
Total[d]	22.9	24.0	6.1	24.3	7.6
Non-Hispanic					
White	20.8	24.3	5.3	22.9	7.4
Black	33.1	23.8	10.1	34.7	8.3
Asian	10.7	15.2	5.6	23.9	6.6
Hispanic	29.8	20.1	10.8	24.5	7.6
Foreign-born adults[e]					
Total[d]	15.7	14.1	6.0	19.8	5.7
Non-Hispanic					
White	14.0	19.3	4.4	20.1	6.9
Black	18.5	8.4	8.2	26.7	3.9
Asian	5.1	11.9	5.6	18.9	4.5
Hispanic	21.0	14.1	7.4	19.3	5.2

Source: National Health Interview Survey, 1998–2003, Data are based on household interviews of a sample of the civilian noninstitutional population

Note: Estimates are age-adjusted to the year 2000 population as standard, using four age groups: 18–34 years, 35–44 years, 45–64 years, and 65 years and over

[a]Obesity is indicated by a body mass index (BMI) equal to or greater than 30.0

[b]Current smokers are those who have smoked at least 100 cigarettes in their lifetime and still smoke

[c]Respondents were asked if they had ever been told by a doctor or health professional that they have, respectively, diabetes or sugar diabetes, hypertension, or coronary artery disease, heart disease, heart attack, stroke, or angina

[d]Persons of other races and unknown race and ethnicity are included in the total

[e]Foreign-born persons are defined as persons living in the United States who are not U.S. citizens by birth, including all naturalized citizens, legal permanent residents, illegal aliens, and persons on temporary long-term visas

Using data from Canada's National Population Health Survey, Newbold (2006) compared health conditions among the immigrant and native populations, and found further support for the "healthy immigrant effect." The foreign born, especially recent arrivals, enjoy better health than their Canadian-born counterparts as measured by prevalence ratios for chronic diseases (e.g., heart disease, asthma, diabetes, arthritis). On the other hand, Klinthall (2008) found that Nordic immigrants to Sweden displayed higher rates of hospitalization from 1990 to 2001 than native Swedes but that immigrants of other ethnic backgrounds did not.

The prevalence of selected chronic diseases among adult white, black, and Asian immigrants does not significantly differ by the length of stay of these groups in the United States. Among adult Hispanic immigrants, however, the percentage with hypertension and cardiovascular diseases significantly increased as length of stay in the United States increased. This suggests that the conditions of "modern life" may contribute to an initial worsening of immigrant health. There is evidence that the health status of many immigrants deteriorates immediately after arrival in the

host country, particularly when the movement is from an underdeveloped country to a developed one. They eat less healthfully and exercise less. They may eat out more or cook their native foods in a less healthful style. After a short period in the host country, the eating habits, food choices, and general lifestyle of the immigrants begin to resemble those of the native population, and the health status of the immigrants tends to converge to that of the native population. Moreover, in the United States, the masses of new immigrants, both legal and illegal, experience a disproportionate share of health problems from life in crowded, deteriorated housing units, and as a result, new and old infectious diseases are spread. Accordingly, during this early stage of acculturation, the health of the immigrants tends to deteriorate. After several years of residence among the natives, the immigrants' socioeconomic status tends to improve and at the same time they experience an improvement in their health status.

In sum, many studies support the finding that immigrants to Western countries have better health than the native population of the destination country at the time of arrival. Numerous explanations have been offered for this phenomenon, including sociological, psychological, demographic, biological, economic, and statistical ones (Deboosere and Gladeyne 2005). The explanation that it is a statistical artifact, that is, due to errors in the data or survey responses, is no longer generally supported. One may seek to explain it as an effect of age distribution since most migrants are in their youth on arrival, when they are rather healthy and do not manifest symptoms of the leading lethal health conditions; but this explanation can be ruled out by age-adjustment of the data.

Markides (2001) and (Markides and Eschbach's (2005) explanation of this finding emphasizes the selective factors mentioned earlier. The immigrants are in fact healthier than their compatriots whom they leave behind, but they are also healthier than the natives of their host countries. This is so because the immigrants are self-selected as healthier; and they have a more healthy diet and lifestyle and suffer less from chronic degenerative diseases than either population. As a group of new residents, they remain healthier also because they "divest" themselves of the unhealthier, often older, members of their group, who have returned to their provenant countries for their final years. Markides and Eschbach note also that Western countries require health screenings of prospective immigrants; that the immigrants are usually prospective workers, who tend to be in relatively good health; and that immigrants tend to have a positive outlook and are motivated to improve their lives ("migrant hope effect"). Recall that U.S. law specifies certain health conditions as grounds for exclusion. In general, the evidence supports the view that the relations between migration and health described here are statistical realities that can be explained by a combination of factors including the selection factors noted.

Immigration and the urban-rural health crossover. Immigration to the United States was not always a positive factor in the country's health situation. The historical crossover of health levels in urban and rural areas a century ago in the United States recalls a time when immigration played a role in exacerbating a poor

Table 10.3 Infant mortality rates by population-size groups in the Birth Registration States for selected dates: 1915–1993

		Places			
Year	United States	Places 100,000+	Places 10,000+	Places under 10,000	Rural
1915	99.9	101.3	103.3		94.4[a]
1930	64.6	61.1	62.8	65.7	66.1
1940	47.0	39.3	42.3	53.4	50.1
		Metropolitan counties		Nonmetropolitan counties	
1950[b]	33.0	30.3		36.2	
1960[b]	27.0	26.0		28.7	
1993	8.4	8.3		8.6[c]	

Source: U.S. NCHS. (1947). *Vital Statistics Rates in the United States*, 1900–1940. By F. E. Linder & R. D. Grove. U.S. Government Printing Office; U.S. NCHS. (1968).*Vital Statistics Rates in the United States*, 1940–1960. By R. D. Grove & A. M. Hetzel, U.S. Government Printing Office; U.S. NCHS. (2002). *Vital Statistics of the United States*, 1993, U.S. Government Printing Office
Infant deaths per 1,000 births for 1915, 1930, 1940, and 1993; infant deaths per 1,000 population under one for 1950 and 1960
[a]Places under 10,000 and rural
[b]Infant death rate, or deaths under age one per 1,000 population under age one
[c]Urban places, 9.3; balance of nonmetropolitan counties, 8.4

health situation. In the nineteenth and early twentieth centuries health conditions in U.S. cities were notably worse than in the rural areas. The tremendous flood of immigrants from abroad to the large cities of the Northeast and Midwest contributed to this situation. At a time when these cities were being radically transformed by the Industrial Revolution and being filled by masses of immigrants from Europe, health conditions in the cities were deplorable. Crowding, poor housing, poor sanitation facilities, unclean water, and other indicators of poor public health contributed to the high rate of infectious diseases and frequent epidemics in the cities, including the influenza epidemic of 1918. With the improvements in urban public health – installation of modern sanitation systems, filtration of the water supply, a safer and more plentiful food supply, improved housing conditions – aided in part by reduced crowding and the slowing of the immigration tide, the situation reversed itself (Cutler and Miller 2005). The data on infant mortality shown in Table 10.3 reflect the changes before and after the 1920s in the relative health situation of the urban and rural areas of the United States. In more recent decades the health situation in urban and rural areas has been converging.

Role in the transmission of diseases and the transfer of illegal drugs. The possibility of the rapid spread of infectious diseases worldwide has increased vastly from the days when it took an oceangoing vessel two months to cross the Atlantic Ocean to the present when it takes an airliner less than six hours to cross the Atlantic Ocean. In recent times, the facile movement of people in intercontinental travel has resulted in the rapid transfer of several infectious diseases from Africa, Asia, and Latin America to North America. Refugee movements and refugee camps are ready breeders and spreaders of infectious diseases.

Some diseases imported from abroad are rare in the United States and are responsible for only one-time limited outbreaks that required an epidemiological sleuth to unravel. For example, Peter Schantz, an epidemiologist at the Centers for Disease Control and Prevention, uncovered the mystery of an attack of neurocysticercosis on a group of Orthodox Jews in Brooklyn, N.Y., in the early 1990s. This disease causes seizures in the victims as a result of a brain infection received from the larvae of the pork tapeworm. This parasite is not endemic to the United States and Orthodox Jews never eat pork. Schantz determined that the victims contracted the disease from their Latin American domestic help, who handled the preparation of their food. The disease was passed along as a result of the poor personal hygiene of the domestic workers, who were carrying the infection from their native countries to the homes of their employers in the United States.

There are several well-documented cases of the decimation of indigenous populations as a result of their exposure to the diseases of invaders, explorers, or colonists. The latter brought diseases with them for which the indigenous population lacked an acquired immunity. Among the more notorious cases are the devastation of the indigenous Hawaiian population in the 18th century after the arrival in Hawaii of Captain James Cook, the British explorer, and the sharp decline of the American Indian population as a result of imported diseases after the colonization of the North American continent by the British, French, and Spanish explorers and colonizers in the seventeenth, eighteenth, and nineteenth centuries. Additional details on the decimation of a few American Indian tribes in this way are presented in a later note.

Finally, I note the special role of immigration in the illegal drug trade to the United States, with its adverse consequences for the health of the U.S. population. In particular, the massive immigration across the Mexican border into the United States in the last few decades has brought in most of the illegal drugs that are imported into the United States. The United States has been flooded with a huge and continuous supply of illegal drugs delivered by illegal Mexican "circular" migrants, who move back and forth at will over the U.S. borders with contraband drugs. While U.S. residents provide the demand for the illegal drugs and U.S. government agencies deal with the cases of drug abuse that result, the health consequences for the U.S. population are considerable though they are preventable or at least treatable. It should also be noted that another segment of the illegal migration serves the U.S. health industry by working as menial labor in hospitals, hospices, and nursing homes.

Internal Migration

Sources and Limitations of U.S. Data

The problems of securing data on the numbers and characteristics of internal migrants are simpler than those associated with securing such data on international migrants. The data are not only more accessible and abundant but their quality

is superior. In the United States the decennial census, the Current Population Survey (CPS), and the American Community Survey (ACS) provide in combination continuing information on the characteristics of inmigrants, outmigrants, and nonmigrants for states and selected areas within states. For example, the U.S. decennial census has questions on place of residence 5 years earlier and state of birth as well as questions on demographic, socioeconomic, and health characteristics, so that one can develop data on the demographic, socioeconomic, and health characteristics of persons who recently moved between states or moved from their state of birth to another state. The questions on health relate to disability of various types – sensory disability, physical disability, mental disability, self-care disability, "personal-business" disability, and employment disability. Disability status reported in censuses, the ACS, or the CPS may be compiled according to migration status for various geographic units, particularly with the public use microdata sample (PUMS) files. Special tabulations from PUMS could be made on the mobility of persons with various types of disability. The data are also especially useful for the study of older persons who move to and from retirement areas in relation to whether or not they have a disability.

The geographic detail is more restricted in the CPS than the ACS and the census because of sample size and design. The American Community Survey is exceptional because it is a "rolling sample;" that is, it is in the field continuously and is designed to produce data to cover places of 65,000 persons or more each year.

The census and survey data are subject to various limitations. The migrant data from these sources understate the actual volume of internal migration in the designated period because they exclude persons who died during the migration period, migrants who returned to their original residence, and the second and higher moves of persons who moved more than once.[5] As mentioned, the health data relate only to certain types of disability status. Since disability status is known only at the end of the mobility period, there is no way of determining the temporal relation between migration and the onset of disability and, hence, the possible cause-and-effect relation between these two events.

Various other sources provide data for investigating the health situation of interstate migrants. These include the vital statistics registration systems for the states, Federal and state records of health conditions for which health coverage is

[5] According to the U.S. classification system, inmigrants to a state represent those persons resident in the state at a given date who resided in another state at an earlier date (e.g., 5 years earlier). Nonmigrants are those who resided in the same state at the earlier date as at the census or survey date. Persons living abroad at the earlier date, either citizens or aliens, are treated separately from inmigrants or nonmigrants. Inasmuch as the migration data represent survivors at the end of the period of persons who were alive at the beginning of the period and identify only the initial and terminal residences of each person, only one net move, if any, is recorded for survivors during the period. Hence, the data understate actual events by the number of migrants who died during the period and by the number of multiple moves during the period. For example, a return to the original state of residence would be recorded as nonmigration and "stage" migrants would be counted only once. Children born during the period are excluded, as are persons who have departed for a residence abroad. The same limitations apply to migration data derived from state-of-birth data.

provided by the Federal or state government (e.g., Medicare and Medicaid records), and the administrative records of certain health providers such as hospitals and nursing homes. The standard state death certificate carries a question on place of birth of the decedent (city, state, or foreign country). Hence, the comparative mortality of persons according to migration between state or country of birth and last state of residence and according to specific migration streams, cross classified by age and other characteristics, can be tabulated from the public use sample file.

The questionnaire of the National Health Interview Survey (NHIS) carries several "migration" questions: State or country of birth, years lived in the present state of residence, and if born abroad, years lived in the United States. These items can be cross-classified with the health and other data on the questionnaire to obtain information on the health status of persons who moved from specific states of birth and on health status according to length of residence in the current state. Data on health status by nativity from the NHIS were referred to earlier in the chapter.

Health Issues as Determinants and Consequences of Migration

Health as a Reason for Internal Migration

Sample surveys concerning internal migration occasionally ask about reasons for moving. This way of securing the data is simpler and likely to provide more accurate data than trying to infer reasons for migration from analyses of migration differentials or from the different characteristics of the area of origin and the area of destination. In employing a direct question on reasons for migration, we do not have to make comparisons of health conditions among nonmigrants and migrants in order to determine whether or not health is an important reason for the change of residence. On the other hand, the question is a subjective one and hence the results are subject to the biases that commonly affect such responses. Moreover, the classification scheme in the question – for example, the decision whether to list only a few primary reasons, clusters of reasons, or detailed categories – can affect the interpretation of the results.

The U.S. Current Population Survey (CPS) has asked reason for moving, of those who changed usual residence in the previous year, annually for several years, and health has been listed as a separate category. Because of the importance of job-related and housing-related reasons at many ages, these reasons are usually distinguished from family-related reasons and other personal or social reasons, including health or wanting a change of climate. In the March 2005 CPS, health was infrequently reported as the reason for moving. In the 1-year period between March 2004 and March 2005, only 1.6% of the intercounty movers reported health as their main reason for moving. Intra-country movers reported health as their main reason even less frequently (0.8%). (U.S. Census Bureau, www.census.gov; Accessed October 2008).

Table 10.4 Health as a reason for moving, for movers in the United States, by age: 2001–2002 and 2004–2005 (Civilian noninstitutional population. Numbers in thousands)

Year	Number of movers		Percent of total		Health as percent of total movers
	Total	Health as reason	Total movers	Health as reason	
2004–2005					
Age					
1+	39,888	633	100.0	100.0	1.6
45–64	5,289	164	13.3	25.9	3.1
65+	1,484	229	3.7	36.2	15.4
65–74	861	75	2.2	11.8	8.7
75+	623	154	1.6	24.3	24.7
2000–2001					
Age					
1+	41,111	510	100.0	100.0	1.2
45–64	5,189	122	16.0	23.9	2.4
65+	1,321	144	3.2	28.2	10.9
65–74	795	51	1.9	10.0	6.4
75+	526	93	1.3	18.2	17.8

Source: U.S. Census Bureau. (2005) *Current population survey. Annual social and economic supplements, 2002.* Available at www:census.gov

Age patterns. Without considering age detail, such results are not unexpected. It is reasonable to surmise, and the CPS confirms, that at the more advanced ages, health is a much more common reason for moving than at the younger ages. With increasing age the share of persons reporting health as their reason for moving rises steadily (Table 10.4). By ages 75 years and over between one-fifth and one-quarter of all movers report health as their reason for moving. Another way of expressing this fact is that a much higher percentage of persons who report health as their reason for moving are at the older ages than of all persons who reported that they moved. For example, at ages 75 years and over in the period 2004–2005, the percentage of health-motivated movers was 24 as compared with 1.6 for all movers.

Movement for health reasons at the advanced ages involves shifts to locations near next of kin, to nursing homes, or to assisted living centers. Hence, some of the "health" movers may have reported "other family reasons." According to the Longitudinal Study of Aging (LSOA) of the National Health Interview Survey (relating to persons 70 years and older in 1984), approximately one-fifth of the respondents who moved between 1984 and 1990 reported that they moved in order to live closer to or with their family, another fifth reported that they moved in order to enjoy better weather or live in a better neighborhood, and one-fifth reported that they moved because of their own or their spouse's flagging health (Bean et al. 1994).

On the basis of the 1984–1986 wave of the LSOA, Zimmerman et al. (1993) found that individuals aged 70 years and over who had suffered moderate declines in IADL are more likely to change their residence than those who had not, and

Worobey and Angel (1990) found that unmarried older individuals who experience a substantial decline in functional capacity are more likely to enter a nursing home than those who suffer no decline. On the other hand, most older single persons continue to live alone even if their health is not good (Bean et al. 1994).

Health can be assumed to enter into some other migration decisions as an associated reason even though it is not explicitly listed. The 1974–1976 Annual Housing Surveys, which asked questions on migration, did not offer health as a reason for migration, but a large share of older women reported in the surveys "wanting to be closer to relatives" (43% of women 55 and over), which could include health as an underlying reason. A response of climate and recreational amenities may incorporate health as a secondary reason. Older persons may be drawn to an area by the availability of health facilities and health-care services. On the other hand, good health can be an assumed condition of migration, as when healthy older retired workers, especially those who want to supplement fixed retirement incomes, are attracted to a local labor market area offering part-time jobs.

The CPS shows a slight spike in the migration rate at ages 85 and over that interrupts its general decline with age after ages 20–24. For example, in the year 1999–2000 4.3% of persons at ages 65–84 moved while 4.7% of persons aged 85 and over moved. Increased mobility at the highest ages of life has been observed over a number of decades. Most persons at advanced ages (*i.e.*, 85 and over) living in group quarters at the time of the 2000 census had moved to these residences, presumably nursing homes and assisted living quarters, in the previous 5 years. While long-distance migration of young elderly persons (*i.e.*, aged 65–74) is usually motivated by the quest for residential amenities and a recreational lifestyle, migration of the old elderly is usually motivated by the need for physical assistance and support, first by family caregivers and then by formal community caregivers. Loss of a spouse accounts in part for the increase in moves at the advanced ages, which often include relocation of persons to institutional residences.

Few international comparisons have been carried out on age patterns of internal migration. In one such study Long and Boertlein(ca. 1975) found an upturn in the percent moving at ages 75 and over as compared with ages 70–74, for Great Britain and the United States but not for Japan. These differences may reflect differences between countries in the tendency for members of families to live together or apart when disability requires supportive care.

Retirement Migration

As we saw, while migration at the younger ages is motivated by educational, labor-market, marital, and housing reasons, migration at the older years is largely motivated by a change in environment, accommodation to changes in income and health, and a desire to be near other family members. Patterns of interstate migration at the older ages, even without specific data on the health of the movers, suggest some health implications and hypotheses regarding the relation of health and migration. At the early ages of retirement, those who can afford it and are

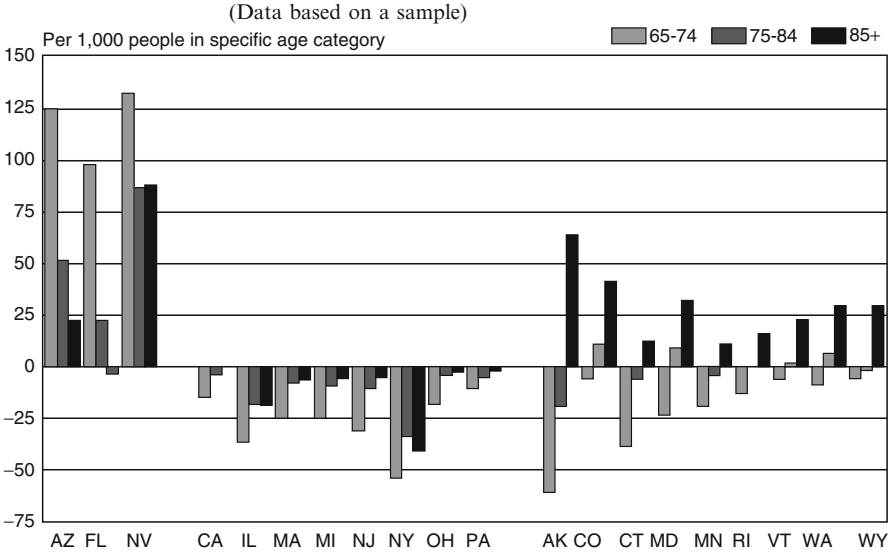


Fig. 10.2 Net migration rates for the population 65 years and over, by age, for selected states: 1995–2000 (Data based on a sample) (Source: U.S. Census Bureau, Census 2000. www.census.gov)

sufficiently healthy move to new homes, often in retirement communities in distant areas, with attractive climates and amenities (Biggar 1980; Speare and Meyer 1988). When the migrants get older and health and income decline, they tend to return to their original communities to be near their relatives and, then, when community care is required, to relocate to a nursing home (Litwak and Longino 1987). The less affluent forego retirement migration and move short distances, if at all, just after the early retirement ages, in order to remain close to, or to move closer to, their relatives, This serves them well when their health begins to decline. Then, when their relatives can no longer care for them, they tend to relocate to a nursing home.

This ideal-typical sequence suggests a considerable homogeneity among the elderly in patterns of residential change, motivations for moving, and health conditions, but there is actually much heterogeneity among them in these respects (Angel et al. 1992; Worobey and Angel 1990). Some moves to retirement centers are made by older persons who are poor. Some who move for "health" reasons move in anticipation of the deterioration of their health. The broader picture is that only about 5% of the elderly move for any reason in a year in the United States and only about 1% move to another state.

A striking illustration of the interconnections between health and internal migration is offered by the retirement migration between the Northeast region and the East North Central States, on the one hand, and the South Atlantic States, on the other, particularly between New York and Florida. (See Fig. 10.2) Many older persons from New York and other states of the Northeast periodically spend their vacations

in Florida and other states of the South Atlantic Division because of the recreational amenities and warm climate of the latter states; others maintain a second residence in these states ("snowbirds"); and still others move south to establish a single, new, permanent residence. Typically on moving to Florida, retirement migrants are in relatively good health, even though their choice of a warmer climate may suggest a desire to improve their health. As they grow older, their health tends to deteriorate. Then, many of them choose to return to New York or other Northeastern state of origin as they anticipate or reach physical dependency, particularly if their children are living in these "home" areas.

The American Community Survey showed 68,163 elderly inmigrants to Florida and 10,154 elderly inmigrants to New York during the period 2004–2005, and hence a net exchange of 58,009 migrants between New York and Florida. Because of the outmigration of many healthy persons of older age to Florida, New York is adversely affected by a burden of less healthy older persons who remain in place. Several years later the relatively unhealthy migrants in Florida return to New York and the other states of origin in the Northeast, and the latter states are again adversely affected. Florida gains at both ends by receiving a supply of relatively healthy older in-migrants and discharging a supply of relatively unhealthy older out-migrants. Thus, the migration "mixing bowl" keeps the Northeast states, particularly New York, at a continuing disadvantage.

In addition to having an impact on the health status of the origin and destination populations, these movements have an impact on their requirements for health services. The exchanges can require expanded health resources in both the areas of origin and destination. Such population movements and exchanges directly affect the demographic characteristics of each area, including population numbers, age-sex distributions, and other characteristics. The age distributions in the states of origin and destination are changed as a result of the differences in size and age distribution among the inmigrants, the outmigrants, and the nonmigrants. An increase in the population alone tends to cause an increase in demand on the health services of a community, but inmigration of the elderly imposes an especially heavy burden on a community because the new group is subject to much greater risks of illness than persons of younger age.

Rural–Urban Migration

In the more developed countries it is typical for young men and women to migrate from rural areas and small towns, particularly rural-farm areas, to metropolitan areas, and for the older population to remain behind (*i.e.*, "age in place") in the rural areas and small towns. This is characteristic in the United States, especially the Great Plains states, and many countries of Europe. As a result, the rural areas often have a disproportionately small share of young adults and a disproportionately large share of older persons. Because of these migration-related, complementary changes in the age compositions of urban and rural populations, the age- and cause-patterns of deaths and diseases of the two residence areas have been diverging. The relative

numbers of elderly persons and young adult persons in many rural counties of the United States have shifted so much that the number of deaths exceeds the number of births in these counties. Nationally there are about 1.7 births for every death in the United States. Because of the migration patterns described, the cause-distribution of deaths and diseases in rural areas has shifted sharply toward the chronic causes of older age, such as heart disease, cancer, stroke, osteoporosis, and osteoarthritis, and away from the acute diseases and other health conditions of early life, such as infectious diseases, pediatric and obstetric conditions, and deaths due to accidents and violence. Urban areas reflect the complementary pattern, with a relative lag in the rise of the age of deaths and shifts in the cause-pattern of death and illness to those of youth and the early adult ages, particularly accidents and violence.

The need for special and expanded health services in the rural areas is great, in the face of the large share of chronic diseases and the very limited resources of these areas. In effect, the flight of youth and the aging-in-place of elders in rural areas have precipitated a crisis in the health situation of many rural communities. They compromise the ability of the local populations to support necessary health services and facilities. The problem is exacerbated by the fact that these are usually low income areas. The population base for economic support is usually low and has been diminished further by the age redistribution described. At the same time the per capita demand for health services has become exceedingly high. Because of the differences in the living conditions and resources of urban and rural areas, the selective migration between them, and the geographic isolation of rural areas, inspite of a more healthful natural environment, health conditions in the rural areas are not better than those in urban areas. The difference is suggested by a comparison of the infant mortality rates in metropolitan counties and nonmetropolitan counties; the figures were 8.3 and 8.6 in 1993.

Urbanization in Less Developed Areas

Currently about half the population of the world and over 80% of the population of the United States lives in urban areas (United Nations 2007). All the regions of the world, particularly the less developed areas, are rapidly urbanizing, and according to the United Nations, by 2050 70% of the population of the world and 90% of the U.S. population will be living in urban areas. In this period nearly all the urban growth will occur in the less developed regions. South Central Asia and Africa are expected to triple their urban populations, but the Latin American/Caribbean region will have the greatest percent urban – 89% – of all the less developed regions. It is notable, however, that rural-to-urban migration now accounts for only about one-third of urban growth in the world; most urban growth is due to natural increase of the urban population and reclassification of areas from rural to urban rather than migration of rural persons to the cities.

To many, rural areas would seem to have a distinct advantage with respect to health; it is taken to be a physical environment that provides fresh air, tranquility, pure spring water, and an abundant supply of fresh food. In reality, rural areas of the less developed regions usually lack clean water and an abundant supply of food,

and are besieged by infectious and parasitic vectors of disease. On the other hand, urban areas have the potential advantage of population concentrations that make delivery of health care, clean water, and food, along with the control of disease, more feasible and efficient. However, commonly rapid urbanization in the less developed countries brings urban slums and desperately poor residents who lack the so-called urban amenities – clean water, sanitary facilities, and health services. In fact, rapid urbanization has brought with it serious health problems both in the rural and urban areas. Rural peasant children and youth migrate to the large cities, often the primate cities, and settle in impoverished urban slums. As in the more developed countries, the rural areas, already poor, become further impoverished by the outmigration of their youth and the concentration of old people.

Few studies exist on the health attributes of internal migrants in the less developed countries, especially studies based on longitudinal survey data. One recent study, that by Lu (2008), explores whether pre-migration health status affects the likelihood of internal migration in Indonesia by comparing persons who do and persons who do not move. His results confirm the "healthy migrant" hypothesis that migrants in Indonesia tend to be selected with respect to better health.

Health services are of poor quality and sparse, or nonexistent, in the rural areas of the less developed countries. The expansion of urban slum settlements complicates the demands made on already scarce health resources in the cities. The density of the slum population, poor sanitation, lack of running water, makeshift housing, and close contact with animals contribute to the rapid spread of infectious diseases, particularly those are transmitted from animals to humans. In addition, this migration trend has fueled the spread of HIV/AIDS in the cities of the LDC (Nguyen 2004).

Influence of Geographic Area of Origin for Lifetime Migrants

In Chap. 6, I presented evidence showing the strong influence of the health experiences in childhood and youth on health and mortality in later life. I carry this "argument" further by posing the question whether the geographic area of birth and residence in early life has an influence on the health in later life of persons who moved to another geographic area. We might reasonably expect this to be the case. Social and environmental conditions differ from state to state. Census data comparing populations distributed according to state of residence and state of birth indicate the share of the current population that bears early life imprints imposed by their being reared in a state different from the state of current residence. Large segments of the population of some states of birth no longer live in these states (Table 10.5). In 10 of the states (and the District of Columbia) less than half the population in 2000 remained in its state of birth and in 35 of the states less than two-thirds of the population remained in its state of birth. In any given year the shares of state populations not living in their states of birth tend to increase with increasing age.

Table 10.5 Percent of population of the states living in state of birth, by rank from highest to lowest: 2000

State	Percent	State	Percent
Louisiana	78.4	Arkansas	61.4
Pennsylvania	78.3	Kansas	61.3
Michigan	74.6	Texas	61.3
Mississippi	74.3	South Carolina	59.2
Ohio	74.2	Georgia	58.9
Iowa	74.0	Connecticut	57.1
Alabama	73.9	Montana	56.8
West Virginia	73.7	Hawaii	56.6
Kentucky	73.5	Vermont	53.3
North Dakota	72.6	New Jersey	52.9
Indiana	71.7	New Mexico	51.3
Wisconsin	71.6	Virginia	50.8
Minnesota	69.9	California	50.6
Maine	67.8	Delaware	50.0
Nebraska	67.4	Idaho	49.9
South Dakota	66.6	Maryland	49.0
Missouri	66.2	Washington	47.5
Illinois	65.8	Oregon	45.8
Illinois	65.8	Dist. of Columbia	43.5
New York	65.6	New Hampshire	43.4
Massachusetts	65.5	Wyoming	41.4
North Carolina	63.5	Colorado	41.1
Utah	63.3	Alaska	37.6
Tennessee	63.0	Arizona	33.9
Rhode Island	62.6	Nevada	23.0
Oklahoma	62.3		

Source: U.S. Census Bureau Census (2000). Supplementary Survey. Available at www.census.gov

The exact health effects of such an environmental shift are not known, but we may speculate about them. It is reasonable to surmise that marked changes over the life course in residential milieu influence the prospects for the quality of one's older years, including one's health and longevity. Geographic areas differ, for example, in the quality of their environments, the amount of air pollution, the quality of their water, and in the distribution of their populations according to education, income, and occupational status.

While life expectations at birth vary moderately over the states of the United States, the variation is small when the figures for each race are considered separately (Table 7.10). Even if health conditions in early life had a strong influence on longevity, once an individual's race is taken into account, the variation in life expectancy among the states is so small that the state of birth could have little differential influence on longevity in the present state of residence. Examination of a historical series of life expectancies at birth, 1969–1971 to 1989–1991, reveals a substantial convergence over this period and increasingly less variation when the data are racially disaggregated.

To explain variations in the incidence of schizophrenia, a number of studies in the 1990s implicate urban birthplace and migrant status as risk factors in incurring the diseases. These studies use subjects from Denmark, Sweden, the Netherlands, and the state of California. In one study of Danes, being born in Copenhagen was associated with a 2.5 fold greater risk of incurring schizophrenia than being born in rural areas. Residence in urban areas through childhood was most risky. In another study using participants in the Kaiser Permanente health plan, conducted by the Mailman School of Public Health, Columbia University, those participants born in the more densely populated neighborhoods of Oakland County have a twofold to threefold greater risk of incurring schizophrenia, as determined by hospital admissions, than persons born in less dense areas of the county, holding race constant (Bresnahan et al. 2007). The difference could be explained by the fact that residents of overcrowded city neighborhoods are more exposed to industrial waste, infectious conditions, and other stresses, and have less access to social supports that would make dealing with the illness more manageable, than those living in the remainder of the county.

In sum, there is limited direct evidences to support the hypothesis of a differential effect of area of birth and early rearing on later life health.

Temporary, Seasonal, and Circular Migration

Temporary and Seasonal Migration

Migration is normally defined as involving a change of usual residence, but many moves are of a different sort. For example, there are vacation trips, business trips, and repeated moves between two residences at different seasons of the year. Changes of usual residence have the characteristic of "permanency" although some people may change their usual residence several times in a lifetime. Changes of the other types are usually short-term and have a fixed duration. They may last a week, a month, or 6 months, and may be one-time events or be repeated from year to year. Depending on the circumstances, these types of moves are temporary or seasonal. One type of temporary and seasonal move that is of considerable interest is the move of elderly persons to vacation and seasonal homes.

Given the considerable volume of temporary and seasonal migration and the concentration of the migration streams in relatively few areas, their impact on the areas of origin and the areas of destination can be considerable. The movers have a tremendous impact on the volume and flow of a wide range of goods and services, including especially health services, health facilities, and recreational facilities, as well as motor vehicle traffic, retail sales, and the supply of housing. Effective government planning and budgeting for many areas receiving temporary and seasonal migrants cannot be done without taking account of their numbers and characteristics. However, data of this kind have not been generally available. This situation makes it difficult to analyze the impact of such migration on the health situation in any area.

Smith and House (2006) have made a pioneering effort to measure seasonal ("snowbirds" and "sunbirds") and temporary migration in Florida through survey methods. Their survey, which was restricted to an analysis of stays of one month or more, was designed specifically for this purpose and secured information on numbers, duration of stay, health status, the determinants of migration, and some other demographic characteristics. A telephone survey was conducted using list-assisted random digit dialing, and from the survey a sample of persons 55 and over was selected each month between September 2000 and December 2003. Some 4.7% of these reported that Florida was not their usual place of residence; these are the so-called snowbirds. Also, some 12% of Florida residents reported that they had spent at least one month outside the state; these are the so-called sunbirds. The remainder, the "stayers," did not leave their usual residences in Florida for one month or more. Information on the role of health and other possible determinants of migration was inferred from an analysis by logistic regression of the relation of the demographic variables.

Snowbirds and sunbirds tend to be more similar to each other than to stayers with respect to health. Snowbirds were older and healthier than sunbirds and both groups were older and healthier than stayers. More than 63% of snowbirds rated their health very good or excellent, compared with 55% of sunbirds and 49% of stayers. Several other studies have found elderly temporary migrants to be healthier than the elderly population as a whole (Monahan and Greene 1982; Sullivan 1985).

The great majority of the snowbirds came from the Northeast or Midwest, but only about half of the sunbirds went to these areas for their temporary stays. Snowbirds flocked to Florida during the winter months; few came during summer. Almost 83% of the snowbirds reported that they went to Florida because of its warm winters; all other reasons were of minor importance. Poor health had a negative effect on the tendency to move to Florida. Sunbirds left Florida mostly during summer. Only a small percentage of the sunbirds said they left their home for primarily weather-related reasons. Yet they left mostly during the intolerably hot summers! The results of Smith and House are consistent with studies of elderly permanent migrants as well, and reasonable, given weather conditions in Florida.

Circular Migration

Circular migration refers to the repeated movement of individuals or groups between two or more locations and a return to the original place of usual residence during some specified period. It does not usually involve a "permanent" change to another usual residence, but it may seem to do so, as when, after several temporary changes of residence, there is a return to the original place of usual residence. In contrast, internal migration and international migration normally refer to a single change of usual residence during a specified period.

Potential Role in Spread of Disease

The path of migration, whether temporary, circular, or "permanent," can reflect the path of an infectious disease. Here are two rather different illustrations of the role of circular migration in the spread of disease. Circular migration of truck drivers in India and other parts of South Asia is a factor in the spread of the HIV/AIDS epidemic. The truck drivers pick up HIV from female sex workers in cities on their routes, transport it back to their homes, and infect their wives. The sex workers maintain that the men are unwilling to use condoms even though the women are willing to supply them. This process has contributed to the spread of HIV/AIDS in neighboring countries that formerly were largely free of the disease since the truck drivers often travel across borders.

The sharp decline in the American Indian population from smallpox mortality after the colonization of the North American continent by the British, French, and Spanish explorers and colonizers in the seventeenth, eighteenth, and nineteenth century provides another example of how circular migration contributed to the spread of a lethal disease.[6] The fur traders traveled by steamboat during the early decades of the 19th century from one trading post to another along the upper Missouri River. This circular migration brought them into periodic contact with the Mandan and Hidatsa tribes. In 1837, several passengers and crewmen on the steamboat St. Peters, who were infected by smallpox, transmitted the disease to the Mandans at the St. Louis trading post. This event triggered a catastrophic epidemic of the disease on the North American continent. As many as 20,000 Indians perished across the high plains, including 90% of the Mandans. The Mandans and other tribes, such as the Ojibwa, Pawnee, and Arikara, were devastated in a previous sweep of smallpox across the high plains from Mexico in the late eighteenth century. By the 1830s their acquired immunity to the disease had largely dissipated and they were not yet inoculated with the new vaccine.

Migration and Mental Health

Few studies of migrants have dealt empirically with the mental health of migrants. The studies of the past years give inconsistent indications of the effect of hypothesized risk factors on the mental health of migrants. A variety of studies have been carried out in recent decades to test the hypothesis that migration and mental health are linked, that is, that migrants suffer to a greater degree than nonmigrants from mental disorders in the area of destination. These studies indicate that migration may be a factor in the onset and progress of mental illness in some circumstances but not in others.

---

[6]This description of events is based on Landon Y. Jones. 2005. Tribal fever. *Smithsonian*, May 2005.

Internal migration has long been considered a stressful personal experience because it tends to be associated with a life-altering change such as divorce, change or loss of a job, settling in a new home and neighborhood, and attending a new school. Immigration would seem to be more emotionally stressful than internal migration and, in fact, studies trying to link immigration and mental health have been more common. The mental stresses associated with immigration are caused by all or some of the following factors: Culture conflict and language changes, social isolation and sense of loss, uprooting, change of social status and social relations, and socioeconomic discrimination. It is important to distinguish peaceful, voluntary migration from migration under pressure of extreme poverty, war, civil strife, persecution, and famine. Even under the former circumstances, the migrant may experience unemployment, racism or ethnic discrimination, cultural and language differences, difficulties in securing a job appropriate for his/her qualifications, and legal restrictions imposed on new immigrants. The circumstances of the involuntary international migrant are likely to be far more dire and combine both the intolerable conditions in the home country with the types of difficulties just noted in the host country.

One problem in the conduct of such studies is the difficulty of agreeing on the indicators of mental illness. Each indicator has some serious flaw(s) and the results are based usually on subjective responses that produce fuzzy data. In spite of the expectation that immigrants would fare poorly from the immigration experience, the studies have both supported and rejected the fundamental hypothesis. This may be because other factors in addition to the move itself are involved: The characteristics of the migrant, the circumstances surrounding migration, and the characteristics of the places of origin and destination. It has been difficult to come to definitive conclusions. For example, a cross-sectional study of Mexican immigrant women in San Diego County by Vega et al. 1987, to determine whether certain factors have value for predicting the symptoms of depression, found that one's risk depended on the socioeconomic characteristics (i.e., education and income) of the individual, the perceived distance between the two points of migration (origin and destination), the degree of loss of interpersonal ties in Mexico, and the perceived economic opportunity.

There is strong evidence of a relation between mental illness and internal migration in the United States. More specifically, the evidence for an adverse effect of migration on mental health is strong for long-distance and interstate migration, and weak for short-distance migration. Pol and Thomas (2001) report that even very high-placed executives and their families often suffer traumatic health effects from mobility although it may mean a substantial career advance, greater prestige, and income. Migration is a particularly traumatic experience for some children, especially if they move more than once. The children lose contact with customary places, objects, and persons, especially friends, classmates, and teachers. This applies particularly to military families, who may have to move several times during their military service.

Conventional theory on the relation between immigration and the emotional responses of the immigrant, particularly the immigrants' adaptation to stress, posits

that, because it disturbs the equilibrium between the migrant and the environment, the act of migration is stressful for the migrant, that the migration experience forces the migrant to modify his/her habits and ways of living, and this process further induces stress and other unhealthful responses, including psychosomatic illnesses and even death (Ben-Sira 1997). The theory is not clearly substantiated by empirical research, as suggested by the following study.

According to the U.S. National Health Interview Survey for 1998–2003, native and immigrant adults were equally likely (3%) to experience symptoms of serious psychological stress (SPD) during the previous 30 days (U.S. NCHS/Dey and Lucas 2006). However, this equality represented offsetting records for the race-Hispanic groups. Native black and native Hispanic adults were more likely to experience symptoms of SPD than foreign-born adults of the corresponding ethnic/racial group. The figures for Hispanics are 4.4% and 3.6%, and the figures for blacks are 3.3% and 1.9%. The non-Hispanic white figures were reversed, with the native whites experiencing symptoms of stress less often than foreign-born whites. Note incidentally that Hispanics experienced more psychological stress than either blacks or whites. To measure SPD, six questions (i.e., feeling that everything was an effort and feeling sad, nervous, restless, hopeless, or worthless) were asked. The frequency of experiencing these symptoms in recent weeks, combined for the various symptoms, provided the basis of the SPD index.[7]

An illustration of research providing stronger evidence of the relation of mental health and immigration is the research on the environmental causes of schizophrenia being conducted in Scandinavian populations, the United Kingdom, and the United States since the 1990s. This research has found that immigrants to European countries are at heightened risk of schizophrenia as compared with native residents of these countries (reported by Bresnahan et al. 2007; Veling et al. 2007). Veling and colleagues describe the higher incidence of psychotic disorders among immigrants in the Hague, Netherlands. In another study, a study of the incidence of schizophrenia among immigrants in the cities of the United Kingdom, Afro-Caribbeans were found to be several times as likely as the general population to be treated for schizophrenia (Jarvis 1998).

These findings have implications for health care systems in countries, states, and local areas experiencing an influx of migrants, whether from abroad or at home. Mental health systems need to expand or at least adapt to the changed circumstances. The types of services may have to be modified to fill the needs of people of a different nationality, language, value system, and lifestyle. Lifestyles of immigrants may differ, especially with respect to health care. Some immigrants may be unaccustomed to Western methods of treatment. The issues may involve inducing them to seek treatment, returning to the treatment facility as scheduled, and following the treatment regimen at home.

[7]The questions asked how often the respondent experienced these symptoms during the previous 30 days. The response codes for each of the six items ranged from zero (none of the time) to 4 (all of the time) and were summed to yield a total score from zero to 24. A total score of 13 or more was used to define serious psychological stress. The index was developed by Kessler et al. (2002).

Health and Mortality of Refugee and Internally Displaced Populations

Refugees and internally displaced persons are migrants according to the conventional definition (*i.e.*, persons who change their usual place of residence across defined political boundaries), but I treat them separately from other migrants because they move under force and they present special problems with respect to health measurement, health analysis, and health management, particularly in the less developed areas.[8] We may also distinguish the cases where, such as in less developed areas, the forced migrants experience poor health because of a lack of basic services and the statistical mechanism for measuring this process is lacking, and the cases where, such as in Finland, the forced migrants experienced poor health because of high levels of social stress and a sophisticated vital statistics reporting system remained in place to measure the process (Saarela and Finnäs 2009).

Background Situation

Millions of refugees and millions of internally displaced persons are located in camps and settlements throughout the world, uprooted from their usual residences mainly because of civil war or political strife in their home countries. The number of refugees and other "persons of concern" to the UN High Commissioner for Refugees (UNHCR) in the world is estimated at 19.2 million in 2005 (UNHCR 2005; Table 10.6). The UN total includes far more refugees (9.2 million) than internally displaced persons (5.6 million). The US Committee for Refugees (USCR) considers these figures gross understatements, particularly the figure on internally displaced persons. According to its World Refugee Survey of 2004, the total number of refugees and internally displaced persons is 35.5 million, of which 23.6 million are internally displaced persons. Part of the difference between the two sets of figures arises from the specific categories of persons included in them but most of the difference arises from the difference in the sources of the estimates, as explained below.

The types of "persons of concern" distinguished by the United Nations are refugees, internally displaced persons, returning refugees, asylum-seekers, and stateless persons. Refugees are persons residing in countries other than their country of usual residence who cannot return to their home country because of a

[8]The Committee on Population of the National Academy of Sciences, National Research Council, established a Roundtable on the Demography of Forced Migration in 1999. The Roundtable organized a series of workshops on specific aspects of the demography of refugees, including mortality patterns, fertility, child health, and demographic assessment techniques. The discussion below is based partly on the results of these workshops. See the List of Suggested Readings and References at the end of the chapter for applicable publications.

Table 10.6 Estimates of "persons of concern," published by the United Nations High Commissioner on Refugees, by class: 2005 and 2004

	2005		2004	
	Number (in mil.)	Percent	Number (in mil.)	Percent
Total	19.2	100	17.0	100
Refugees	9.2	48	9.7	57
Internally displaced persons	5.6	29	5.3[a]	31[a]
Asylum seekers	0.8	4	0.9	5
Returned refugees	1.5	8	1.1	6
Stateless and others of concern	2.1	11	[a]	[a]

Source: UNHCR (2005)
For definitions of classes of "persons of concern," see text
[a] Figure for internally displaced persons includes estimate for stateless persons

well-founded fear of persecution or imprisonment for political, ethnic, racial, or religious reasons. Internally displaced persons are persons residing in their usual country of residence who have been displaced from their homes because of civil war or other political strife. Returning refugees are persons who were of concern to UNHCR when they were outside their country of origin and who remain persons of concern for a limited time even though they have returned to their country of origin. Asylum-seekers are those in a refugee-like situation in a country other than their country of usual residence who have not actually been granted refugee status by that host country. Defining these groups is a subject of much debate and controversy, and numerous variations on these definitions are employed by governments, international governmental agencies, and private international relief organizations. For example, private international relief organizations include in their count of internally displaced persons those who have been forced to move because of crop failure and famine, but the official organizations tend to exclude them from the counts of "persons of concern."

According to the UNHCR, the countries of Asia have produced the largest number and proportion of internally displaced persons, but there are large numbers also in Africa and Europe (Table 10.7).

Securing basic demographic and health data, such as data on population size, age-sex composition, deaths, and health conditions, for persons in refugee camps or in settlements of internally displaced persons, is extremely important for determining their requirements for health services, food and water supplies, and other forms of humanitarian assistance. Mortality data are essential for understanding the health conditions in the camps and settlements, but they are also useful for developing estimates of the size of the populations in them. We readily recognize that, in conditions of crises, it may not be possible to secure precise estimates of population size and mortality, and rough approximations may be adequate, especially during the early critical period. More exact estimates may later be necessary for historical analysis, reconstruction of the sequence of critical events, and preparation for response in future crises.

Table 10.7 Estimates of "persons of concern," published by the United Nations High Commissioner on Refugees, by region: 2005 and 2004

	2005		2004	
Region	Number (in mil.)	Percent	Number (in mil.)	Percent
Total	19.2	100	17.0	100
Africa	4.9	26	4.3	25
Asia	6.9	36	6.1	36
Europe	4.4	23	4.2	25
Latin America and Caribbean	2.1	11	1.3	8
Northern America	0.85	4.4	1.0	6
Oceania	0.08	0.4	0.07	0.4

Source: UNHCR (2005)

Several destabilizing social, political, and economic conditions create refugees and internally displaced persons. Among these conditions are continuing armed violence, civil war, political and military interference by neighboring states, economic exploitation of resources, and the absence of a central state authority in parts of a country. For example, in the Democratic Republic of the Congo, a civil war broke out in 1996 and lasted to 2001, or about 5 years. These conditions led to excessively high mortality, the depletion of the national treasury, poverty of the households affected, and a lack of public funds necessary for the support of the health system. Under these circumstances, providing basic health services to the population affected is very difficult, if not impossible. International governmental and nongovernmental organizations (NGOs) are unable to follow existing conventions regarding the delivery of health services to the target populations and try to deal with the health situation as best they can.

In addition to the usual problems in poor countries – poor roadways, insufficient funds available to the government and relief organizations, high demand for health services, and a shortage of adequately trained local health workers – there are the special conditions in the refugee camps and displaced-persons settlements. Among these are continuous insecurity, rapidly shifting membership, dire health conditions, and occasional chaos. The refugee and internally displaced populations are rapidly changing, even volatile, populations, with sharp fluctuations in numbers as a result of deaths, births, and new and departing "migrants."

The above circumstances describe the characteristics of a complex humanitarian emergency as recognized by epidemiologists specializing in chronic crises. A complex humanitarian emergency is formally defined as a relatively acute situation in which a large civilian population is affected by war, civil strife, or a violent attempt to restructure the state, leading to large-scale population displacement, deterioration of living conditions (e.g., shortages of food and potable water), and excess mortality (Toole and Waldman 1997). Epidemiologists commonly use two measures of "excess mortality" as cited in this definition: A daily crude mortality rate equal to or greater than 1 death per 10,000 population per day, and a daily death rate for children under 5 years of age equal to or greater than 2.4 deaths per 10,000 children. (These numbers correspond to an annual crude death rate equal

to or greater than 36 per 1,000 and a child death rate equal to or greater than 88 per 1,000.) Examples of complex humanitarian emergencies in recent years are the forced migrations associated with the wars, "ethnic cleansing," and genocide that occurred or are occurring in Bosnia, Darfur, East Timor, Kosovo, Republic of Congo, Rwanda, Sierra Leone, and Somalia.

Deriving Demographic and Health Data

Estimates of Total Population and Its Age-Sex Composition

Estimates of the population and its age-sex composition are required to measure the demographic extent of the emergency and particularly to measure the extent of the health problem. Because of the physically dangerous and chaotic conditions and the complex political situation usually prevailing in the affected areas, developing such estimates is a very problematic undertaking. Under the circumstances standard procedures for collecting demographic data and making population estimates are not usually applicable and novel techniques must be employed.

Because of the different applications of the terms refugee and internally displaced person by different organizations, the grave difficulties in the measurement of these populations associated with the unsettled environment, and the manipulation of the figures by governments, warring factions, and relief organizations, multiple, widely disparate estimates of the size of these populations are usually available. Of the two types of populations, internally displaced persons are more difficult to count because of their isolation and lack of official status. Those internally displaced persons who are "self-settled" in particular are exceedingly difficult to track and count because they have intermingled with the local people, are indistinguishable from them with respect to ethnicity, and may be assisted by them.

As noted above, the two major agencies that publish statistics on refugees and other displaced persons are the UN High Commissioner on Refugees (UNHCR) and the US Committee on Refugees (USCR). The UNHCR statistics are obtained from national governments' own records and from its own annual statistical survey. It also relies on local governments, religious authorities, local and international relief agencies, other UN agencies, and other secondary sources. USCR produces its own estimates, examining all the available estimates and then making an independent assessment. The methods of estimation used range from the most sophisticated to the most informal, depending on the situation. The USCR also relies on its own site visits to "sort the figures out" (USCR 2005). It is apparent that it is necessary to critically evaluate any estimates obtained for validity and reliability because of the unconventionality of some of the sources.

In the absence of dependable secondary sources, one can consider using primary sources or methods, that is, registration, enumeration, and sample surveys, to measure the number of refugees and displaced persons. Registration is not suitable if the group is very temporary or is "self-settled." Since registration statistics are

likely to be counts of the people who are being assisted by the international aid community, they are likely to understate the number of people internally displaced, especially if the displaced persons are living in the community. An enumeration may be considered, but it is generally unsuitable because the people are often on the move and the situation is unstable. Moreover, the people in the field may not be well trained in counting techniques and so there may be great variability in the reporting. On the other hand, refugee populations in a crisis situation are commonly overestimated because food distribution is linked to camp size, and internally displaced and local persons seek to be counted as refugees in order to receive food aid and other services.

During the very early stages of a crisis, rapid assessments are normally undertaken by U.S. AID teams. Estimates may have to be made by conjectural methods, that is, methods that do not depend on any body of data that directly reflect population size. These teams use various techniques from aerial photography to mapping (by either manual or geographic positioning systems) to simple visual assessments, in order to estimate promptly the number needing aid at the onset of a crisis. The area and population of a camp may be estimated on the basis of satellite images and in-situ data. Starting in 1997, U.S. companies have provided high-resolution images on a commercial basis. Sources such as graveyards, mullahs who distribute burial shrouds, and community informants and family members may also be examined or consulted.[9]

Registration of individuals or households, simple counting or censuses of individuals or households, or household surveys may be considered as possible methods of securing demographic and health data after the initial crisis phase of an emergency is over. Sample surveys may be a practicable and financially feasible way of collecting data, provided that conditions are relatively stable and secure within the country and the camps. In employing a sample survey, one has to choose between various types of area probability samples or multistage cluster samples (National Research Council/Reed 2002). One form of area probability or spatial sampling is the quadrate method, which divides the total area equally in small square blocks and then counts the population in a random sample of the blocks. In multistage cluster sampling, sample clusters of areas are first selected and then sample households are selected randomly from the clusters.

As stated earlier, estimates of the age and sex composition of the refugee and internally-displaced populations are also needed for planning services and for calculating mortality rates. Yet it is not common to find estimates of the age-sex composition of forced migrants. Some organizations use standard distributions to divide the total population into age-sex groups. That is, if the total size of the population is known or satisfactorily estimated, the approximate magnitudes of the various age-sex groups can be estimated by applying model age distributions to the total (see Chap. 12).

[9]Based on information obtained from Tom Argent, U.S. Committee for Refugees, Brenton T. Burkholder, U.S. Centers for Disease Control and Prevention, and Susanne Schmeidl, Center for Refugee Studies, York University.

Estimates of Mortality and Excess Mortality

In considering the difficulties of securing death statistics for forced migrants, we should bear in mind that a vital registration system or a population register does not usually exist, and if one did exist, it would probably break down in the crisis situation or be seriously flawed. Second, people are unwilling to report deaths because they fear that their family's food rations would be reduced if they report a decrease in their family size. Then, too, there may be cultural taboos about reporting death (National Research Council 1999). Moreover, the age-sex estimates of the population are likely to be flawed, so that the age-sex specific death rates are likely to be in serious error. Given the probable underreporting of deaths (i.e., the numerator of death rates) and the probable over reporting of population (i.e., the denominator of death rates), death rates are likely to be grossly understated. However, the mortality rates for the refugees in camps will in general be less subject to error than the mortality rates for displaced persons, who are dispersed over large areas in their home countries.

In the event of complex emergencies, the level of the mortality rate varies with the phase of the emergency. Complex emergencies have several well-established phases: a beginning phase, a peak phase, a phase when humanitarian assistance arrives, and a final phase when there is a return to stability (National Research Council/Keely et al. 2001). The mortality rate rises moderately in the first phase, then rises more steeply and begins to fall sharply in the second, falls off gradually in the third, and continues to fall gradually to baseline in the fourth. In most complex emergencies, governments and private organizations intervene to provide a measure of stability for the refugee population, helping to reduce mortality and sickness and thereby shortening the first two phases of the emergency and hastening a return to baseline mortality.

The methods available for estimating total population are generally applicable for measuring deaths. The mortality risks of certain age groups, e.g., children under 5 years of age, pregnant women, and the elderly, need to be separately measured because of the special vulnerability of these groups. Following the initial crisis, a sample survey may serve to provide estimates not only of total population and its age-sex distribution but also of mortality and health conditions. For example, the US Committee for Refugees (USCR) sponsored several sample surveys by the International Rescue Committee (IRC) in the camps of the Democratic Republic of the Congo during the 1996–2001 civil war. USCR used two-stage cluster sampling for conducting the surveys. Households responded to retrospective "verbal autopsies" relating to deaths in the household during the crisis. USCR also carried out a series of monitoring and evaluation surveys, such as of malaria prevalence and nutrition (National Research Council/Dikjzeul and Lynch 2006).

In areas of chronic crisis characterized by poverty, violence, and war, the level of mortality is well above normal. The excess mortality results from several related causes: The direct toll of war; the flight of people from their villages and/or their hiding places, often in the forest at night, where they may be attacked; the collapse of the health system; and the inability of people to buy the health services that remain

(National Research Council/Dikjzeul and Lynch 2006). By fleeing or hiding, their access to health care is limited, they are exposed to the elements and parasites, and they suffer from exhaustion and malnutrition. These conditions result in epidemics of diseases such as tuberculosis, cholera, meningitis, and malaria.

Epidemiologists have a special interest in measuring the excess mortality resulting from such complex emergencies. To do this, the concept of "baseline mortality" is employed. Baseline mortality is the "normal" mortality level of a given population. It would ordinarily be the mortality level before the emergency began and/or the level to which the mortality rate returns after the crisis has subsided. More exactly, it is the "normal" mortality level for the year in question, assuming that there was no humanitarian emergency and allowing for changes in trend to the current year. It is often difficult to establish such a baseline level in the affected countries because of the continuing unstable conditions and it is even more difficult to measure excess mortality for such areas under such conditions. In simple terms excess mortality is the excess of mortality during a year over normal mortality for that year.

It was estimated, on the basis of surveys taken by the International Rescue Committee (IRC) in the eastern sector of the Democratic Republic of the Congo in 2000, 2001, and 2002, that, between August 1998 and November 2002, 3.3 million (c.i., 3.0 to 4.7) excess deaths occurred in a population of approximately 20 million (Roberts et al. 2003). Another estimate based on a more recent IRC survey showed that by April 2004 the whole country had experienced 3.8 million (c.i., 3.5 to 4.4) excess deaths (Coghan et al. 2004). Because of the unsettled conditions in the Democratic Republic of the Congo, by the late 1990s the eastern sector of the country was described as an "unchecked incubation zone for diseases," with extremely high rates of excess mortality (Roberts 2000).

Estimates of daily crude mortality rates (CMRs) and annual crude death rates (CDRs) in selected refugee populations and internally displaced populations for 1990–1994 are shown in Tables 10.8 and 10.9 As a result of complex humanitarian emergencies in the 1990–1994 period, mortality was excessive (i.e., daily CMRs were equal to or greater than 1.0) in all these areas. The data in Tables 10.8 and 10.9 also reflect the wide variation in the burden of death on the refugee and displaced populations as well as the heavy burden on the assistance services.

Another example of an effort to measure mortality under crisis conditions is based on the experience in the Republic of Guinea in 2001. Since 1990 the Republic of Guinea has accepted 390,000–450,000 refugees from Sierra Leone and Liberia (CDC 2001). Subsequent attacks by armed factions in Guinea in late 2000 led to the sheltering of 58,000 refugees in camps in northwest Guinea. During the crisis in Guinea in January–May, 2001, NGOs collected and reported mortality data for camps that were accessible for on-site visits by international agencies. Deaths were listed by line entries, one line for each death. Monthly death rates for the camps usually were calculated by dividing the sum of all (monthly) deaths in the camps by the camps' (monthly) midpoint population and then dividing by the number of days in the month (for a per diem death rate) or the mean number of days the camps were open. Estimates of camp population were provided by the government of Guinea,

Table 10.8 Estimated daily crude mortality rates (CMRs) and annual crude death rates (CDRs) in selected refugee populations: 1990–1994

Period	Asylum country	Origin country	Daily CMR[a]	Annual CDR[b]
July 1990	Ethiopia	Sudan	2.3	84.0
June 1991	Ethiopia	Somalia	4.6	167.9
March-May 1991	Turkey	Iraq	4.1	149.7
March-May 1991	Iran	Iraq	2.0	73.0
March 1992	Kenya	Somalia	7.3	266.5
March 1992	Nepal	Bhutan	3.0	109.5
June 1992	Bangladesh	Burma	1.6	58.4
June 1992	Malawi	Mozambique	1.2	43.8
August 1992	Zimbabwe	Mozambique	3.5	127.8
December 1993	Rwanda	Burundi	3.0	109.5
August 1994	Tanzania	Rwanda	3.0	109.5
July 1994	Zaire	Rwanda	19.4–30.9	708.1–1,127.8

Source: Toole and Waldman (1997: Table 10.2)
[a] Deaths per 10,000 population per day
[b] Deaths per 1,000 population per year

Table 10.9 Estimated daily crude mortality rates (CMRs) and annual crude death rates (CDRs) in selected internally displaced populations: 1990–1994

Period	Country	Daily CMR[a]	Annual CDR[b]
January–December 1990	Liberia	2.3	84.0
April 1991–March 1992	Somalia (Merca)	4.5	164.3
April–November 1992	Somalia (Baidoa)	16.7	609.6
April 1992–December 1993	Somalia (Afgoi)	5.4	197.1
April 1992–March 1993	Sudan (Ayod)	7.6	277.4
April 1992–March 1993	Sudan (Akon)	4.5	164.3
April 1992–March 1993	Bosnia (Zepa)	1.0	36.5
April 1993	Bosnia (Zarajevo)	1.0	36.5
May 1995	Angola (Cafunfo)	8.2	299.3
February 1996	Liberia (Bong)	5.4	197.1

Source: Adapted by Keely et al. (2001) from Toole and Waldman (1997: Table 10.3)
[a] Deaths per 10,000 per day
[b] Deaths per 1,000 per year

UNHCR, NGOs, and refugee and other organizations. Since these estimates varied widely, the lowest estimates for all camps were used to calculate the mortality rates.

Because of the excess malnutrition among the camp and settlement residents, there is a strong interest in gathering up-to-date data on the nutrition conditions there, especially among the children. Estimating nutrition rates among displaced populations that are moving in and out of the camps and settlements at varying rates with varying accessibility is a demanding and even daunting task. Nutrition surveys were used to measure the prevalence of acute malnutrition among displaced children aged 6–48 months in northwestern Guinea. Nutrition screening data were collected

for all children entering new camps in this region of the country. Such surveys could not be conducted in less accessible camps. Of 4,771 children who were screened in the new camps using weight-for-height measurements as an indicator of acute malnutrition during February–May, 2001, 119, or 2.5%, were acutely malnourished.

In the above section I have described the prevailing experience with forced migration in the world today. A much different experience is represented by the history of the Karelians in Finland. Many decades after being forcefully displaced from their historical region of birth and residence in eastern Finland, the men experienced mortality rates 20% above the expected time trend and above the level of their neighbors born in eastern Finland. This excess mortality was presumed to be a result of the extended effects of psychosocial stress occasioned by the breakup of the Soviet Union and the movement to recover the region of their birth, which was ceded to the Soviet Union after World War II. According to the authors (Saarela and Finnäs 2009), this case illustrates how, when internally displaced persons must adjust to situations for which appropriate coping behaviors are unknown, psychosocial stress might arise several decades after the territorial displacement.

References and Suggested Readings

International Migration

Abraido-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The Latino mortality paradox: A test of the 'salmon bias' and healthy migrant hypotheses. *American Journal of Public Health*, *89*, 1543–1548.

Antecol, H., & Bedard, K. (2006). Unhealthy assimilation: Why do immigrants converge to American health status levels? *Demography*, *43*(2), 337–360.

Ben-Sira, Z. (1997). *Immigration, stress, and readjustment*. West Port, CT: Praeger Publishers.

Bresnahan, M., Begg, M. D., Brown, A., Scaeffer, C., et al. (2007). Race and schizophrenia in a U.S. birth cohort: another example of health disparity. *International Journal of Epidemiology*, *36*(4), 751–758.

Chen, J., Ng, E., & Wilkins, R. (1996). The health of Canada's immigrants. *Health Reports*, *7*, 33–45, 47–50.

Cho, Y., Frisbie, W. P., Hummer, R. A., & Rogers, R. G. (2004). Nativity, duration of residence, and the health of the Hispanic adults in the United States. *International Migration Review*, *38*, 184–211.

Courbage, Y., & Khlat, M. (1996). Mortality and causes of death of Moroccans in France, 1979–81. *Population-E*, *8*, 59–84 (INED).

Deboosere, P., & Gladeyne, S. (2005). Adult migrant mortality advantage in Belgium: Evidence using census and register data. *Population-E*, 60(5–6), 655–698 (INED).

Donovan, J. M., Espaignet, E., Merton, C., & van Ommeren, M. (Eds.). (1992). *Immigrants in Australia: A health profile* (Australian Institute of Health and Welfare. Ethnic Health Series, No. 1). Canberra, Australia: AGPS.

Elo, I. T., Turra, C.M., Kestenbaum, B., & Ferguson, R. F. (2004). Mortality among elderly Hispanics in the United States: Past evidence and new results. *Demography*, *41*, 109–128.

Glicksman, A. (2005, November 18–22). *The Chernobyl nuclear disaster in Ukraine*. Paper presented at the annual meeting of the Gerontological Society of America, Orlando, FL.

Jarvis, E. (1998). Schizophrenia in British immigrants: Recent findings, issues, and implications. *Transcultural Psychiatry*, *35*(1), 39–74.

Jasso, G., Massey, D. S., Rosenzweig, M. R., & Smith, J. P. (2004). Immigrant health: Selectivity and acculturation. In N. B. Anderson, R. A. Bulatao, & B. Cohen (Eds.), *Critical perspectives on racial and ethnic differences in health in late life*. Washington, DC: National Academy Press.

Kessler, R. C., Andrews, G., Colpe, L.J., et al. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*, 959–976.

Keys, A. (1980). *Seven countries: A multivariate analysis of death and coronary heart disease*. Cambridge, MA: Harvard University Press.

Klinthall, M. (2008). Ethnic background, labour market attachment and severe morbidity: Hospitalization among immigrants in Sweden: 1990–2001. *Journal of International Migration and Integration*, *9*(1), 45–61.

Kouris-Blazos, A. (2002). Morbidity mortality paradox of 1st generation Greek Australians. *Asia Pacific Journal of Clinical Nutrition, 11*, 569–575.

Markides, K. S. (2001). Migration and health. In N. J. Smelzer & P.B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 14, pp. 9799–9803). NewYork: Elsevier.

Markides, K. S., & Eschbach, K. (2005). Aging, migration, and mortality: Current status of research on the Hispanic paradox. *Journal of Gerontology*: Series B, *60B*(Special Issue II), S68–S75.

Marmot, M. G., & Syme, S. L., (1976). Acculturation and coronary heart disease in Japanese-Americans. *American Journal of Epidemiology, 104*, 225–247.

McDonald, J. T., & Kennedy, S. (2004). Insights into the 'healthy immigrant effect': Health status and health service use of immigrants to Canada. *Social Science and Medicine, 59*, 1613–1627.

Newbold, K. B. (2006). Chronic conditions and the healthy immigrant effect: Evidence from Canadian immigrants. *Journal of Ethnic and Migration Studies*, *32(5)*, 765–784.

Ng, E., Wilkins, R., & Berthelot, J. -M. (2004, April 1–3). *Healthy immigrant effect in Canada: A longitudinal perspective using national population health surveys*. Poster presentation at the annual meeting of the Population Association of America, Boston, MA.

OECD Observer. (2004). *234*, May 2004.

Padareth, A., Chamberlain, C., McCoy, D., Ntuli, A., et al. (2003). *Health personnel in Southern Africa: Confronting maldistribution and brain drain* (Equinet discussion paper number 3). Regional Network for Equity in Health in Southern Africa (EQUINET), Health Systems Trust (South Africa), and MEDACT (UK).

Palloni, A., & Arias, E. (2004). Paradox lost: Explaining the Hispanic mortality advantage. *Demography, 41(3)*, 385–415.

Pew Hispanic Center. (2006). *Size and characteristics of the unauthorized migrant population in the United States: Estimates based on the March 2005 current population survey*. By J. S. Passel. Washington, DC: Pew Hispanic Center.

Razum, O., et al. (1998). Low overall mortality of Turkish residents in Germany persists and extends into a second generation: Merely a healthy migrant effect? *Tropical Medicine and International Health*, *3*(4), 297–303.

Sevak, P., & Schmidt, L. (2008). Immigrant – native fertility and mortality differentials in the United States (Working Paper WP2008-181). Ann Arbor, MI: University of Michigan Retirement Research Center.

Singh, G. K., & Miller, B. A. (2004). Health, life expectancy, and mortality patterns among immigrant populations in the United States. *Canadian Journal of Public Health*, *95*, 114–121.

Singh, G. K., & Siahpush, M. (2001). All-cause and cause-specific mortality of immigrants and native-born in the United States. *American Journal of Public Health, 91*(3), 392–399.

Uitenbroek, D. J., & Verhoeff, A. P. (2002). Life expectancy and mortality differences between migrant groups living in Amsterdam, the Netherlands. *Social Science and Medicine, 54(9)*, 1379–1388.

United Nations. (2006). *International migration report, 2006*.

U.S. National Center for Health Statistics. (1994). Health of the foreign-born population: United States, 1989–90. By E. H. Stephen, K. Foote, G. E. Hendershot, & C. A. Schoenborn. *Advance Data. Vital and Health Statistics,* 241.
U.S. National Center for Health Statistics. (2006). Physical and mental health characteristics of U.S. and foreign-born adults: United States, 1998–2003. By A. N. Dey & J. W. *Lucas. Advance Data. Vital and Health Statistics*, 369.
Vega, W. A., Kolody, B., & Valle, J. R. (1987). Migration and mental health: an empirical test of depression risk factors among immigrant Mexican women. *International Migration Review, 21*(3), 512–530.
Veling, W., Susser, E., van Os, J., Machinbach, J.P., et al. (2007). Ethnic density of neighborhoods and psychotic disorders among immigrants. *American Journal of Psychiatry*. ajp.psychiatryonline.org. Accessed October 2009.
World Bank. (2008). *Migration and remittance factbook*. Accessed August 1, 2008, from http://econ.worldbank.org
World Health Organization, Regional Office for Africa. (2004). *Migration of health professionals in six countries: A synthesis report*. By M. Awases, J. Nyoni, A. Gbary, & R. Chatora. Geneva, Switzerland: World Health Organization.

Internal Migration

Angel, R. J., Angel, J. L., & Himes, C. L. (1992). Minority group states, health transitions, and community living arrangements among the elderly. *Research on Aging*, *14*, 496–521.
Bean, F. D., Myers, G. C., Angel, J. L., & Galle, O. R. (1994). Geographic concentration, migration, and population redistribution among the elderly. In L. G. Martin & S. H. Preston (Eds.), *Demography of aging*, Washington, DC: National Academy Press.
Biggar, J. C. (1980). Who moved among the elderly: 1965–1970? A comparison of types of older movers. *Research on Aging*, *2*, 73–91.
Cutler, D., & Miller, G. (2005). The role of public health improvements in health advances: The twentieth-century United States. *Demography*, *42*(1), 1–22.
Litwak, E., & Longino, C. F. (1987). Migration patterns among the elderly: A developmental perspective. *Gerontologist*, *27*, 266–272.
Lu, Y. (2008). Test of the 'healthy migrant hypothesis': A longitudinal analysis of health selectivity of internal migration in Indonesia. *Social Science and Medicine, 67*(8), 1331–1339.
Monahan, D. J., & Greene, V. L. (1982). The impact of seasonal population fluctuations on service delivery. *Gerontologist*, *22*, 160–163.
Nguyen, L. T. (2004, April 1–3). *Health status, urbanization, and migration in Vietnam: Preliminary findings from the1997–98 living standard survey*. Paper presented at the annual meeting of the Population Association of America, Boston, MA.
Pol, L. G., & Thomas, R. K. (2001). *The demography of health and health care* (2nd ed.). New York: Kluwer Academic/Plenum Publishers, esp. Chaps. 7–11.
Smith, S. K., & House, M. (2006). Snowbirds, sunbirds, and stayers: Seasonal migration of elderly adults in Florida. *Journal of Gerontology: Social Sciences, 61B(5)*, S232–S239.
Speare, A., & Meyer, J. W. (1988). Types of residential mobility and their determinants. Journal of Gerontology: *Social Sciences, 43*, S74–S71.
Ssengonzi, R., DeJong, G. F., & Shannon, C. (2002). The effect of female migration on infant and child mortality in Uganda. *Population Research and Policy Review*, *21*(5), 403–431.
Sullivan, D. A. (1985). The ties that bind. *Research on Aging*, *7*, 235–250.
United Nations. (2007). *World urbanization reports: The 2007 revision, executive summary*.
U.S. Census Bureau. ca. (1975). The geographical mobility of Americans: An international comparison. By L. H. Long & C. G. Boertlein. *Current Population Reports*, Series P-23, No. 64.

U.S. Census Bureau. (2001). Why people move: Exploring the March 2000 current population survey. By J. Schachter. *Current Population Reports*, P23–204.
Wiseman, R. F. (1980). Why older people move. *Research on aging*, *2*, 141–154.
Worobey, J. L., & Angel, R. J. (1990). Functional capacity and living arrangements of unmarried elderly persons. *Journal of Gerontology: Social Sciences*, *45*, S95–S101.
Zimmerman, R. S., Jackson, D. J., Longino, C. F., & Bradsher, J.E. (1993). Interpersonal and economic resources as mediators of the effects of health decline on the geographic mobility of the elderly. *Journal of Aging and Health*, *5*, 37–57.

Refugees and Internally Displaced Persons

Burkholder, B. T., Spiegel, P., & Salama, P. (2001). Methods of determining mortality in the mass displacement and return of emergency-affected populations in Kosovo, 1998–99. In *Forced migration and mortality*. (pp. 86–101) Washington, DC: National Academy Press.
Coghan, B., Brennan, R., Ngoy, P., Dofara, D., et al. (2004). *Mortality in the democratic republic of the Congo: Results from a nationwide survey*. Melbourne and New York: Burnet Institute and International Rescue Committee.
Keely, C. B., Reed, H. E., & Waldman, R. J. (2001). Understanding mortality patterns in complex humanitarian emergencies. In H. E. Reed & C.B. Keely (Eds.), National Research Council, *Forced Migration and Mortality* (pp. 1–37). Washington, DC: National Academy Press.
National Research Council. (1999). *The demography of forced migration: Summary of a workshop*, H. Reed, J. Haaga, and C. Keely (Eds.). Washington, DC: National Academies Press.
National Research Council. (2001). *Forced migration and mortality*. H. E. Reed and C.B. Keely (Eds.). Washington, DC: National Academy Press.
National Research Council. (2002). *Demographic assessment techniques in complex humanitarian emergencies: Summary of a workshop*, H. E. Reed (rapporteur). Washington, DC: National Academy Press.
National Research Council. (2003). *Malaria control during mass population movements and natural disasters*. By P. B. Boland & H. A. Williams. Washington, DC: National Academies Press.
National Research Council. (2006). *Supporting local health care in a chronic crisis: Management and financing approaches in the Eastern Democratic Republic of the Congo*. Summary by D. Dijkzeul & C. Lynch. Roundtable on the Demography of Forced Migration, Committee on Population, and Program on Forced Migration, Mailman School of Public Health, Columbia University. Washington, DC: National Academies Press.
Roberts, L. (2000). *Mortality in Eastern DRC: Results from five mortality surveys*. New York: International Rescue Committee Health Unit.
Roberts, L., Zantop, M., Ngoy, P., Lubula, C., et al. (2003). *Mortality in the Democratic Republic of Congo: Results from a nationwide survey*. New York: International Rescue Committee.
Saarela, J., & Finnäs, F. (2009). Forced migration and mortality in the very long term in Finland. *Demography*, *46*(3), 575–588.
Toole, M. J., & R. J. Waldman. (1997). The public health aspects of complex emergencies and refugee situations. *Annual Review of Public Health, 18*, 283–312.
United Nations High Commissioner for Refugees. (2005). *Refugees by numbers* (2005 ed.). New York: UNHCR.
U.S. Centers for Disease Control and Prevention (CDC). (2001). Surveillance of mortality during a refugee crisis – Guinea, January–May 2001. *Morbidity and Mortality Weekly Report*. Atlanta: Centers for Disease Control and Prevention.
U.S. Committee on Refugees. (2005). *World refugee survey, 2004*. New York: International Rescue Committee.

Chapter 11
Special Health Issues in Less Developed Countries

Introduction

Sources and Quality of Data for Less Developed Countries[1]

In the Less Developed Countries (LDC) the information on illness that is secured about the population is usually provided by the individuals themselves or their caregivers. The information on death is usually provided by caregivers in verbal autopsies, that is, interviews with caregivers or relatives who are familiar with the circumstances or conditions that led to the death of the decedent. This method has been employed widely in the LDC, especially in sub-Saharan Africa, to secure information on the cause of death, particularly the deaths of children (Snow et al. 1992; Oosterbaan 1995). Verbal autopsies are useful for securing cause-of-death data where vital registration systems are lacking and where, as in sub-Saharan Africa, most deaths occur outside of a hospital or clinic (Snow and Marsh 1992).

Much of the information secured is inaccurate or incomplete. Errors may be made inadvertently or intentionally, depending on the circumstances. Some health conditions like HIV/AIDS (i.e., human immunodeficiency virus/acquired immune deficiency syndrome) are stigmatized and, hence, the caregiver may instead report some associated condition, such as tuberculosis. Very commonly the cause of death is not reported. For example, in some parts of sub-Saharan Africa more than one-third of the deaths recorded are shown as cause not reported.

There are many variations in the mode of collection of data on deaths. A Demographic Surveillance System (DSS), as conducted by the Navrongo Health Research Centre in northern Ghana, may collect data on deaths and their causes in a limited geographic area. The deaths recorded in the DSS may be followed up by verbal autopsies to determine or verify the cause of death. This system is

[1]This section on sources of data for Less Developed Countries supplements the discussion of general sources in Chap. 2 and the discussion of special sources on reproductive health in Chap. 9.

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_11, © Springer Science+Business Media B.V. 2012

like a system of sample death registration areas. In the Demographic and Health Surveys (DHS), more formal biometric or anthropometric devices may be employed to establish the cause of illness or death. For example, a sample of blood may be taken to distinguish between HIV/AIDS and tuberculosis. To measure stunting of a child that may have resulted from undernutrition, height may be measured and matched with age.

Inasmuch as most people in the Less Developed Countries, particularly the Least Developed Countries, do not spend their days of serious illness or their final days in a hospital, many deaths are not recorded. Perinatal deaths are commonly buried without a record. There are no legal requirements for reporting fetal losses and the remains may be buried privately and informally in a nearby field. Deaths of infants aged less than 1 week may be given the same informal treatment. Deaths of older persons are frequently not registered and reporting in health surveys may be incomplete.

The most comprehensive source of information on the extent and causes of death, disease, and injuries in the Less Developed Countries is *The Global Burden of Disease* study, published in 1996 by Harvard University on behalf of the World Health Organization and the World Bank (Murray and Lopez 1996) (See also Chap. 8.). The *Global Burden of Disease* study elaborated the available data as of 1990 to provide the first systematic picture of death, disease, and injuries in all the regions of the world and developed projections of the world health situation in 2020. In the years following 1996 the methodology for measuring the disease burden was refined, the supply of data on death and disease in the Less Developed Countries increased, the quality of data improved, and the world health situation changed. Given these developments, an updated Global Burden of Disease study was undertaken in these years and published by Lopez et al. (2006). In addition, the World Health Organization produced a more limited assessment of the global burden of disease for 2002 in the *World Health Report* for 2004. Revised projections of global mortality and the burden of disease, consistent with the second Global Burden of Disease study, were prepared for 2020 and 2030 (Mathers and Loncar 2006).

General Context for the Health Situation in the Less Developed Countries

All the health issues of concern in the MDC are applicable to the LDC but are exacerbated in the latter areas by several factors. These factors include the large proportion of children and the rapid growth in their numbers, the unprecedented pace of the increase in the number of elderly persons, the lack of a functioning social security system, the low or negative economic growth, the general poverty of much of the population, the difficult and even harsh geographic conditions, the inadequate public resources, the poorly developed health systems, lack of access to health care and health facilities, and the unstable political environment.

These countries suffer from a shortage of health facilities, physicians, nurses, and other health providers. At a time when physicians and nurses are most needed in the LDC, they have been recruited away to the West, have left their countries in order to take advantage of the greater opportunities abroad, or have fled intolerable conditions at home. Most countries of the West have over 25 doctors for every 10,000 persons, but many countries of sub-Saharan Africa have only a few doctors for every 10,000 persons. This shortage is exacerbating the problems created by HIV/AIDS and the many endemic diseases in these countries. Moreover, the physicians practicing in the LDC work mostly in the large cities, and yet a large share of the population inhabiting these countries lives in rural areas. This is especially true in sub-Saharan Africa, where the rural population exceeds 60% of the total population of many countries.

The health situation in the poorer countries has been becoming both more complex and more burdensome. The rapid changes in their demographic situation have contributed to this development. The massive urbanization that has been occurring has been bringing with it serious health problems, particularly the spread of infectious diseases. These countries have had considerable success in eradicating some infectious and parasitic diseases, but other such diseases persist. The health situation in many of the LDC has greatly deteriorated as a result of the spread of HIV/AIDS. In addition, these countries now have to deal with the rise of the noncommunicable diseases, such as heart disease, stroke, cancer, and chronic obstructive pulmonary disease, as well as with other conditions characteristic of the MDC, such as obesity, mental disorders, and road traffic accidents. The epidemiological transition has been occurring far more rapidly in the LDC than it did in the MDC. Infectious diseases continue to be a major cause of death for poor people in the LDC, while the chronic degenerative diseases are the leading killers of the more affluent persons in these countries. The health systems of the poor countries cannot handle the burden of both infectious diseases and noncommunicable diseases at the same time. The monumental health problems of the LDC are related in large part to their general lack of resources and their ill-developed public health systems.

In addition to the wide differences between the health conditions of the MDC and the LDC, the various socioeconomic classes in the latter countries differ greatly in health status from one another and these differences have been increasing. Inequality in the health status of the segments of the LDC populations can be extreme, as is the case in China.

Least Developed Countries

Because of the great variations in the economic status and the health status of the inhabitants of the LDC, both among and within them, it is desirable to distinguish the least developed countries (including the landlocked LDC and small-island less developed states) from the remaining less developed countries. The United Nations has defined such a sub-set of the LDC countries. This group of countries corresponds

Table 11.1 Gross national income per capita ($US) for world regions and selected countries: 2007 (Represents gross national income (GNI) in purchasing power parity (PPP) divided by midyear population. Gross national income is converted to "international" dollars by a conversion factor representing the amount of goods and services one could buy in the United States with a given amount of money)

World	$9,600
More Developed	31,200
Less Developed	4,760
Least Developed	1,060
Northern Africa	4,760
Sub-Saharan Africa	1,830
Western Africa	1,480
Eastern Africa	940
Burundi	330
Middle Africa	1,550
Southern Africa	9,140
Northern America	44,790
United States	45,850
Latin America/Caribbean	9,080
Western Asia	10,160
South Central Asia	2,940
Tajikistan	1,710
Southeast Asia	4,440
Cambodia	1,690
East Asia	8,380
China	5,370
Northern/Western Europe	34,766
Eastern Europe	13,210
Moldova	2,930
Russia	14,400
Southern Europe	26,230
Albania	6,580
Oceania	23,910
Solomon Islands	1,400

Source: Population Reference Bureau (2008). Reprinted with permission of the Population Reference Bureau. Primary source: World Bank

approximately to the group of countries considered very low-income countries (Table 11.1). Most of the countries of sub-Saharan Africa (*e.g.*, Malawi, Niger, Chad), some countries in South Central and Southeast Asia (e.g., Bangladesh, Myanmar, Cambodia), and some countries of the Caribbean and Oceania (*e.g.*, Haiti, Solomon Islands) are in the Least Developed Countries class (Exhibit A.2). Most countries of South and Central America and East Asia are in the more "affluent" LDC class. The least developed countries, especially, suffer from the combination of endemic poverty, political instability associated often with civil war, masses of internally displaced persons and refugees, negative economic growth, high levels of illiteracy among both youths and adults, and a rapid increase in the numbers of children and elderly persons. As a result, these countries are characterized by low life expectancy, high rates of infectious diseases, including HIV/AIDS, increasing rates of chronic diseases, and very high rates of maternal and perinatal mortality.

General Burden of Disease by Region

In this section I describe selected aspects of the depressed health situation in the Less Developed Countries, or the Low and Middle Income Countries. Many of these topics will be developed more fully in later sections.

Life Expectancy, Maternal Mortality, and Undernutrition

There is a considerable gap in the health situation of the rich and poor countries and it has been widening. The commonest measures of this gap are life expectancy at birth, infant mortality, and maternal mortality. In general, the world's poorer countries have low life expectancies and high infant and maternal mortality rates, and the world's wealthier countries have high life expectancies and low infant and maternal mortality rates. In 2008 life expectancy at birth lagged 10 years in the LDC (67 years) behind the MDC (77 years); the Least Developed Countries lagged 22 years (55 years). (See Population Reference Bureau 2008.) While infant mortality and maternal mortality rates in the LDC have been improving, they continue to be distressingly high. For example, in 2007 in the LDC about 1 in 75 women, and in the Least Developed Countries 1 in 22 women, died from pregnancy-related causes (Population Reference Bureau 2008). The LDC relative frequency for maternal mortality is 80 times that in the MDC. The very poor health status of the Least Developed Countries is suggested further by the data in Table 11.2, which presents estimates of life expectancy and infant mortality rates for selected Least Developed Countries in 2007. Life expectancy at birth in these countries is often below 50 years and the infant mortality rate is often above 100 deaths per 1,000 births.

The percent of the population that is undernourished may be taken as an indirect summary indicator of health differences between these groups of countries. Less than 2.5% of the MDC population, 17% of the LDC population, and 35% of the population of the Least Developed Countries is undernourished (Population Reference Bureau 2008). That is, one out of three persons in the latter group of countries consumes fewer than the minimum number of calories required to pursue a healthy active life.

Health Expectancy

As described in Chap. 8, by extending conventional life tables to take account of health status, life expectancy can be disaggregated into healthy expected years and unhealthy expected years. From the World Health Organization's estimates of healthy and unhealthy life expectancy for its member countries for 2002, we can draw the following conclusions. For both males and females in the LDC, unhealthy

Table 11.2 Life expectancy and infant mortality rate for selected least developed countries: 2007 (Includes landlocked less developed countries and small island less developed states)

	Life expectancy	Infant mortality rate
More Developed	77	6
Less Developed	66	57
Africa		
Northern Africa		
Sudan	58	69
Western Africa		
Guinea-Bissau	46	117
Liberia	45	138
Niger	56	126
Togo	58	91
Eastern Africa		
Burundi	49	107
Ethiopia	49	77
Malawi	40	96
Zambia	38	100
Middle Africa		
Angola	41	141
Central African Rep.	43	102
Equatorial Guinea	49	105
Southern Africa		
Botswana	34	56
Lesotho	36	91
Asia		
South Central Asia		
Afghanistan	42	166
Bangladesh	62	65
Nepal	62	51
Southeast Asia		
Cambodia	63	71
Laos	55	85
Myanmar	60	75
Western Asia		
Yemen	63	64
Latin America and the Caribbean		
Caribbean		
Haiti	58	57
Oceania		
Kiribati	59	43
Solomon Islands	62	48

Source: Population Reference Bureau 2007.*World Population Data Sheet*, Washington, DC: Population Reference Bureau, 2007. Reprinted with permission of the Population Reference Bureau

Table 11.3 Percentage of total life expectancy at birth lost as unhealthy years, by sex, for selected less developed and more developed countries: 2002

	Male	Female
Algeria	11.7	13.5
Bangladesh	11.7	14.8
Bolivia	13.2	14.8
Brazil	13.0	13.6
Burundi	13.7	14.4
China	9.3	10.4
Ethiopia	13.0	15.6
India	11.3	13.6
Malawi	12.1	14.3
Nigeria	14.1	15.6
Philippines	12.4	14.3
South Africa	11.3	13.8
Thailand	13.0	15.2
Australia	9.0	10.4
Canada	9.2	10.0
Germany	7.8	9.3
Italy	7.8	9.5
Norway	7.8	9.9
Russian Federation	9.4	10.7
Sweden	7.9	9.5
United States	9.9	10.7

Source: Based on World Health Organization (2004), Annex Table 4. Reprinted with permission of the World Health Organization

years make up a greater share of total years of life than in the MDC, and in both the LDC and MDC, females lose more years as unhealthy years than males (Table 11.3). The share of life expectancy at birth lost as unhealthy years for females runs from about 13% to 16% in the LDC and from about 9% to 11% in the MDC. China (as well as the Russian Federation) is an exception, with figures resembling those in the MDC more than those in the LDC.

Burden of Disease

Inasmuch as the regions of the world display substantial variation in their rates of mortality and disease, it may be expected that they differ greatly in their overall burden of disease. As may be recalled, disability-adjusted life years (DALY) was the measure used by the Global Burden of Disease (GBD) studies of 1996 and 2006 to measure the burden of disease. The first GBD study defined eight geographic regions, of which six are less developed regions and two are more developed regions, and applied several health measures, among them death rates, the years of life lost prematurely (YLL), years lived with a disability (YLD), and their combination,

i.e., disability-adjusted life years (DALY).[2] The second GBD study defined seven geographic regions, six constituting the low-income and middle-income group of countries and one the high-income group (mainly the industrial countries of North America, Western and Central Europe, Asia, and Oceania), and the same four health measures.

The 1996 GBD study found that in 1990 the age-standardized death rate and the disability-adjusted life years per 1,000 population were considerably higher in the less developed regions than in the more developed regions. The MDC, composed of the Established Market Economies and the Formerly Socialist Economies of Europe, show the lowest YLL rates and YLD rates; the LDC regions of China, Latin America and Caribbean, and the Middle Eastern Crescent, show intermediate rates; and Sub-Saharan Africa shows the highest rates. The Global Burden of Disease study of 1996 summarized the regional imbalances in the burden of disease in the world in 1990 as follows:

> The peoples of Sub-Saharan Africa and India together bore more than four-tenths of the total global burden of disease in 1990, although they make up only 26% of the world's population. By contrast, the Established Market Economies and the Formerly Socialist Economies of Europe, with about a fifth of the world's population between them, together bore less than 12% of the total disease burden. China emerged as substantially the most 'healthy' of the developing regions, with 15% of the global disease burden and a fifth of the world's population. Put differently, about 579 years of healthy life were lost for every 1,000 people in Sub-Saharan Africa, compared with just 124 for every 1,000 people in the Established Market Economies. This assessment demonstrates clearly the glaring inequalities of world health at the end of the 20th century.
>
> The rates of premature death varied sharply between regions, with rates seven times higher in Sub-Saharan Africa than in the Established Market Economies. By contrast, the rates of disability were less varied, with Sub-Saharan Africa having twice the rate of YLDs as the rich countries. (See Fig. 11.1).

Leading Causes of Death and Disease

Noncommunicable Diseases

Heart disease, cerebrovascular disease, and chronic obstructive pulmonary disease (COPD) – all leading causes of death in the industrial countries in 1990, ranking first, second, and fifth – were among the ten leading causes of death in the LDC in that year as well (Table 11.4). In the LDC their ranks were also high – second, third, and seventh. The first two of these causes figured among the ten leading causes of the LDC burden of disease (DALY) as well, although their ranks are lower, at the

[2] See Chap. 8 for a further discussion of the DALY as a measure of the burden of disease, including the method of computation.

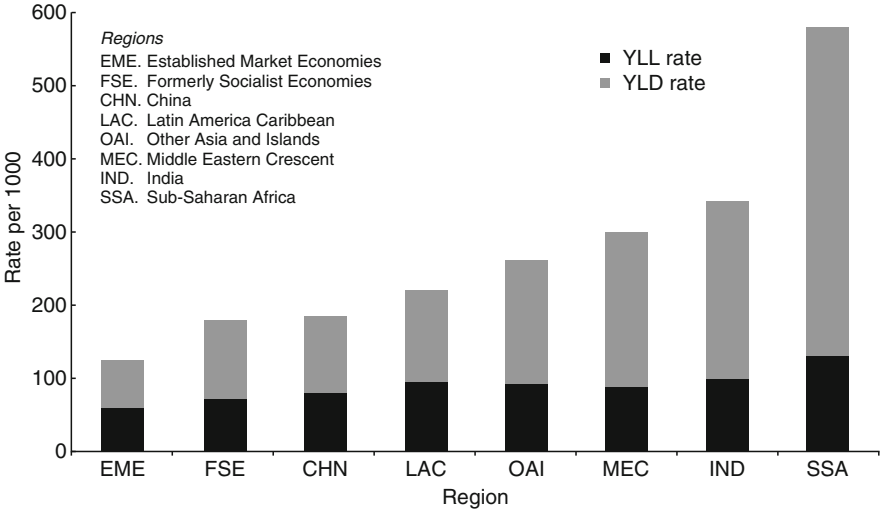


Fig. 11.1 Disability-adjusted life years (DALY) per 1,000 population, disaggregated into years of life lost (YLL) and years lost to disability (YLD), for world regions: 1990 (Source: Murray and Lopez 1996, Figure 12, p.23. Copyright © 1996. World Health Organization. Reprinted with permission)

eighth and tenth positions, because of the greater role of infectious diseases and perinatal diseases than noncommunicable diseases in contributing to disabled years (Table 11.5).[3]

The number of deaths from cardiovascular diseases has been increasing so rapidly in the LDC that they have become the leading causes of death in many of these countries and will soon become the leading causes in several more (Reddy and Yusuf 1998). As is the case in North America, cardiovascular diseases are already the leading cause of death in Latin America and the Caribbean, South Asia, and Eastern Europe and Central Asia. Cardiovascular diseases now make up about one-quarter of all deaths in the LDC as compared with nearly two-fifths in the United States, and are expected to show a sharp increase in their share of total deaths in the next few decades. As a result of these developments and the large share of the world's population in the LDC, the vast majority of deaths from cardiovascular diseases in the world occur in the LDC (Table 11.6). Of the deaths in the world due to the three major categories of cardiovascular diseases, 83% occurred in the low and middle income countries in 2001. Of special note is the fact that deaths from

[3] As will be evident in several tables, we can often derive an approximate but informative picture of the health situation in the less developed regions or the low and middle income regions from estimates for the world as a whole because these regions compose over four-fifths of the world's population and many leading health conditions are concentrated in these areas.

Table 11.4 The ten leading causes of death for the less developed and the more developed countries: 1990

	Less developed countries			More developed countries	
	Deaths	Percent of total		Deaths	Percent of total
All causes	39,554	100.0	All causes	10,912	100.0
1. Lower respiratory infections	3,915	9.9	1. Ischemic heart disease	2,695	24.7
2. Ischemic heart disease	3,565	9.0	2. Cerebrovascular disease	1,427	13.1
3. Cerebrovascular disease	2,954	7.5	3. Trachea, bronchus,	523	4.8
4. Diarrheal diseases	2,940	7.4	4. Lower respiratory infections 385	3.5	
5. Conditions arising during the perinatal period	2,361	6.0	5. Chronic obstructive pulmonary disease	325	3.0
6. Tuberculosis	1,922	4.9	6. Colon and rectal cancers	277	2.5
7. Chronic obstructive pulmonary disease	1,887	4.8	7. Stomach cancer	241	2.2
8. Measles	1,056	2.7	8. Road traffic accidents	222	2.0
9. Malaria	856	2.2	9. Self-inflicted injuries	193	1.8
10. Road traffic accidents	777	2.0	10. Diabetes mellitus	176	1.6

Source: Murray and Lopez (1996), Table 2, p.18. Copyright © 1996 World Health Organization. Reprinted with permission
Deaths in thousands. "All causes" includes remaining causes not listed among the first ten causes

cardiovascular diseases are significantly higher among the working-age population in low and middle income countries than in high-income countries.

This change is occurring in part because of the growing "Westernization" of the LDC populations (Reddy and Yusuf 1998). They are adopting less healthful diets, becoming less physically active, and beginning to use tobacco products in great numbers. These lifestyle changes have led to a widespread rise in their health indicators; their blood pressure, body weight, blood sugar levels, and lipid concentrations are going up. In addition to cardiovascular diseases, these changes in lifestyle have led to an increase in the prevalence of cancer, chronic obstructive pulmonary disease, and diabetes. The chronic diseases are becoming increasingly prevalent in the LDC also because of the continuing rise in life expectation, which is contributing to the emerging trend of rapid population aging. Obesity, a contributing factor in heart disease, diabetes, and possibly some cancers, is also emerging as a health problem in the LDC as well as in the MDC.

Table 11.5 Ten leading causes of the disease burden (DALYs) for the less developed countries: 1990 (DALYs: Disability-adjusted life years. One DALY equals one lost year of healthy life)

	Number	Percent of all causes
All causes	1,218.2[a]	100.0[a]
1. Lower respiratory diseases	110.5	9.1
2. Diarrheal diseases	99.2	8.1
3. Conditions arising during the perinatal period	89.2	7.3
4. Unipolar major depression	41.0	3.4
5. Tuberculosis	37.9	3.1
6. Measles	36.5	3.0
7. Malaria	31.7	2.6
8. Ischemic heart disease	30.7	2.5
9. Congenital anomalies	29.4	2.4
10. Cerebrovascular disease	29.1	2.4

Source: Murray and Lopez (1996), Table 4, p. 26. Copyright © 1996 World Health Organization. Reprinted with permission

[a]Includes causes not shown

Whether we examine the data for 1990, 2001, or 2002, in each year the noncommunicable diseases have a lesser impact as causes of years of life lost prematurely and the burden of disease than as causes of death, while the communicable diseases, perinatal conditions, and nutritional deficiencies have an increased effect. (Compare Tables 11.4 and 11.5 for 1990; see Tables 11.6 and 11.7 for 2001, and compare Figs. 11.2 and 11.3 for 2002). According to the Global Burden of Disease (GBD) study of 1996, noncommunicable diseases accounted for 56% of the deaths in the world in 1990, but for 31% of the years of life lost prematurely. On the other hand, infectious and parasitic diseases made up about 27% of total deaths but about 39% of total DALYs. Similarly, injuries, which affect mainly young people, accounted for only 10% of the deaths, but for 15% of years of life lost prematurely. These differences result primarily from the fact that the noncommunicable diseases affect mainly older people and older persons have fewer additional years to live than younger persons, but the communicable diseases and injuries have the opposite age effects. For 2002 these relations appear again. (Compare Figs. 11.4 and 11.5.) For example, ischemic heart disease made up 13% of total deaths but only 4% of total DALYs in that year. In addition to the influence of the leading ages of impact of the

Table 11.6 Distribution of deaths and the burden of disease by cause for low- and middle-income countries and high income countries: 2001

	Deaths		Burden of disease[a]	
	Low/middle-income	High-income	Low/middle-income	High-income
All causes				
Total number (1,000)	48,351	7,891	1,386,709	149,161
Rate per 1,000 population	9.3	8.5	265.7	160.6
Age-standardized rate per 1,000[b]	11.4	5.0	281.7	128.2
Selected cause groups	*Percent of total for income class*			
I. *Communicable diseases, maternal and perinatal conditions, and nutritional deficiencies*	36.4	7.0	39.8	5.7
Tuberculosis	3.3	0.2	2.6	0.1
HIV/AIDS	5.3	0.3	5.1	0.4
Diarrheal diseases	3.7	<1	4.2	0.3
Measles	1.6	<1	1.7	<1
Malaria	2.5	0.0	2.9	<1
Lower respiratory infections	7.0	4.4	6.0	1.6
Perinatal conditions	5.1	0.4	6.4	0.9
Protein-energy malnutrition	0.5	0.1	1.1	<1
II. *Noncommunicable conditions*	53.8	87.0	48.9	86.7
Selected cancers[c]	4.7	12.3	2.5	7.6
Diabetes mellitus	1.6	2.6	1.1	2.8
Unipolar depressive disorders	<1	<1	3.1	5.6
Alcohol use disorders	0.1	0.3	0.8	2.8
Vision and hearing disorders	0.0	0.0	4.9	4.9
Heart disease[d]	13.4	18.9	5.9	9.1
Cerebrovascular disease	9.5	9.9	4.5	6.3
Chronic obstructive pulmonary disease	4.9	3.8	2.4	3.5
Cirrhosis of liver, nephritis, and nephrosis	2.5	2.9	1.7	2.0
Osteoarthritis	<1	<1	1.0	2.5

(continued)

Table 11.6 (continued)

	Deaths		Burden of disease[a]	
	Low/middle-income	High-income	Low/middle-income	High-income
Congenital anomalies	1.0	0.4	1.7	1.0
Alzheimer and other dementias	0.4	2.6	0.7	5.0
III. *Injuries*	9.8	6.0	11.2	7.5
Road traffic accidents	2.2	1.5	2.3	2.0
Falls, self-inflicted injuries, violence	3.3	2.8	3.6	3.2

Source: Lopez et al. (2006), Table 1.1, p. 8. Copyright © 2006 World Bank. Reprinted with permission

Some percentages may be slightly overstated because in a few cases individually reported causes have been combined

[a]DALYs, or disability-adjusted life years. DALYs here are DALYs (3.0) and represent the version of the DALY based on a 3% annual discount rate and uniform age weights

[b]Age-standardized using the WHO world standard population

[c]Cancers include stomach, colon, rectal, liver, and trachea, bronchus, and lung cancers

[d]Hypertensive heart disease and ischemic heart disease

Table 11.7 Twelve leading causes of death for low/middle income countries and high income countries: 2001

Cause of death	Low/middle-income	Cause of death	High-income
Ischemic heart disease	11.8	Ischemic heart disease	17.3
Cerebrovascular disease	9.5	Cancers[a]	12.3
Lower respiratory infections	7.0	Cerebrovascular disease	9.9
HIV/AIDS	5.3	Lower respiratory infections	4.4
Perinatal conditions	5.1	Chronic obstructive	
Chronic obstructive		Pulmonary disease	3.8
Pulmonary disease	4.9	Diabetes mellitus	2.6
Cancers[a]	4.7	Alzheimer and other dementias	2.6
Diarrheal diseases	3.7	Self-inflicted injuries	1.6
Tuberculosis	3.3	Hypertensive heart disease	1.6
Malaria	2.5	Road traffic accidents	1.5
Road traffic accidents	2.2	Cirrhosis of the liver	1.5
Measles	1.6	Nephritis and nephrosis	1.4

Source: Based on Table 1.1, page 8 in Lopez et al. (eds.) (2006), Chapter 1. See also Table 3.6 in this volume. Copyright © 2006 World Bank. Reprinted with permission

[a]Four leading causes of cancer only

various health conditions, the duration of the illnesses plays a role in accounting for these differences in the proportions of deaths and years of life lost prematurely.

A leading conclusion of the GBD studies and their sequelae is that sense organ disorders, injuries, and particularly neuropsychiatric illnesses, such as depression

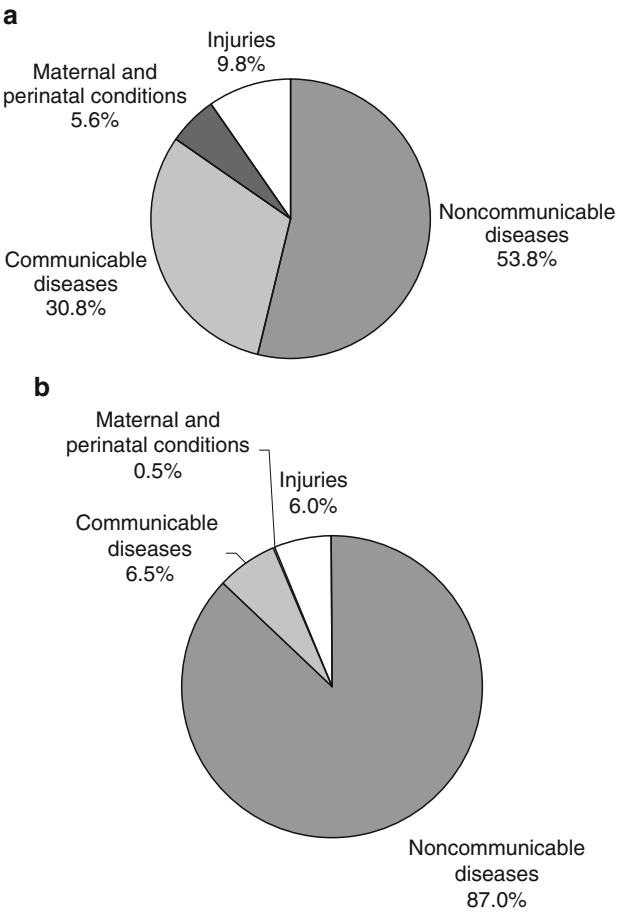


Fig. 11.2 (**a**) Distribution of deaths in the world for low- and middle-income countries by causal group: 2002; (**b**) Distribution of deaths in the world for high-income countries by causal group: 2002 (Source: Based on World Health Organization (2004), Annex Table 2. Used with permission of the World Health Organization)

and bipolar disorder, have been seriously underestimated as contributors to the worldwide burden of disease, especially in the LDC. This is because investigations have traditionally focused on deaths and not disability. Neuropsychiatric disorders were responsible for only a little over 1% of deaths in the world in 1990, but accounted for 11% of the burden of disease in that year. The WHO study showed higher results for 2002 – 2% and 13%. When the various diseases in the LDC were ranked on the basis of disability-adjusted life years (DALYs) for 1990, unipolar major depression was fourth among the causes of the total disease burden (Table 11.5). When, however, disability years (YLDs) only were considered, rather than DALYs (using data for the world as a proxy for the LDC), the list of the

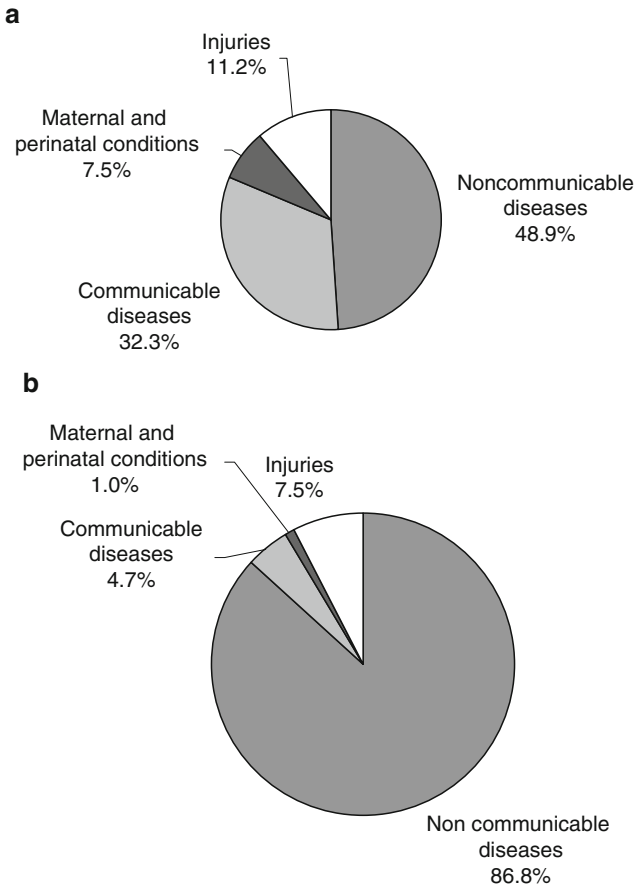


Fig. 11.3 (**a**) Distribution of the burden of disease in the world for low- and middle-income countries by causal group: 2002; (**b**) Distribution of the burden of disease in the world for high-income countries by causal group: 2002 (Note: DALY, or disability-adjusted life years, is a summary measure of population health that combines the number of years of life lost to a premature death and the number of years of healthy life lost to a disability; Source: Based on World Health Organization (2004), Annex Table 4. Used with permission of the World Health Organization)

ten most disabling diseases included five mental disorders, ranked as indicated: depression (1), alcohol abuse (4), bipolar disorder (6), schizophrenia (9), and obsessive-compulsive disorder (10). (See Murray and Lopez 1996.) In the MDC, eight of the top ten disability-causing diseases were mental disorders. Again, the GBD-study data for 2001 showed similar results for the regions when considered in two broad income groups (Lopez et al. 2006).

The GBD and WHO studies underscore the current and prospective importance of neuropsychiatric disorders in the total disease burden. By 2020 Murray and Lopez (1996) expect a major shift to occur in the diseases holding the first three places among the leading causes of the disease burden in the world. The burden

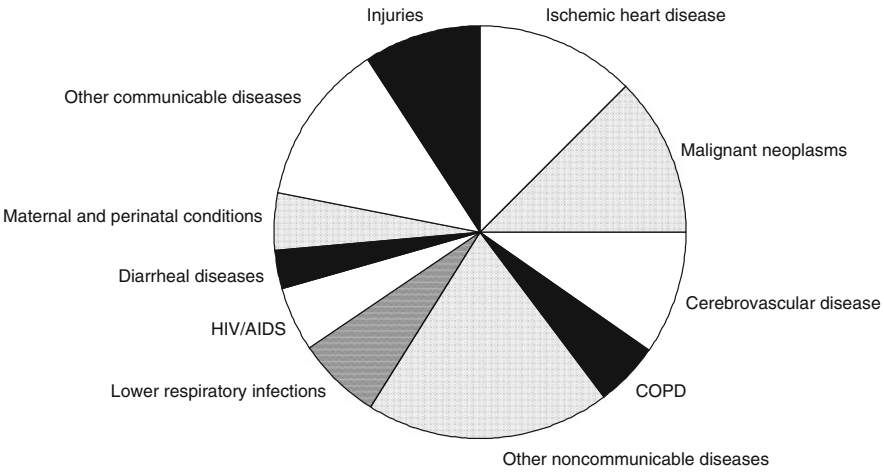


Fig. 11.4 Global deaths by leading cause: 2002 (Note: UNAIDS/WHO revised the estimate of the number of deaths from HIV/AIDS. Its current estimate is 1.9 million rather than the earlier 2.8 million. It is not known whether the difference should be removed from the total number of deaths or assigned to another cause; Source: Based on World Health Organization (2004), Annex Table 2. Used with permission of the World Health Organization)

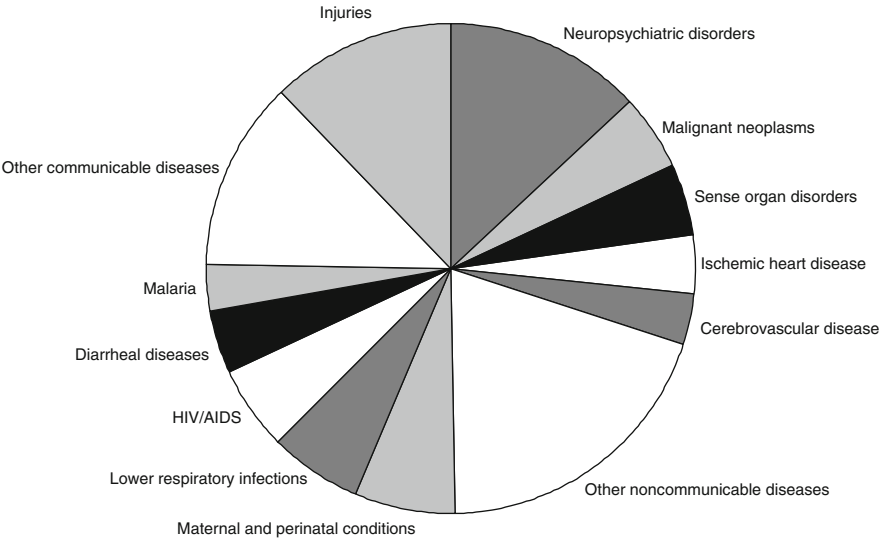


Fig. 11.5 Global burden of disease by leading cause: 2002 (Notes: DALY, or disability-adjusted life years, is a summary measure of population health that combines the number of years of life lost to a premature death and the number of years of healthy life lost to a disability. UNAIDS/WHO revised the estimate of the number of deaths from HIV/AIDS. Its current estimate is 1.9 million rather than the earlier 2.8 million. It is not known whether the difference should be removed from the total number of deaths or assigned to another cause; Source: Based on World Health Organization (2004), Annex Table 3. Used with permission of the World Health Organization)

of noncommunicable diseases is expected to rise sharply during these years, as is that for injuries, with the result that ischemic heart disease, depression, and traffic accidents are expected to replace lower respiratory infections, diarrheal diseases, and perinatal conditions, the previous incumbents of these ranks.

For their more recent projections of global mortality and the global burden of disease from 2002 to 2030, Mathers and Loncar (2006) built on the WHO estimates for 2002, using the method of structural modeling to make their projections. They too found a major shift in the distribution of deaths from communicable, maternal/perinatal, and nutritional causes to noncommunicable causes. In their middle series the proportion of deaths due to noncommunicable diseases is projected to rise from 59% in 2002 to 69% in 2030. The three leading causes of the burden of disease are expected to include two noncommunicable causes in addition to HIV/AIDS, namely, unipolar depressive disorders and ischemic heart disease. Fourth in the list is road traffic accidents. As Mathers and Loncar indicate, the projections enable us to visualize the implications of current population and health trends, such the aging of the population, the progress of the epidemiological transition, and the HIV/AIDS epidemic in the LDC for the future health situation of the world.

Communicable Diseases

The leading communicable diseases in the world in 1990 were lower respiratory infections, diarrheal diseases, tuberculosis, measles, and malaria (Murray and Lopez 1996). These diseases led as causes of death and also as causes of the burden of disease, as measured by DALYs – disability-adjusted life years.[4] When the data are shown separately for the MDC and the LDC, the diseases listed are leading causes only for the LDC and not for the MDC. The LDC are more vulnerable to communicable diseases such as measles and diphtheria, whooping cough (pertussis), and tetanus (i.e., the triad labeled DPT) because of the generally poorer sanitary conditions, lack of clean water, and lower levels of immunization.[5]

HIV/AIDS and Other Sexually Transmitted Diseases

Sexually transmitted diseases, including HIV/AIDS, are common throughout the world, especially in the less developed countries. According to UNAIDS/WHO, at the end of 2007 an estimated 33.2 million persons were infected with HIV/AIDS.

[4] AIDS does not appear as a leading cause of death or a leading cause of the burden of disease in the 1990 list of the GBD study. It does appear in the 2002 WHO list and the 2020 and 2030 lists of causes as projected.

[5] Informative summaries of the world situation with respect to the prevalence of infectious diseases and their trends are given in Olshansky and Carnes (1997) and in Kent and Yin (2006).

Table 11.8 Estimated HIV/AIDS cases for world regions: Year-end 2007

Region	Number (millions)	Percent of world total
World total	33.2	100.0
Sub-Saharan Africa	22.5	67.8
South/Southeast Asia	4.0	12.0
Latin America/Caribbean	1.8	5.4
Eastern Europe/Central Asia	1.6	4.8
North America	1.3	3.9
East Asia/Oceania	0.9	2.7
Western/Central Europe	0.8	2.4
North Africa/Middle East	0.4	1.2

Source: UNAIDS/WHO 2007

About 2.5 million persons became infected during the year, and about 2.1 millions persons died of the disease during the year. Since a cure for HIV/AIDS does not exist, these figures imply a net addition of 0.4 million cases to the world total living with the disease during the year. The estimates are derived from population-based surveys, sentinel surveillance at sites in selected countries, and mathematical models representing the natural history of untreated HIV infection in low- and middle-income countries.

Geographic distribution. After HIV/AIDS emerged in the West in the late 1970s and early 1980s, the virus spread widely to all parts of the world and pandemics developed in many regions, especially in sub-Saharan Africa. Of the estimated 33.2 million people in the world living with HIV/AIDS at year-end 2007, more than 90% reside in the LDC (Table 11.8). Two-thirds (68%) live in sub-Saharan Africa and one-eighth (12%) live in South Asia and Southeast Asia. Latin America and the Caribbean countries also show a high incidence of AIDS, falling just behind South Asia. Most of the earlier cases occurred in Africa, but most of the newer cases have been occurring in Asia, the Caribbean, South America, and Eastern Europe. Although the number of deaths from HIV/AIDS has been falling in several regions and the number of new infections has been falling in much of the world, the prevalence ratio for special groups in some regions, for example, intravenous drug users in Russia, Ukraine, India, and China, is growing. The epicenter of the epidemic, however, remains in sub-Saharan Africa, where nearly 90% of children infected with HIV live and more than three out of four (76%) deaths due to AIDS occur.

The countries of sub-Saharan Africa most affected by HIV/AIDS in 2007 are Botswana, Lesotho, Malawi, South Africa, Swaziland, and Zimbabwe (Table 11.9). According to UNAIDS/WHO (2007), Swaziland has the highest HIV/AIDS prevalence ratio in the world; over one-quarter of the adults 15–49 years in Swaziland are estimated to be HIV-positive. The other countries of southern Africa – Botswana, Lesotho, and South Africa – all have extremely high infection rates, ranging from 16% to 26% of adults 15–49 years of age. With an AIDS/HIV prevalence ratio

Table 11.9 Percent of persons 15–49 years of age infected with HIV/AIDS for world regions, 2007 and 2001, and selected countries in sub-Saharan Africa and other regions, 2004–2006

World Regions	2007	2001	
Total	0.8	0.8	
Sub-Saharan Africa	5.0	5.8	
Middle East and North Africa	0.3	0.3	
South and South-East Asia	0.3	0.3	
East Asia	0.1	<0.1	
Oceania	0.4	0.2	
Latin America	0.5	0.4	
Caribbean	1.0	1.0	
Eastern Europe and Central Asia	0.9	0.4	
Western and Central Europe	0.3	0.2	
North America	0.6	0.6	
Sub-Saharan Africa		Asia	
Benin (2006)	1.2	Cambodia (2005)	0.6
Botswana (2004)	25.2	India (2005–2006)	0.3
Cameroon (2004)	5.5	Indonesia (2005)	0.1
Central African Republic (2006)	6.2	China (2005)	0.1
Chad (2005)	3.3	Pakistan (2005)	0.1
Côte d'Ivoire (2005)	4.7		
Ethiopia (2005)	1.4	Eastern Europe	
Guinea (2005)	1.5	Russian Federation (2005)	1.1
Lesotho (2004)	23.5	Ukraine (2005)	1.4
Malawi (2004)	12.7		
South Africa (2005)	16.2	Caribbean	
Swaziland (2006–2007)	25.9	Haiti (2005–2006)	2.2
Uganda (2004–2005)	7.1		
Zimbabwe (2005–2006)	18.1		

Sources: Regions: Estimates for Dec. 2007 by UNAIDS/WHO (2007), *AIDS Epidemic Update*; Countries of sub-Saharan Africa, Cambodia, India, and Haiti from population-based HIV surveys, UNAIDS/WHO 2007 *AIDS Epidemic Update*; Other countries: Estimates for Dec. 2005 by UNAIDS. Reprinted with permission of UNAIDS

of 16%, South Africa's figure was at the lower end of the array of figures for the countries of southern Africa, but it has the largest number of residents infected with HIV/AIDS. Its 7.8 million cases may be compared to India's far smaller number of 3.2 million cases. In India, with a prevalence ratio of 0.3%, the disease is very thinly spread, much more so than in South Africa. China's prevalence ratio is still very low at 0.1% but, given its massive population, over one million people are infected. The disease is spread very thinly in China as in India, but the number of cases is growing rapidly there as in India.

The 2007 UN year-end survey of HIV/AIDS showed only a slightly greater number of persons in the world living with the disease than in 2006 (about 0.4 or 0.5 million), but a considerably larger number than at year-end 2001 (4.2 million).

During this period HIV/AIDS prevalence declined in many African countries, mainly in eastern Africa. On the other hand, the countries of southern Africa, South and Southeast Asia, East Asia, and Eastern Europe and Central Asia have made little progress in stemming the tide of the disease.

Some demographic characteristics of HIV/AIDS cases. Originally, in the Western countries, HIV/AIDS was primarily a condition of young homosexual men, but it spread into the heterosexual population. Now generally throughout the world, and in Africa in particular, married men are as likely to acquire the disease as homosexual men, and women are as likely to acquire the disease as men. The world total for the infected population in 2007 includes about 2.5 million children under 15 years old. These children mostly acquired HIV through their mothers before or at birth, or through breastfeeding. Breastfeeding accounts for about half of the new cases among children in sub-Saharan Africa.

Many countries of sub-Saharan Africa, especially in southern Africa, have continuing epidemics of HIV/AIDS. In these countries the disease is concentrated among segments of the population that are most at risk, such as men who have sex with men, persons who have multiple partners, needle-sharing intravenous drug users, sex workers, and their clients. The disease has spread rapidly in these countries because it is common to have multiple and concurrent sexual partners, male circumcision and condom use are relatively rare, and other sexually transmitted diseases are quite prevalent–a situation that serves to increase the risk of HIV transmission (Bongaarts et al. 2008). Men and women living in monogamous unions or without sexual partners are least at risk. High-risk behavior remains pervasive in the countries of sub-Saharan Africa and preventive efforts have been only moderately effective.

The general historical pattern by which HIV/AIDS spread over the last few decades began with a slow diffusion of the infected cases in the early 1980s; then followed a period of rapid expansion; and now the course of the virus is experiencing a relatively stable level or a plateau. A plateau is reached when the virus is concentrated among small but select, vulnerable segments of the population, such as those identified above. In its expansion the virus spreads quickly among the groups of highest risk but slows when a substantial portion of these persons become infected or die, while low-risk persons remain uninfected. The social diffusion process reaches a plateau when the vulnerable groups have been penetrated to the maximum. The disease typically takes hold in the large cities and then spreads out into the surrounding rural areas. This geographic diffusion process tends eventually to result in a rather even prevalence of the disease in a country.

The HIV/AIDS epidemic shows signs of leveling off although the disease can be considered still in its growing stages (Bongaarts et al. 2008). The global prevalence ratio of HIV infection is stabilizing everywhere except in Eastern Europe and Central Asia, the annual number of new infections globally is falling, and as a result of improved access to treatment, HIV-associated deaths have been falling in many regions. However, the global number of persons living with AIDS is rising because of the extension of the survival periods of new infections and the general growth

of the population. The disease will continue to spread before the number of cases levels off. Since there have been 2.5 million or more new infections annually in recent years, the death toll is likely to remain high. As of the end of 2007 more than 75 million persons have been infected by HIV in the world since the disease was first identified.

Demographic effects of HIV/AIDS. HIV/AIDS has had an enormous negative impact on life expectation and population growth in several countries of sub-Saharan Africa and it has begun to exact a similar toll in other regions of the world. By the end of 2005 25 million lives have been lost to AIDS in the world and, of these, about three-fifths have been lost to AIDS in sub-Saharan Africa. In 2007 some 2.1 million people died of HIV/AIDS in the world and of these about three-quarters (1.6 million) died in sub-Saharan Africa. The extremely low level of life expectancy – below 50 – in Malawi, Zambia, Angola, and other Sub-Saharan African countries is largely a direct consequence of the ravages of HIV/AIDS (Table 11.2). Life expectancy has fallen by as much as 20 years in some sub-Saharan African countries (e.g., Lesotho, Zimbabwe, Botswana, and Swaziland) and, as a result, the longevity gains of most of the last century in these countries have been largely lost (United Nations 2007; Fig. 11.6).[6] In Swaziland and Zimbabwe life expectancy at birth has fallen back to its pre-1940 level of about 40–45 years. No other region of sub-Saharan Africa has experienced as large a loss in life expectancy as southern Africa although a few countries of eastern Africa and central Africa have had similarly sharp losses (Fig. 11.7 and Table 11.2).

As a demographic influence, the AIDS epidemic has a direct effect on the death rate and life expectancy and indirect and smaller effects on fertility and migration. AIDS mortality tends to select young men and women who are fathers, mothers, and potential parents, and hence, the birth rate is reduced. However, the effect on the death rate is greater than the effect on the birth rate and the growth rate is reduced. Population growth has been "stunted" in many countries although never enough to cause a decline in population. The differences between the AIDS- and no-AIDS scenarios on a regional basis are greatest in sub-Saharan Africa. Consider the case of South Africa, where the number of cases of HIV/AIDS has been especially high. According to the United Nations, as of 2005 the population would have been 5% higher if the AIDS epidemic had not occurred (Fig. 11.8). The deficit is expected to grow rapidly in the next few decades so that by 2025 the population would have been 25% greater if the HIV/AIDS epidemic had not occurred.

A scourge of such a magnitude resulting from a sexually transmitted disease cannot fail to have other pervasive demographic consequences, affecting the balance of the sexes, the age-sex structure of the population, family composition, marriage,

[6]The United Nations prepared a "No-AIDS" scenario as part of its report on 2006 World Population Prospects (United Nations 2007). The No-AIDS scenario is the medium projection variant modified to include the additional deaths in the population total that result when the mortality rates of uninfected individuals are applied to the entire population, including the affected individuals.

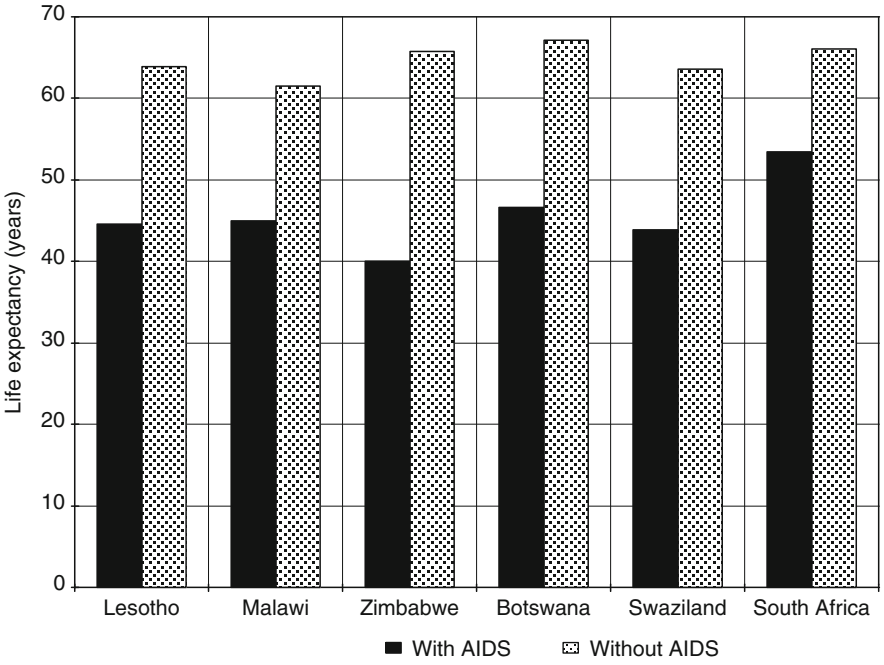


Fig. 11.6 Effect of AIDS on life expectancy in selected countries: 2006 (Source: United Nations, Population Division (2007))

parent-child relations, and sexual practices. Most of the victims have been young men, so that there has been a rapid increase in female-headed households. The deaths of parents from AIDS have produced millions of orphans. Some 15.4 million children (under 18 years of age) were orphaned by the disease as of 2007. Of this total, some 12 million were created by the deaths of adults from HIV/AIDS in sub-Saharan Africa. Because the number of AIDS cases is expected to grow, we may expect these effects to intensify.

The effect of the AIDS epidemic on the age distribution is still very small in all regions except sub-Saharan Africa. In that region the age range bearing the greatest impact shifts as the phases of the AIDS epidemic change. In general, however, it is the childbearing ages and the earliest childhood ages that are experiencing the greatest toll at present and will do so in the coming few decades.

HIV/AIDS has been a powerful agent in causing high death rates from malaria, tuberculosis, pneumonia, and other infectious diseases. It weakens the immune system until the immune system can no longer fight infection, leaving the way open for opportunistic infections to take hold. An estimated 22% of the TB cases in Africa occur in people living with AIDS. Conversely, areas where infectious and parasitic diseases are endemic are more hospitable for AIDS epidemics to take hold.

Management of HIV/AIDS. Prevention is the best policy for managing HIV/AIDS since there is no cure for this disease. Prevention programs for reducing HIV

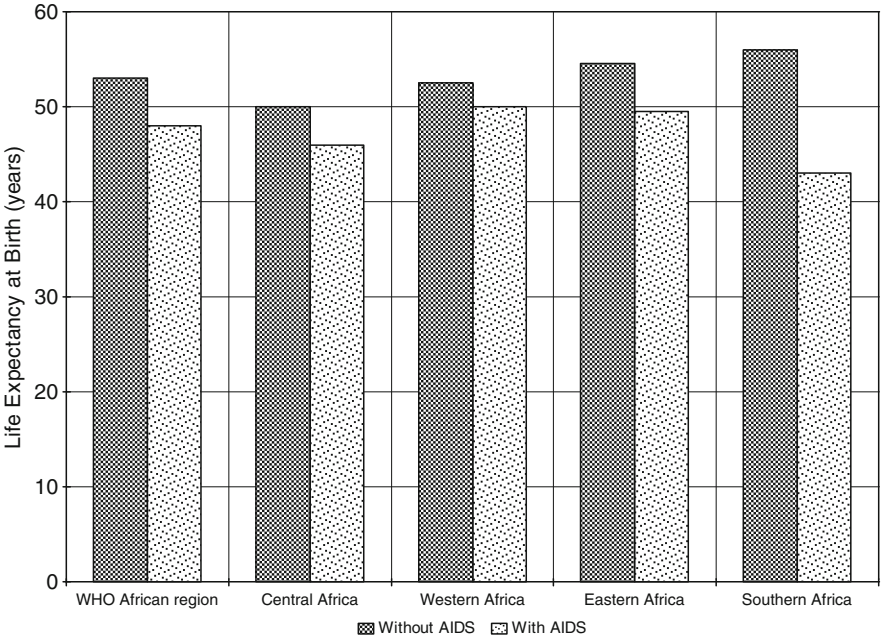


Fig. 11.7 Life expectancy with and without AIDS: Regions of Africa, 2002 (Source: Based on World Health Organization (2004), Figure 1.4, page 7. Reprinted with permission of the World Health Organization)

incidence have been only moderately successful, however, and a large number of adolescents and adults is still likely to become infected (Bongaarts et al. 2008). The commonest basis of infection (about three-quarters of the time) is unprotected sexual intercourse, followed far less commonly by injecting drug use and blood transfusions, so that the condom is viewed as the primary means of preventing transmission of the disease. Although sub-Saharan Africa has the highest rates of HIV/AIDS, the condom is infrequently used there, in spite of the tremendous efforts to encourage its use. Condom use is apparently more common among casual sexual partners and sex workers than among spouses and established partners. Poverty may account for the limited use of the condom among regular partners for preventing transmittal of AIDS. Moreover, there are cultural and religious beliefs in the LDC that limit the use of condoms. Millions of women cannot protect themselves from HIV infection because they are powerless to demand use of a condom or refuse sex intercourse (Lamptey et al. 2006). They are at risk of HIV infection from husbands who have other sexual partners. They could, however, use microbicides if their partners do not use condoms – and without knowledge of their sexual partners – but a safe and effective microbicide has yet to be developed. Clinical trials of microbicides and antiretroviral prophylactic agents are under way.

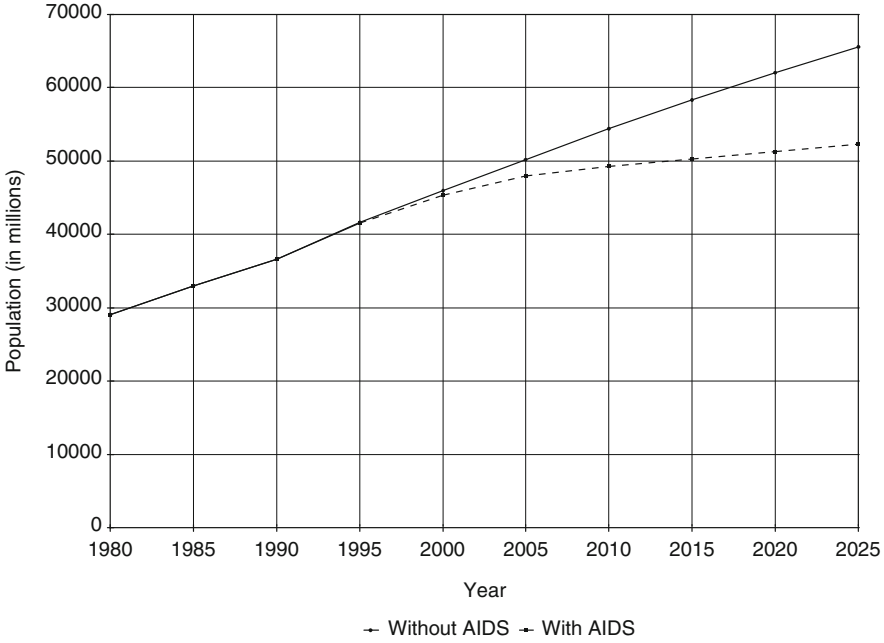


Fig. 11.8 South Africa's population without and with AIDS: 1980–2005 (Source: United Nations, Population Division (2007). Reprinted with permission of the Population Division, United Nations)

More effective means have to be found to promote use of the condom by people who are at high risk of infection from AIDS but who do not understand the risks, such as adolescents. The A (abstinence), B (be faithful), C (condom use) policy in Ghana has had some success, but it is important to realize that a moralistic policy that depends only on A and B, without C, is destined to fail. Other intervention strategies have to be found that might be effective in reducing disease transmission. One practice that is currently in clinical trials and shows early signs of success is circumcision of adult men. This surgical operation does not provide full protection against the disease but infection ratios among young men who have been circumcised appear to be sharply reduced as compared with uncircumcised men.

It has become clear that HIV/AIDS in Africa is a socioeconomic and cultural problem and not a medical problem. Millions of people in South Africa infected by HIV reject the use of the antiretroviral medications that have worked so well in reducing mortality from the infection in the West. They favor some form of traditional medicine. In rural South Africa the local healer is a prominent and respected figure whom many villagers consult for health problems and are likely to visit for medical treatment of HIV/AIDS. In this practice, they are influenced by concern for the adverse side-effects caused by the medications and by earlier government announcements that AIDS is not caused by HIV infection. In sub-Saharan Africa only about 10% of the persons afflicted with HIV/AIDS are

receiving treatment. This low participation is, in part, due to the shortage of inexpensive antiretroviral drugs, the lack of medical resources, the disorganized state of the health systems, lack of dedication by many governments to the AIDS problem, and the stigma attached to the disease. Sex is a major taboo and subjects like sex and condoms cannot be talked about openly in this region. In such male-dominated societies, men have the last word about sex including how, when, and where to have sex, and they generally refuse to use condoms. Ways have to be found to amplify the effectiveness of the limited health personnel available, possibly by use of the newer information technology. Contacts by cell phones could be used to remind patients to renew prescriptions, keep appointments with health workers, and observe appropriate hygiene and safe sex practices. Botswana, about one-quarter of whose adult population had HIV as recently as 2006, has made considerable progress in expanding treatment to HIV/AIDS patients by developing a medical infrastructure to effectively dispense antiretroviral drugs and, with the help of foreign medical teams, by establishing numerous HIV clinics throughout the country.

Most of the people who become infected and die from the disease now do so because of inadequate access to HIV prevention and treatment services. The relative ability of the more developed countries and Africa to manage the HIV/AIDS problem is suggested by a comparison of the ratio of the number of deaths to the number of cases under treatment in the MDC and Africa in 2002:

	Treated	Deaths	Ratio
More developed countries	500,000	25,000	.05
Africa	30,000	2.2 million	73.3

Source: UNAIDS. (2002). Executive Director, Dr. Peter Piot, BBC online, July 2

The difference in the ratios is striking and point to the vastness and urgency of the problem.

The development of antiretroviral medications has transformed some acute conditions like HIV/AIDS into chronic conditions, so that the patients may live a long time after the initial infection before the infection and its symptoms compromise and then destroy the integrity of one or more of the patient's vital organs. Thus, unlike many other infectious diseases, which are acute conditions that result in early recovery or early death, HIV/AIDS has been converted by modern medicine into a chronic infectious condition.

The greater volume of funds dedicated to AIDS management and the greater acknowledgment of the problem by national governments, international agencies, and private foundations suggest that control of the disease is likely to make more rapid progress in the future. Progress is linked to the strength of the national economy, an effective public health system and local health infrastructure, including a corps of trained nurses and physicians, empowerment and independence of women, increased literacy, and dedicated resources. Improvement in managing the

HIV/AIDS problem will contribute to management of other major health problems, such as tuberculosis, malaria, lower respiratory infections, measles, and maternal and child health.

Other sexually transmitted diseases. There are several other sexually transmitted diseases common in the less developed areas of the world: Syphilis, gonorrhea, chlamydia, chancroid, genital warts, and hepatitis A, B, and C. (Unlike the other diseases, hepatitis is usually transmitted in other ways than sexual intercourse.) Chlamydia figured among the 15 leading causes of the disease burden for women of childbearing ages in the LDC in 1990 (Murray and Lopez 1996). Some of these conditions are risk factors for other diseases. Genital warts results from the extremely common human papillomavirus (HPV) and can lead to cervical cancer. Only in some cases does infection by the papillomavirus lead to cancer, however. Cervical cancer kills about 250,000 women worldwide.

Other Common Infectious Diseases

Lower respiratory infections, diarrheal diseases, malaria, tuberculosis, and measles and other childhood diseases are common infectious diseases in the LDC. The five named diseases were among the ten leading causes of death (Table 11.4) and among the 10 leading causes of the disease burden (DALY) in the LDC in 1990 (Table 11.5). In that year, in general, these diseases moved up a rank or two as causes of the disease burden from their rank as causes of death. It is an indication of the force of the epidemiological transition that by 1990 only one communicable disease, lower respiratory infections, appeared among the top ten causes of death in the MDC and that three chronic noncommunicable diseases appeared among the top ten causes of death in the LDC.

According to the WHO study for 2002, the total contribution of the communicable, maternal, perinatal, and nutritional conditions to the total death toll tended to remain about the same in 2002 as in 1990 in the LDC, accounting for about one-third of all deaths (Fig. 11.2 and Table 11.6). While the share for most leading communicable causes declined in this period, HIV/AIDS appears prominently in the WHO list for 2002 for the first time, accounting for 5% of all deaths. Otherwise, the rank order of the leading communicable diseases among themselves remained roughly unchanged.

In the WHO study the ten leading causes of the disease burden (i.e., disability-adjusted life years) in the world included lower respiratory-tract infections as first in rank, then HIV-AIDS as second in rank, diarrheal diseases as fourth, malaria, and tuberculosis. If the 13 major "neglected tropical diseases" were listed as a single group, they would rank sixth (Hotez et al. 2007). The seven most prevalent ones are ascariasis, trichuriasis, hookworm infection, schistosomiasis, lymphatic filiariasis, trachoma, and onchocerciasis. In the following sections diarrheal diseases, tuberculosis, malaria, and a few tropical diseases are briefly considered in order to illustrate the nature and scope of the problem of infectious and parasitic diseases in the LDC.

Diarrheal diseases. While diarrheal diseases are still an important cause of death in the LDC, deaths from this cause have fallen sharply in recent decades. The immediate cause of death when diarrheal diseases strike is dehydration. Major public health drives have been launched to prevent and treat dehydration and they have been succeeding. In just a dozen years, from 1990 to 2002, the number of deaths from diarrheal diseases was cut from 2.9 million to 1.8 million, or by 39%, and the burden of disease was cut from 99 million DALYs to 62 million DALYs, or by 38%.

Diarrheal diseases are mainly a condition of young children and they accounted for 17% of all childhood deaths as of 2002. Almost all of these deaths occur in the LDC, especially in sub-Saharan Africa and South and Southeast Asia. The prevalence of diarrheal diseases has remained at a high level even while mortality from diarrheal diseases has fallen. Diarrheal sickness rates are highest among infants 6–11 months, a period when breastfeeding can be part of the treatment, along with normal feeding and appropriate immunizations (Fontaine and Boschi-Pinto 2006). In dehydrating the body, diarrhea causes loss of electrolytes (sodium, chloride, potassium, and bicarbonate). While a dehydrated child at first may show no symptoms, with increasing dehydration he or she may exhibit symptoms affecting behavior, appearance, and physiological responses. Dehydration may lead to shock, associated with a feeble pulse and low or no blood pressure; and with these, a young child can succumb quickly. Diarrheal episodes are also associated with weight loss and stunting (i.e., low height for age) because of eating less food and absorbing fewer nutrients at a time when the child has increased nutrient requirements. Early childhood undernutrition magnifies the risk of poorer physical fitness, poorer health, and lower work productivity in later life.

Most deaths attributed to diarrhea are linked to malnutrition, most commonly in the form of undernutrition, and they are mutually reinforcing as cause and effect. The underlying circumstances predisposing to the disease are poverty, associated as it is with poor sanitation, poor access to clean water, improper hygiene, poor housing, and crowding. The exposure to an unhealthy environment is often accompanied by weakened resistance to disease. The lack of access to necessary health providers exacerbates the situation. These conditions are characteristic of the poorest segment of the population of the LDC, especially in sub-Saharan Africa.

Treatment of dehydration, even in its acute stage, is easy, with administration of increased fluids plus continued feeding, including increased frequency of breastfeeding in the earliest months of infancy accompanied by regular feeding in the later months of infancy, oral rehydration salts solution, and zinc and vitamin supplementation. Apart from a carefully designed and implemented protocol for treatment of individual patients, community action to improve the supply of safe water, install clean sanitation facilities, and educate households on personal and domestic hygiene would be part of a total prevention program. In spite of the fact that death from dehydration is preventable, it remains a leading cause of death among young children in the LDC and, even more, a leading cause of sickness among them.

Tuberculosis. One third of the world's population is infected with either the passive (latent) or active forms of tuberculosis. Some 4 million new cases of the most infectious form of the disease appeared in 2002. In that year there were an estimated 1.6 million deaths from tuberculosis, reflecting a modest decline from the 1.9 million in 1990. The WHO/World Bank (Murray and Lopez 1996) ranked tuberculosis in 1990 as sixth among the leading causes of death and fifth among the leading causes of the burden of disease in the LDC. With the rise of the noncommunicable diseases, TB fell back to the tenth and ninth places, respectively, in 2002.

TB is particularly difficult to diagnose because most persons infected with it have the latent form of the disease, are asymptomatic, and do not spread the disease. The victim can spread the disease only after he or she becomes sick; then, if untreated, the person with active TB can propel TB bacilli through the air and infect other persons. If the victim is also infected with HIV, as is not uncommon in the LDC, TB is even harder to diagnose. Although persons infected with the latent form of TB have less than a one-fifth chance of acquiring the active form, this risk does not progress randomly to the active form. Lack of treatment is a major factor in hastening this progression. In spite of the availability of good methods of treatment, only about one-third of TB cases receive proper treatment. This shortfall may be attributed to limited resources, ineffective health care systems, and inadequate supplies of modern drugs.

Given that a stigma is attached to HIV/AIDS and TB is a risk factor for HIV/AIDS and often accompanies it, it is possible that TB is over reported as a cause of death and AIDS is correspondingly underreported. Inasmuch as most persons with AIDS in the LDC do not die in hospitals, the cause of death is usually obtained from reports of caretakers of decedents. They sometimes choose to misreport the actual cause of death, and as a result, the cause-of-death data are biased in favor of the companion disease, TB. Infection by HIV is also a risk factor for TB and hastens its progress by weakening the victim's immune system. Thereby, HIV contributes to the disease's progression to the active form and increases the likelihood of death. From this perspective HIV/AIDS may be overstated as a cause of death at the expense of TB.

The vast majority of cases living with TB – some 95% – and of TB deaths – some 98% – occur in the LDC (Yin 2006). About one-third of newly diagnosed cases of TB in Africa are also infected with HIV. Because of TB's association with HIV/AIDS, the disease has spread most rapidly in sub-Saharan Africa and Eastern Europe. In "sync" with the waxing and waning of the HIV/AIDS epidemic, the most rapid expansion of the TB epidemic occurred during the 1990s and a contraction appears to have occurred during the 2000s.

The incidence of multidrug-resistant TB (MDR-TB) has increased sharply in recent years because patients do not follow drug-treatment regimens fully or consistently, the wrong drug may be prescribed by the health worker, or the supply of drugs may be undependable. This type of TB is much harder and far more expensive to treat than conventional TB. Progress in overcoming tuberculosis, as with HIV/AIDS, requires not only an adequate supply of economically priced

medicines but also a considerable degree of social and economic development and an effective public health system to get patients, the supply system, and the health system working successfully together.

Malaria. Malaria is another major cause of disease and death in the world. WHO has estimated that each year there are 300 million to 500 million new cases and some 1.3 million deaths in the world from this disease (WHO 2004). It is endemic in about 90 countries, half of which are in Africa (WHO 1993). About 90% of all malaria deaths in the world occur in Africa. Malaria accounts for one in five childhood deaths in Africa and 20–30% of all infant deaths (Yin 2006; Bawah and Binka 2005; Molyneaux 1985). Bawah and Binka (2005) cite a report of the Ghanaian Ministry of Health concluding that in Ghana about 25% of all deaths among children under age 5 are attributable to malaria.

Malaria is both a leading cause of death and a leading cause of the burden of disease in the world, especially Africa. In the original Global Burden of Disease (GBD) study, the WHO and the World Bank (Murray and Lopez 1996) ranked malaria as eleventh among the leading causes of death in the world, ninth among the leading causes of death in the LDC, and seventh among the leading causes of disability-adjusted life years in the LDC. That study provided an estimate of 32 million disability-adjusted life-years lost from premature death and disability as a result of malaria in the LDC in 1990. The WHO reported a sizeable increase in this measure to 46 million in 2002. Malaria has retained the rank of 11th among the causes of disability-adjusted life-years in the world over this period and so it remains a major public health world challenge.

Malaria is endemic only in tropical and subtropical areas of the world, specifically sub-Saharan Africa excluding southern Africa, the northern part of South America, Central America, part of the Caribbean, south and southeastern Asia, and part of the western Pacific. Within these areas malaria disproportionately victimizes the poorest segments of the population. With the continued rapid growth of the population in the endemic areas, the number of persons exposed to the risk of infection by the malaria parasite has grown even though the endemic area has been contracting (Yin 2006). For example, the explosion of the population of sub-Saharan Africa in the last few decades has added several hundred million people to the population at risk.

Malaria weakens the victim and confers on him/her an increased vulnerability to other diseases and an increased risk of death. The disease has special risks for pregnant women, resulting in miscarriages, perinatal deaths, low birth-weight babies, and even severe anemia and death.

The disease has placed a massive social and economic burden on the principal countries affected. The medical, economic, and social costs of malaria are colossal, especially in Africa. These costs take the form of millions of patient-days spent in hospitals and millions of work-days and school-days lost, and sharply reduced productivity and production at work and at home. Malaria is causing sub-Saharan Africa's gross domestic product to grow by some 1.3 percentage points less per year than it would if malaria were eliminated (WHO 2000).

Much effort has been expended in trying to control and even eliminate malaria in Africa, but these efforts have generally failed. Early successes in trying to control the mosquito population, the vector for the parasites that cause this disease, through large-scale spraying in the 1950s and 1960s in parts of Africa by DDT, the insecticide employed in eliminating mosquito breeding sites, have been frustrated by a number of factors. They include the emergence of new strains of malaria-carrying parasites that are resistant to the standard insecticide treatments and the standard drugs, concerns about the safety of and the eventual banning of DDT, and the breakdown of national control programs. In other areas of the world, where malaria was presumably eradicated, as in Asia and Latin America, it has reappeared because of the emergence of drug-resistant strains. Countries like Singapore and Malaysia, where economic conditions have been much more favorable, were able to eliminate the disease after a period of resurgence, but sub-Saharan Africa lacks the health infrastructure and resources to emulate this example. As drug resistance grows, the cost of treating patients has risen. Malaria is once again a major cause of disease throughout the world and therefore a major international public health problem.

The new approaches to the control of malaria include the reintroduction of DDT. It is used not for large-scale outdoor spraying but for controlled indoor spraying of homes and nets. The use of nets treated with a long-lasting toxin/insecticide may be a low-risk and cost-effective means of preventing the spread of malaria in the most susceptible areas.

Measuring the prevalence of malaria. We get some insight into the problem of measuring the prevalence of malaria in the communities where it is endemic, and the demographic consequences of its high incidence in such areas, from the following description of efforts to collect data on deaths in a rural district in Ghana and to analyze them. Bawah and Binka (2005) analyzed the mortality data for 1995 for the Kassena-Nankana district of Northern Ghana, a "hyperendemic" rural area severely afflicted with malaria. This area depends on subsistence farming, and has high levels of childhood mortality as a result of malnutrition, poverty, endemic infectious and parasitic diseases, and an inadequate health infrastructure. Bawah and Binka found that 44% of the deaths of this area were due to infectious and parasitic causes and surmised that at least one-third of all deaths were due to malaria.[7] After allocating a share of the deaths of unknown cause (36% of all deaths) to the reported causes, they arrived at a considerably higher estimate of the percentage of deaths as due to infectious and parasitic causes – 69% – and decided that most of these deaths should be attributed to malaria. They constructed a multiple decrement table to

[7]Cause-of-death data were collected at the Navrongo Health Research Center through the Navrongo Demographic Surveillance System. The data are for 1995 and are restricted to the rural segment of the population. All deaths were followed up with verbal autopsies to determine the cause of death. Three physicians coded the records independently and, when at least two of them agreed on a particular cause for a death, that cause was assigned as the most probable cause. A large share of the deaths, over one-third, were classified as of unknown cause.

estimate the total number of person-years that would have been saved had malaria been eliminated from the area, given the age- and cause-specific death rates of 1995. They found that, under the assumption that the causes of death are independent, life expectancy at birth would increase by more than 6 years if malaria was eliminated as a cause of death. The gain would be much greater – 9 years – if an appropriate proportion of the deaths of unknown cause were allocated to malaria.

Other diseases. According to the Global Burden of Disease study of 1996, a few other infectious diseases, in addition to HIV/AIDS, diarrheal diseases, tuberculosis, and malaria, were identified by the researchers among the (15) leading causes of the disease burden (DALYs) in the world in 1990 or expected to be leading causes in 2020. They were lower respiratory infections and measles. All of these diseases, except HIV/AIDS, were among the leading causes of death or premature death (YLLs) in 1990 and most were among the leading causes of disability in that year. HIV/AIDS appears in the list of leading causes in the new century.

A number of common bacterial, viral, and parasitic infections are risk factors for other major diseases that appear in the list of leading causes of death or the disease burden. Among these are hepatitis B, schistosomiasis, trachoma, onchoceriasis, and Chagas' disease. Infection with hepatitis B virus increases the risk of developing liver cancer and cirrhosis of the liver; about 400 million people around the world are chronically infected with hepatitis B virus. The last four of the diseases listed are among the so-called neglected tropical diseases. The neglected tropical diseases are among the most common infections in the less developed regions of the world. They occur primarily in rural areas and some poor urban areas in the poor countries.

Onchocerciasis ("river-blindnesss") causes nodular swellings on the skin, lesions of the eyes, and blindness. An estimated 37 million people are infected with onchocerciasis in the world and 90 million people live in tropical areas of the world threatened by it (Hotez et al. 2007). The disease is caused by infestation of the blood and lymphatic tissues with filarial nematode worms of the genus *Onchocerca.* Schistosomiasis, another tropical disease caused by a parasitic worm (the trematode worm of the genus *schistosoma),* has a global prevalence of some 207 million people. It is widespread in the rural areas of sub-Saharan Africa, Latin America, and the Caribbean. People catch the parasitic worm by using contaminated water in which snails carrying the schistisomes live. Eventually the disease affects various organs in the genito-urinary, gastro-intestinal, and nervous systems. Trachoma affects 84 million persons in sub-Saharan Africa, the Middle East, and North Africa, and Chagas' disease affects 8–9 million persons in Latin America and the Caribbean (Hotez et al. 2007).

Management, Trends, and Prospects

The LDC suffer from a lack of late-model drugs and vaccines to treat such tropical diseases. Diseases specific to poor countries do not generally inspire pharmaceutical companies to develop new drugs. According to the *OECD Observer*

(Parris, May 2004), only 13 of the 1,223 new medicines commercially produced between 1975 and 1997 were designed to treat tropical diseases. In spite of this, in the American continent outside U.S and Canada, the prevalence of some major infectious diseases have shown spectacular declines, and deaths from infectious diseases, in general, except for HIV/AIDS, have fallen steadily. Measles has nearly disappeared in Latin America although it continues to be an important cause of death of children elsewhere in the less developed world. Deaths from malaria are a small fraction of what they were a quarter century earlier; they have fallen from 8 deaths per 100,000 population to 2 deaths per 100,000 population.

According to the projections of the Global Burden of Disease studies, major declines in the relative importance of the leading communicable diseases in the world, with two major exceptions, as reflected in DALY scores, are expected to occur by 2030. Lower respiratory diseases, diarrheal diseases, measles, and malaria are expected to drop too much lower ranks. However, tuberculosis is expected to maintain its rank and HIV will enter the list and make a strong showing.

Leading Causes of Disability[8]

Disabled years (YLD) are a major part of the total burden of disease. The leading causes of disability tend to be different from the leading causes of death and the leading causes of years of life lost to premature death (YLL). This is especially evident in the case of the mental disorders, which, as we saw, are infrequent causes of death but are common causes of disability. When we focus on the (10) major causes of disability (YLD), as determined by the GBD studies, we see a rather different list of causes than those given in the list for causes of death and causes resulting in years of life lost to premature deaths (YLL). The disability list is dominated by mental disorders. A leading cause of disability in the world is unipolar major depression, which accounted for 11.5% of the total years lost to disability in 1990. The list of ten major causes of disability (YLD) also includes bipolar disorder, alcohol abuse, schizophrenia, and obsessive-compulsive disorders. These neuropsychiatric illnesses accounted for 22% of all years lived with a disability (YLDs), compared with only 1.9% of all deaths and 1.1% of years of life lost prematurely (YLLs). In the MDC, neuropsychiatric disorders account for an even greater proportion of years lost to disability. The burden of mental illness as a group is greatest in the high-income countries (i.e., mainly the group of industrialized countries of the West, Asia, and Oceania), but mental disorders are important causes of disability in the other regions as well (except sub-Saharan Africa). The Global Burden of Disease report by Mathers and Loncar (2006) predicts that

[8]Disability is defined here as a disease or injury, following the usage of the Global Burden of Disease studies, not as in any of the other technical senses used in other parts of this book.

neuropsychiatric disorders will continue to rise in importance and major depression will be the second leading cause of disability worldwide in 2030.

Again, to emphasize the point, while neuropsychiatric disorders account for a very small proportion of deaths in the world, they account for a substantial share of the disability in the world. If the burden-of-disease analysis had been confined to mortality alone, several major causes of disability would not have been included and would not have received the attention they merit in the total disease burden. The addition of a disability component to the mortality component in the DALY measure makes it possible to highlight the very important role of neuropsychiatric disorders in the total burden of disease.

We may attribute the difference in the extent of mental illness between the LDC and MDC either to differences in lifestyle and the associated greater stresses and complexities of life in advanced societies or to the less complete data for the LDC. The relative role of these two factors is unknown since research on and clinical intervention into mental illness in the LDC have been neglected because of the negative attitudes, prejudices, and stigma that surround these disorders in those regions of the world. A report of the National Academy of Sciences, Institute of Medicine (2001) on neuropsychiatric disorders in the LDC concludes that there is a considerable underestimation of the incidence of such conditions in the LDC and a neglect of their consequences both for the individual and society.

The list of ten leading causes of disability in the world also includes several other noncommunicable health conditions, namely, iron-deficiency anemia, falls, chronic obstructive pulmonary disease (COPD), congenital anomalies, and osteoarthritis, with shares of the total years lost to disability for each of these conditions ranging from 4.7% to 2.8%. We see in these leading causes of disability the effects of malnutrition, possibly due to infectious or parasitic disease (iron-deficiency anemia, congenital anomalies) and risky behavior (falls, road traffic accidents, COPD).

Leading Risk Factors Accounting for Disease, Injury, and Disability

A great variety of factors can be enumerated that at different levels of analysis and in different degrees contribute to the excessive burden of death and disease in poor societies. I highlight only some of them, those that have special application to the LDC, such as the often exclusive pursuit of traditional medicine, the close familial relation with animals, poor water supply and sanitation, and the marked food insecurity resulting from the extremes of weather and political instability. The GBD studies have explored this question as well. It has measured the contribution of a selected list of risk factors to the disease and injury burden in the world. These risk factors include: Malnutrition; poor water supply, sanitation, and personal/domestic hygiene; unsafe sex; substance abuse (that is, tobacco use, excessive alcohol use; illicit drug use); occupation (that is, exposure to hazards through work);

hypertension; physical inactivity; and air pollution. This list greatly resembles the list given in Chap. 6, which was designed to describe the conditions in the MDC. The present list differs mainly in featuring some factors more distinctive of the LDC, such as undernutrition, air pollution, unsafe sex, and poor water supply, sanitation, and personal hygiene.

Tables 11.10 and 11.11 present summaries of the findings of the GBD studies with respect to the consequences of these risk factors for the burden of disease and injury in the world. In the earlier GDB study six risk factors taken as a group accounted for more than one-third of the world's disease burden in 1990. These were: malnutrition; poor water, sanitation, and hygiene; unsafe sex; alcohol; tobacco; and occupational risks. Of these, malnutrition (16%) and poor sanitation (7%) accounted for almost one-quarter of the total. All other factors individually accounted for 3.5% or less.

As of 2001 undernutrition was still a leading risk factor (Lopez et al. 2006). Alcohol was also playing an important role in increasing the rates of death and disability from disease and injuries in regions such as Sub-Saharan Africa and Latin America and the Caribbean. An estimated 45% of global mortality and 36% of the global disease burden are attributable to the joint hazardous effects of the 19 risk factors studied. These risk factors have unequal effects on the sexes and vary among the regions of the world. For example, unsafe sex practices are of primary importance for young adult women in Sub-Saharan Africa, whereas tobacco use and alcohol abuse affect mostly men in the more developed countries. Because risk-taking sexual behavior is more common among the poor than among the non-poor in the LDC, HIV/AIDS and other sexually transmitted diseases are relatively more prevalent among the poor in these areas. The poor in the LDC are also less likely to participate in the antiretroviral treatment programs or other treatment programs for HIV/AIDS that are most likely to benefit them. This is because of the unavailability of the drugs or the programs, their cost, or the limited knowledge they have about them.

Tobacco products are being widely marketed in the LDC and these populations are taking up the use of these products in great numbers. In part, this explains why these countries are experiencing rapid increases in cardiovascular diseases, cancer, and other endogenous diseases. The upsurge in tobacco use in the LDC foreshadows a sharp rise in disease and deaths from this source in the next few decades. Tobacco-associated deaths are projected to rise from 5.4 million in 2005 to 8.3 million in 2030. Furthermore, tobacco use is projected to kill 50% more people in 2015 (6.4 million) than HIV/AIDS (4.3 million) and to be responsible for 10% of all deaths globally (Mathers and Loncar 2006).

Culture/Lifestyle, Poverty, and Disease

Culture and lifestyle factors affect the prevalence of disease by the ways that members of a society think about the origins of disease and try to treat it, their

Table 11.10 Global burden of disease and injury attributable to selected risk factors, by years of life lost prematurely and years lived with a disability: 1990

	Number (thousands)				Percent of total			
Risk Factor	DALYs	YLLs	YLDs	Deaths	DALYs	YLLs	YLDs	Deaths
Total	NA	NA	472,700	50,466	100.0	100.0	100.0	100.0
Malnutrition	219,575	199,486	20,089	5,881	15.9	22.0	4.2	11.7
Poor water supply, sanitation, and personal and domestic hygiene	93,392	85,520	7,872	2,668	6.8	9.4	1.7	5.3
Unsafe sex	48,702	27,602	21,100	1,095	3.5	3.0	4.5	2.2
Tobacco use	36,182	26,217	9,965	3,038	2.6	2.9	2.1	6.0
Alcohol use	47,687	19,287	28,400	774	3.5	2.1	6.0	1.5
Occupation	37,887	22,493	15,394	1,129	2.7	2.5	3.3	2.2
Hypertension	19,076	17,665	1,411	2,918	1.4	1.9	0.3	5.8
Physical inactivity	13,653	11,353	2,300	1,991	1.0	1.3	0.5	3.9
Illicit drug use	8,467	2,634	5,834	100	0.6	0.3	1.2	0.2
Air pollution	7,254	5,625	1,630	568	0.5	0.6	0.3	1.1

Source: Murray and Lopez (1996), Table 5, p. 28. Copyright © 1996 World Health Organization. Reprinted with permission
DALY disability-adjusted life years, *YLL* years of life lost by premature death, *YLD* years of life lived with a disability, *NA* not available

Table 11.11 Distribution of deaths and the burden of disease attributable to risk factors for the low- and middle-income countries and high-income countries: 2001

	Deaths		Burden of disease[a]	
Risk Factor	Low/middle-income	High-income	Low/middle-income	High-income
All Cause				
Total number (1,000)	48,351	7,891	1,386,709	149,161
Rate per 1,000 population	9.3	8.5	265.7	160.6
Age-standardized rate per1,000[b]	11.4	5.0	281.7	128.2
Risk Factor	*Percent of total*			
All selected risk factors combined	45.6	44.0	36.1	34.3
Childhood and maternal undernutrition[c]	12.3	0.1	14.2	0.5
Other nutrition-related risk factors and physical activity[d]	30.8	45.1	15.0	28.7
Addictive substances				
Smoking	6.9	18.5	3.9	12.7
Alcohol use	3.9	0.3	3.6	4.4
Illicit drug use	0.4	0.5	0.6	1.4
Sexual and reproductive health				
Unsafe sex	5.8	0.4	5.8	0.6
Non-use and use of ineffective methods of contraception	0.3	0.0	0.5	<0.1
Environmental risks				
Unsafe water, sanitation, and hygiene	3.2	<0.1	3.7	0.2
Urban air pollution, indoor smoke from household use of solid fuels	5.2	1.0	3.6	0.4
Other selected risks	0.9	<0.1	1.0	0.5

Source: Lopez et al. (2006), Chapter 1, Table 1.2, p. 10. Copyright © 2006 World Bank. Reprinted with permission

Some percentages may be slightly overstated because in a few cases individually reported causes have been combined

[a]DALY, or disability-adjusted life years

[b]Age-standardized using the WHO World Standard Population

[c]Includes childhood overweight, iron-deficiency anemia, vitamin A deficiency, and zinc deficiency

[d]Includes high blood pressure, high cholesterol, overweight and obesity, low fruit and vegetable intake, and physical inactivity

literacy and sophistication about health matters, their dietary habits including the quantity and types of food they eat, their use of other ingested products that affect their state of well-being such as tobacco and drugs, their relations with one another as social and sexual beings, and their relations with domesticated animals. I distinguish this group of factors, the culture/lifestyle factors, from another group of influences, the environmental factors. The latter encompass geographic conditions, weather conditions, the food supply, availability of clean water, and the management of waste. These two groups of factors are not mutually exclusive in their impact on health and another group of factors draws from each of them – the cultural, financial, and geographic barriers to accessing health care.

Usually when culture/lifestyle and disease are linked, socioeconomic differences between and within populations are also involved. Closely associated with the considerable and persistent variations in the prevalence of disease observed between countries and within countries are variations in income, wealth, and education. It is difficult to assign low income or little education to the roles of cause, correlate, or risk factor for disease, but they may play all of these roles under different circumstances. In any case they are so closely linked to health status that they should serve as principal factors to be exploited in the effort to reduce the levels of and differences in disease prevalence intranationally and internationally. Gross national income levels per capita in 2007 were only 1/7 as great in the LDC as in the MDC, and 1/29 as great in the Least Developed Countries as in the MDC (Population Reference Bureau 2008/World Bank). Per capita income was $330 in Burundi as compared with $45,850 in the United States (Table 11.1). These differences are huge and have correspondingly tremendous consequences for health. Even where economic inequality is less dramatic, as in the poorest population groups of the Least Developed Countries (older adults in rural Cambodia, for example), health differences appear among the economic segments of the population (Zimmer 2006). Poor people are disadvantaged in their efforts to maintain good health through their limited education, poor nutrition, lack of knowledge of good health practices, and lack of access to health services.

Traditional Medicine

Endemic poverty, traditional health practices, and the everyday unhygienic behavior of the members of a society contribute to the persistence of poor levels of health or even the deterioration of health conditions. This combination of elements is characteristic of many LDC, particularly the least developed countries. There old age, extreme poverty, and poor health are inextricably bound together with one another, and this nexus is aided by the intergenerational transfer of poverty.

Many societies follow their own traditional healing practices, and are ignorant of or ignore modern Western medicine. Usually these alternative medical systems serve as ways of bringing health care to poor and illiterate persons in rural areas in the face of a lack of public resources, doctors, and nurses. In Asia traditional

healers are used more often as a last resort, particularly in rural areas, where health care is not readily available (UNFPA 2002). In sub-Saharan Africa, on the other hand, traditional healers are universally considered a part of the health care system. Traditional healing also serves as a complementary medical system in the more sophisticated parts of these societies and in the MDC, which have borrowed elements of traditional medicine. In India a traditional Hindu healing system known as ayurveda has been widely practiced. Traditional Chinese medicine continues in much of rural China, and in Africa the tribal healer often provides medical advice and potions. In the Caribbean, especially Haiti, voodooism, a combination of Roman Catholic ritual elements and African animism, serves as a healing program for believers through the interpretation of dreams and trances and use of fetishes.

Ayurveda is the traditional medical system of India, with a history going back several millennia. It is based on the theory that illness results from imbalance of the body's life force and employs dietary modification, herbal preparations, yoga, massage, internal cleansing, and other measures similar to those now labeled alternative health measures in the West.

Traditional Chinese medicine also originated millennia ago and continues as part of Chinese medical practice to this day, mainly in the rural areas of China. It is based on the theory that illness results from the improper flow of life force through the body and employs herbal remedies, meditation, acupuncture, and massage. To supplement this tradition, the program of "barefoot doctors" was launched in China in 1965 as part of the Cultural Revolution and as a backlash against Western-style "elitist" medicine. Barefoot doctors are Chinese peasants who have been trained to administer basic health care in the countryside, including first aid, immunizations, and health education, while continuing their farm work. By 1975 there were about 1 million barefoot doctors in China, but by the 1990s the program had largely ended because of lack of government support. In the 1970s the World Health Organization and the leaders of some less developed countries thought that China's program might serve as a model of a way of delivering health care to rural populations in those countries where public resources and medical personnel were inadequate. In measuring the success of the Chinese health program, we may note that it contributed to making considerable reductions in such infectious diseases as smallpox and polio, and to making spectacular reductions in schistosomiasis in some regions.

Relation to Animals and Zoonotic Diseases

Disease among animals and close contact of diseased animals with humans contribute to the risk that the disease will move readily from animals to people. Diseases that readily move from animals to people are called zoonotic diseases. Persons in less developed countries often live among animals ridden with parasites. In such countries, humans may have a personal relation with animals and share living areas and facilities with them. For example, the nomadic sheepherders of Kenya feed their

dogs the viscera of sheep that they have slaughtered nearby. The viscera are often infected with the *Echinococcus* tapeworm and the families of the sheepherders, who live in close contact with the dogs, indirectly ingest the tapeworm eggs eliminated in the dogs' feces. This happens because the eggs are transferred from the feces to the coats of the dogs, then to human hands and mouths. The newly hatched larvae cause hydatid disease, which produces cysts in the liver, lung, brain, or other organs of the humans infected.

In some countries pigs roam the streets and feed on human feces. The family latrine may be situated over the pig pen wherein the waste is disposed of and the pigs are fed. The chain connecting elimination and feeding completes the life cycle for the tapeworm, *taenia solium*, which develops in the muscle of pigs that consume human feces containing tapeworm eggs. The persons who eat undercooked, contaminated pork can become infected with taeniasis. The larvae form cysts in major body organs, causing the debilitating syndrome neurocysticercosis. Some five million people in the world are affected by this disease.

Avian or bird flu (virus H5N1) illustrates the emergence of a new disease being acquired by humans from animals, although it may not yet qualify as a zoonotic disease. As of May 2005 sick ducks and chickens in South East Asia were responsible for the infection of 97 persons and the deaths of 53; by June 2006 the virus had spread to a few countries in Europe, 224 cases were recorded, and 127 deaths had occurred. With such a high reported mortality rate, an understandable fear has developed that the virus would be transmitted to humans from humans and a flu pandemic like that of 1918 would be repeated. This virus has shown up in tigers, leopards, and pigs, which are mammals that often serve as conduits of influenza from animals to humans. The path by which the H5N1 virus could become a major menace to humans is through merger with a strain of human flu, the product of which could pass readily from human to human. One suspected case of human-to-human transmission has been reported, but to date the virus remains an animal (bird) virus.

Considering humans' lack of baseline immunity from this disease, the additional vulnerability of the groups in the population with reduced immune systems and one or more major chronic diseases, the highly contagious character of the disease, and the high lethality of the reported cases compared to other diseases, this disease presents a serious public health threat for a new flu pandemic. (A pandemic disease is one that has spread over a wide geographic area and has affected a substantial proportion of the population in this area.) In the face of uncertainty about this risk, preparations are being made to develop, manufacture, and stockpile a new vaccine,[9] to extend the influenza surveillance system internationally, and to exchange research results among laboratories around the world. As of mid-decade, a global influenza surveillance system is being set up. Such a surveillance system could locally detect

[9]U.S. health officials and South Asian health officers are conducting trials on an avian-flu vaccine in order to be prepared for the further spread of the disease and the possible outbreak of an epidemic. There are serious concerns, however, about the capacity of the pharmaceutical manufacturing system to produce sufficient supplies of any vaccine developed if there is a world wide pandemic.

and report the first wave of an impending pandemic from a virulent human strain of the virus, affording authorities an opportunity to prepare for a full response to later waves of the disease (Yin 2006). Fortunately the H5N1 (avian flu) virus has not mutated to become transmissible from human to human and has not emerged as the cause of the pandemic that the specialists feared.

A new virus, however, the H1N1 (swine flu) virus, has emerged and has raised new fears of a global pandemic. It spreads easily and makes persons sick, but it has rarely been life threatening. H1N1 appeared in Mexico as a combination of two strains of swine flu, a human strain and an avian strain. Cases have now appeared over much of the world. By fall 2009 about 5,000 persons have died worldwide from the swine flu, vastly fewer than in the previous pandemic of H1N1 in 1918. These deaths have not been disproportionately concentrated in the less developed countries. Sickness rates and death rates from the disease have been high among healthy young adults but low among persons over age 60. Apparently older persons have a residual acquired immunity from exposure to an earlier form of H1N1. A vaccine has been developed and has been distributed, and the supplies are now adequate for the demand, partly because there is a reluctance on the part of many persons to get vaccinated. So far a pandemic has been avoided although that may still occur. Health officials have been assiduously disseminating simple instructions to the public on ways to keep the infection rate low and these efforts have been effective.

Lifestyle Factors

In Chap. 6 and earlier in this chapter I discussed the role of certain lifestyle factors and behaviors in affecting an individual's risk of death, disease, and injury. These factors and behaviors included the use of alcohol, tobacco, and illicit drugs, risk-taking social and sexual behavior, insufficient physical activity, and poor or limited dietary choices, often resulting in malnutrition, i.e., overnutrition or undernutrition. Inasmuch as this topic has already been discussed, I will merely present a brief note here on the subject as it pertains to the LDC.

There is ample evidence that lifestyle and behavioral factors are leading contributors to the total disease burden in the LDC. Recall that the original Global Burden of Disease (GBD) study, having assessed the effect of an array of ten risk factors on the total disease burden in the world in 1990, found that the most consequential of these lifestyle and behavioral factors was undernutrition at 16% of the total, and that unsafe sex practices, alcohol abuse, and tobacco use were the third, fourth, and sixth largest risk factors, respectively, in the total disease burden (3.5%, 3.5%, and 2.6%). (See Table 11.10; Murray and Lopez 1996.) For tobacco use, YLLs were a larger contributor to the total disease burden (DALYs) than YLDs; for alcohol use YLDs were a much larger contributor to DALYs than YLLs. Physical inactivity and illicit drugs contributed 1% or less to the total disease burden.

There are sharp differences between men and women in the disease burden caused by most risk factors. The adverse health consequences of unsafe sex practices are borne disproportionately by women in all regions via infections and the complications of unwanted pregnancies. The relative disease burden of the use of tobacco and alcohol largely affects men and, in the last few decades, this burden has been shifting from men in the MDC to men in the LDC. A rapid rise in the use of tobacco and alcohol by men is occurring at present in all the less developed regions, and smoking has become a habit of poorer, less educated men throughout the world (World Bank 1999).

As I mentioned, a sharp rise in the burden of the noncommunicable diseases is anticipated for the LDC in the future. Apart from the fallout from the obesity epidemic, two factors explain this prospective epidemiological change, the aging of the population and the delayed effects of the current heavy and widespread use of tobacco (Warner and Mackay 2006). A paramount role will be played by tobacco use in the burden of disease in the LDC in the next few decades. In 1990 tobacco use was estimated to account for only 2.6% of the total global disease burden (Murray and Lopez 1996); its toll in 2001 was estimated at 4.7% (Mathers et al. 2006); and in 2015 it is expected to account for 10% of the global disease burden (Mathers and Loncar (2006). By 2020 the role of tobacco may be greater than that of any other single factor or disease including malnutrition. For men between the ages of 15 and 60 in the Formerly Socialist Economies of Europe and Asia, the risk of dying is expected to be higher in 2020 than in 2000 because of the tobacco epidemic.

Alcohol use will also remain a leading menace in the LDC. In 1990 it accounted for 3.5% of the global disease burden and in 2001 its toll was about the same, 3.6%. In some areas its share of the total burden is much greater, however. Already by 1990 in Latin America and the Caribbean area alcohol use accounted for almost 10% of the total disease burden. In this region, as in other less developed areas, any supposed protective effects of alcohol use for ischemic heart disease are far outweighed by its harmful effects in increasing rates of death and disability from disease and injuries among men.

Family Planning, Poverty, and Disease

The nexus between low levels of family planning, poverty, and disease is one that can be marshalled to improve health in the LDC. Extension of family planning in such areas would reduce family size and contribute to family income and assets and national economic growth. Improving access of the poor in the LDC to family planning services is an important tool, therefore, in confronting the poverty problem and in turn the health problem of these countries. Family planning is, in short, one of the most cost-effective, least expensive interventions for improving health among the poor people of the world. Furthermore, its impact is one of the more long-lasting interventions for improving health among poor people.

Environment, Poverty, and Disease

Lack of Clean Water, Poor Sanitation, and Air Pollution

Lack of Clean Water and Poor Sanitation

Many of the infectious and parasitic diseases are caused by a lack of access to clean water, poor sanitation, poor personal hygiene, or a combination of these conditions. Unclean water may contain not only microbial pathogens but also chemical contaminants such as arsenic, lead, and other industrial pollutants. These conditions lead to a contaminated food supply. In 2006, 14% of the people in the world, or nearly a billion (950 mil.) people, did not have access to an improved source of water and nearly all of them lived in the Less Developed Countries (880 mil.). (See Population Reference Bureau 2008.) In the Least Developed Countries in 2006 only 62% of the population had access to improved water sources. In 2002 about 2.6 billion people in the world did not have adequate sanitation facilities (Rheingans et al. 2006; see also Parris 2004). The disease burden resulting from a lack of a safe water supply and inadequate sanitation facilities is borne primarily by the poorest countries.

Most of the rural population of the LDC is quite poor, and it is the rural LDC population that has the greatest problems of installing sanitation facilities and securing safe drinking water. In 2002 only about two-thirds of the rural population of the LDC and much less than half of the rural population of sub-Saharan Africa had access to an improved source of drinking water (Table 11.12). Less than two-fifths of the rural population of eastern Africa and middle Africa (for example, Somalia, Zambia, Chad, and Equatorial Guinea) have access to improved drinking water. The situation with respect to the availability of improved sanitation is even more dire. Less than one-third of the rural population of the LDC overall and just over a quarter of the rural population of sub-Saharan Africa have access to improved sanitation. The rural parts of some countries of sub-Saharan Africa have virtually no improved sanitation (for example, Burkina Faso, Guinea, Chad, Somalia, and Namibia) and, around the world, half or less of the rural populations in each of the less developed regions of the world have access to improved sanitation.

The lack of clean water and sanitary facilities in the slums of Luanda, the capital of Angola, undoubtedly accounts for the cholera epidemic that struck that country in 2006. Inadequate water supply, sanitation, and personal and domestic hygiene are major causes of diarrheal diseases. Most of the total burden of disease (DALYs) of the young children in the world results from the mortality (YLL) and disability (YLD) of children suffering from diarrheal diseases. Fortunately the prevalence of the diarrheal diseases has been declining. Rheingans et al. (2006) caution, however, that reducing the proportion of people without access to safe drinking water and sanitary facilities will not automatically result in proportional reductions in the related disease burden. Eliminating the disease burden of young children in the LDC

Table 11.12 Percent of the population with access to improved drinking water and improved sanitation, for the urban and rural parts of world regions and selected countries of Africa: 2002

	Drinking water[a]		Sanitation[b]	
Region and country	Urban	Rural	Urban	Rural
More developed countries	100	NA	100	92
Less developed countries	92	69	73	31
Northern Africa	93	79	82	49
Libya	72	68	97	96
Morocco	99	56	83	31
Tunisia	94	60	90	62
Sub-Saharan Africa	82	45	55	26
Western Africa	78	50	54	25
Burkina Faso	82	44	45	5
Guinea	78	38	25	6
Togo	80	36	71	15
Eastern Africa	84	39	50	26
Malawi	96	62	66	42
Somalia	32	27	47	14
Zambia	90	36	68	32
Middle Africa	79	33	47	22
Angola	79	40	56	16
Chad	40	32	30	0
Equatorial Guinea	45	42	60	46
Southern Africa	98	72	84	41
Namibia	98	72	66	14
South Africa	98	73	86	44
Northern America	100	100	100	100
Central America	97	76	88	45
Caribbean	95	74	82	51
South America	95	64	83	42
Northern/Western Europe	100	NA	NA	NA
Eastern Europe	99	81	94	70
Southern Europe	NA	NA	NA	NA
South Central Asia	94	80	66	25
Southeast Asia	91	71	79	50
East Asia	94	69	74	32
Western Asia	95	73	94	50
Oceania	99	53	98	58

Source: Improved drinking water: Population Reference Bureau (2005). *2005 World Population Data Sheet.* By C. Haub. Washington, DC. Reprinted with permission of the Population Reference Bureau. Primary source: WHO/UNICEF

Improved sanitation: Population Reference Bureau (2006). *2006 World Population Data Sheet*. By C. Haub. Washington, DC. Reprinted with permission of the Population Reference Bureau. Primary source: WHO/UNICEF

[a]The data indicate whether people have reasonable access to an adequate amount of safe water for domestic purposes. Improved sources include a household connection, public standpipe, borehole, and protected well or spring, or rainwater collection. Unimproved sources include vendors, tanker trucks, and unprotected wells or springs

[b]The data indicate whether people have access to facilities likely to insure privacy and hygienic use. Improved facilities include connection to a public sewer or a septic system, pour-flush latrines, simple pit latrines, and/or ventilated improved pit latrines. Unimproved facilities include public or shared latrines, open pit latrines, or bucket latrines

will require a comprehensive health-care program including effective health-care infrastructures and additional resources as well as environmental cleanup.

Poor water, sanitation, and personal and domestic hygiene caused 2.2 million deaths and 82 million disability-adjusted life years (DALYs) in the world in 2002 (WHO 2004). This environmentally-induced health situation had been improving, however. The death and DALY figures from these factors were much higher – 2.7 million and 93 million, respectively – in 1990 (Table 11.10). (See also Prüss et al. 2002; Prüss-Üstün and Corvallán 2006.) The long-term prospects with respect to water and sanitation for the LDC are not favorable. The areas most in need of fresh water and improved sanitation today are the areas with the fastest growing populations or with the greatest increase in numbers. This is particularly true of sub-Saharan Africa. India is expecting a massive fresh water shortage in the next several decades as it adds millions to its population. The future situation may be even more problematic in China.

Air Pollution and Global Warming

People account for the increase in greenhouse gases and so cause climate change, and its "component" global warming, and people experience the adverse consequences of the changes. It is not just population size that accounts for climate change, however, but population size in combination with two other factors, technological efficiency and consumption levels. The impact of these factors is affected by a variety of population characteristics such as age-sex structure, urbanization, education, and household size. The worst consequences are experienced by areas with the least adaptive capacity, i.e., low education levels, low economic growth rates, and high population growth rates (Lutz 2009/2010a). These are the least technologically efficient areas and the ones with the lowest consumption levels.

Now for some concrete details. Many less developed countries endure polluted-air conditions as a result of inadequately controlled fuel emissions even though their consumption of energy is far below that of the industrialized countries. In 2002 the LDC were responsible for the emission of 2.1 metric tons per capita of carbon dioxide as compared with 11.7 metric tons per capita for the MDC (Population Reference Bureau 2007). Alternatively viewed, the MDC account for 56% of the world's emissions of carbon dioxide but have only 18% of the population. The "carbon footprint" of Ethiopia is 0.1 metric ton of carbon dioxide per capita and that of Tunisia is 2.3 tons, as compared with 20 tons for the United States. The huge population growth of China and India is leading to a convergence of the absolute volume of emissions with the United States, but as of 2007–2008 the carbon footprint of the United States is five times that of China and over fifteen times that of India (UN Development Programme 2008).

In general, the negative health impact on countries is inversely related to the level of emissions. Hence, it is greater in the LDC than the MDC and greater in Ethiopia than in Tunisia. This is because in the poorer countries the fumes of coal-burning stoves and automobiles are often discharged directly into homes

or into the immediate environment without controls that convert them into less noxious or innocuous chemical products. The situation in the LDC is deteriorating as automobile use increases. Automobile use in the LDC is expected to rise sharply in the next few decades and carbon dioxide emissions are expected to quadruple between 2000 and 2025 in these areas (UN Intergovernmental Panel on Climate Change 2007).

Global warming, a climatic product of carbon dioxide emissions, is already beginning to manifest itself in the health and survival of the populations of the world and will do so increasingly. Climate change is causing more frequent and intensive meteriological events of great destructive power, such as floods, hurricanes, and rising temperature and humidity. We recall the heat wave in Europe in 2003 and the extreme floods and hurricanes of the years since 2003. Climatic change will expand the areas where certain diseases are common and endemic, especially infectious and parasitic diseases. Malaria is of particular concern because the areas that are tropical or semitropical will widen as a result of global warming and hundreds of millions of additional people will become exposed to this major killer (UN Intergovernmental Panel on Climate Change 2007). The greater health impacts will be felt in the LDC because of the limited ability of the public health systems to respond, the great poverty of the population, and the poor initial health of the population. The MDC will also be affected, of course, but they are better able to prepare for and cope with the effects of extreme temperatures or other climatic shocks produced by global warming. The impact of these extreme weather events in terms of increased morbidity and mortality will depend on the countries' vulnerability and adaptability, which in turn depend on their bank of human capital – itself dependent on education and health levels.

Chemical and Biological Contamination of the Food Supply

In the LDC there has been a considerable rise in the use of chemical pesticides and chemical fertilizers in recent decades, such as occurred in the MDC after 1940. Controlling pests is a major issue in food storage and production, and the use of chemical pesticides is the traditional way of dealing with the problem. Chemicals fed to plants get into the soil and are then washed into underground streams and rivers. These chemical pesticides may be carcinogenic. The effects of the use of pesticides on human health are not well known because it takes many years for these effects to be reflected in actual health conditions and death rates. Nevertheless, this practice would seem to be deleterious to human health. Moreover, food is spoiled by microorganisms, and the additives used to help preserve food may also be carcinogenic. The World Health Organization has estimated that about 20% of the world's food supply is lost to spoilage from microorganisms (Weeks 1999).

Extremes of Weather, Famine, and Malnutrition

Consequences of Extremes of Weather and Crop Failure

In many less developed countries the extremes of weather combine with abject family poverty to produce crop failures, food insecurity, famines, and continuing widespread undernutrition. Periodic drought, flood, and insect infestation result in periodic crop failures. Chronic famine associated with drought is a characteristic of many of the less developed countries, especially those in sub-Saharan Africa. Occasionally these countries experience a very severe famine. Between the mid-1980s and the early 1990s a severe famine associated with drought affected much of the area south of the Sahara desert, including Chad, Niger, Ethiopia, and Somalia. Famines can also occur as a result of other weather conditions than a drought. As T. Dyson (1991), a leading expert on famines in the LDC, has noted, famines periodically occur in South Asia not only when there is a lack of monsoon rains but also when there is flooding from extremely heavy rainfall. Thus, the two extremes of weather, drought and flooding, can destroy crops and contribute to a food shortage and famine. In some cases, as in Ethiopia, there are two other problems that contribute to food insecurity. One is the lack of an infrastructure to retrieve and store water for use by the large rural population when a drought occurs. The second is the progressive subdivision of small farms resulting from rapid population growth; these farms are too small to support large families and the families face a food shortage when water is lacking.

In the usual chain of events, bad weather results in reduced agricultural production and a fall in the food supply, which in turn brings about a rise in malnutrition, a rise in the death rate, and a decline in the population growth rate. According to Galloway (1984), long-term historical data for China and Europe confirm the fact that bad weather contributed to poor harvests and famine in these areas, and then to a rise (or a slowing of the decrease) in mortality rates and modest drops in population growth rates. These events did not mean a sharp drop in the growth rate of these regions because birth rates tended to remain high and compensate for the rise in death rates. In recent decades the famines have produced only moderate declines in death rates without affecting birth rates, and hence the effect on growth rates of famines has been small.

Underlying these episodic occurrences of drought and flood in specific regions are several more general trends affecting the world food supply and its distribution that portend more widespread food shortages, food insecurity, and famines. These factors have been alluded to in the parts of the discussion in this chapter dealing with the environmental influences on health in the less developed regions, especially sub-Saharan Africa. The factors are rapid population growth and three environmental factors, increasing soil erosion, spreading water shortages, and rising temperatures ("global warming"). (See Brown 2008.) Currently the world's population is increasing by about 80 million persons a year. They not only require more stocks of food but want to consume highly grain-intensive livestock products in greater numbers.

In addition, there is a heavy demand to divert U.S. grain to the production of ethanol fuels for automobiles. Of the environmental factors the most immediate threat is the spread of water shortages. The heavy use of water for irrigation is depleting the supply of water from underground wells and causing water tables in China, India, and the United States to fall. As a result, China's wheat and rice crops have been falling and India's food supply is in imminent jeopardy (Brown 2009). Successful agriculture depends on fertile topsoil but the topsoil is eroding faster than new soil can form in a substantial share of the world's cropland. Rising surface temperatures reduce crop yields; even a minor rise in temperature can have a deleterious effect on food production.

The consequences of food shortages, food insecurity, and famine for health appear in the form of hunger, malnutrition, stunting of growth, wasting of the body, and premature mortality. In the less severe stages, these conditions account for multitudes of unhealthy children who are hungry, unable to attend school, or if attending, cannot learn, and large numbers of unhealthy employees or, more commonly, unhealthy adults who are unemployed. Countries so affected have little prospect of sustained social and economic development.

Defining Terms and Measuring the Health Consequences of Famine

To measure and compare the frequency and scope of famines and their consequences, operational definitions are needed for the relevant terms. First, note that food shortage, food insecurity, and famine are population concepts (i.e., they apply to populations), whereas terms such as hunger, malnutrition, and undernutrition are individual concepts (i.e., they apply to individuals) though they may be summarized for populations. Dyson (1991) has defined a famine as a "food shortage accompanied by a significant increase in deaths." This definition is too general for use in comparative analytic studies; yet there is no universally accepted operational definition of the term. In general, the situation is similar for the other terms listed here.

Food insecurity is antecedent to and the conduit to undernutrition. The commonly used definitions of food security are those of the UN Food and Agricultural Organization and the US Department of Agriculture. The basic elements of these definitions are the availability (i.e., sufficiency) of, access to, and quality (i.e., safety/nutrient content) of food. The food must meet the dietary needs and food preferences of a household for an active and healthy life, and must be acquired in a socially acceptable way. Accordingly, food insecurity may be defined as a household's lack of a sustainable physical or economic access to safe, nutritious or socially acceptable food for a healthy and productive life. In short, a household is considered food-secure when its members do not live in hunger or fear of hunger. Food insecurity may be temporary, seasonal, or chronic.

Concepts of malnutrition, undernutrition, and overnutrition. Malnutrition may be defined as a condition of impaired health due to a dietary deficiency, excess, or imbalance. With this condition, the body has not received the right amount of

vitamins, minerals, and other nutrients that it needs to maintain healthy bodily functions. Malnutrition may be considered as having three forms: Primary undernutrition, secondary undernutrition, and overnutrition. Primary undernutrition, or micronutrient malnutrition, develops when nutrient intake is insufficient to provide for normal physiological needs and to enable one to carry out light physical activity. In adults, primary undernutrition is invariably due to a lack of food. Secondary undernutrition occurs when an underlying disease process (e.g., TB or HIV) increases metabolic demands and/or decreases food intake or utilization. The nutritional deficiency occurs because there is loss of appetite, inability to eat, or disruption of digestion, absorption, or metabolism of food particles, or because the bodily use or excretion of nutrients occurs more rapidly than they can be replaced. Here undernutrition is secondary to infection. The treatment of primary and secondary undernutrition may be quite different. Overnutrition results from the chronic intake of nutrients at levels above ranges that are adequate or safe. It could involve consumption of too much of the right food, consumption of too much of the wrong food, or taking too many vitamins or dietary supplements.

Acute undernutrition has also been distinguished from chronic undernutrition. In 1988 the International Dietary Energy Consultative Group proposed a definition of chronic adult undernutrition, calling it "chronic energy deficiency." It is "a steady state at which a person is in an energy balance. . . at a cost either in terms of increased risk to health or an impairment of functions and health." Acute adult undernutrition, or "acute energy deficiency," was defined as "a state of negative energy balance, i.e., a progressive loss of body energy." The distinction between acute adult undernutrition and chronic adult undernutrition is important because persons with these conditions function differently. Persons with chronic undernutrition with very low body mass indexes may function reasonably normally. In contrast, persons with acute undernutrition, who have similarly low body mass indexes and who have been plunged into the condition suddenly, show extremely poor physical and mental responses. It is difficult, however, to distinguish the two forms of undernutrition in population surveys.

Measurement of malnutrition and its variations. I refer briefly to some basic measures of malnutrition in this section and consider them more fully in a later section. In general, undernutrition is measured as the relation of weight to height for adults, and weight to age, or height to age, for children. The usual measure is the body mass index (BMI), or weight in kilograms divided by height in meters squared. Following the usual cutoffs, a BMI under 18.5 is deemed underweight and a BMI at or over 25.0 is deemed overweight. These translate into undernutrition and overnutrition. If the subject is a child, undernutrition is defined as low weight for the age of the child. More specifically, a child is suffering from undernutrition if his or her weight is two or more standard deviations below the median weight for his/her particular age according to the World Health Organization's Child Growth Standards.

With chronic undernutrition in young children, stunting, or a low height for the age of the child, is likely to result. The Demographic and Health Surveys (DHS)

classifies as stunted those children whose height-for-age Z-score is below (i.e., greater than) two standard deviations of the median reference standard as established for their age by the WHO. Stunting reflects a failure of the child to reach its "natural" size. Stunting usually occurs after nutrient deprivation *in utero* or within the two first years of life. (The DHS figures are based on children under age 3, 4, or 5 depending on the country.) It may result from an insufficient amount or quality of food, infectious disease with its common sequel of diarrhea, or a combination of these two causes. At any age, particularly as an adult, a person who experiences major recent radical weight loss is likely to be suffering from wasting, or a low weight for his/her height. Sudden weight loss can result from an extreme lack of food caused by an environmental catastrophe or a severe attack of a disease that "kills" the appetite, prevents absorption of nutrients, and dehydrates the body. Hence, undernutrition can cause either stunting or wasting, and both of them are associated with being grossly underweight.

Extent of undernutrition. Food insecurity is a worldwide problem. According to the Food and Agriculture Organization, 14% of the people in the world, or about 940 million people, were suffering from undernutrition in 2002–2004 (Population Reference Bureau 2008/FAO); Table 11.13). In Haiti, for example, where the gross national income per capita is $1,050, only 30% of the land is arable, and nearly 70% of the population depends on subsistence agriculture, levels of undernutrition are approaching half the total population. Eight other countries in the world outrank Haiti in their level of undernutrition: Liberia, Sierra Leone, Burundi, Comoros, Eritrea, Zimbabwe, Democratic Republic of the Congo, and Tajikistan. More than 46% of each of their populations are undernourished. All but one of these countries is located in sub-Saharan Africa. The largest numbers of undernourished persons, however, live in the more populous underdeveloped countries: India (230 million), China (160 million), Democratic Republic of the Congo (49 million), Bangladesh (44 million), and Pakistan (41 million). Whether or not the gross numbers on global undernutrition are now rising or falling is not clear, but it is quite likely that they are rising in the Least Developed Countries, and in sub-Saharan Africa in particular.

Large inequalities in health exist among the socioeconomic classes within each less developed country. An analysis of data from the Demographic and Health Surveys (DHS) for 53 less developed countries conducted by the World Bank in 2003 revealed huge differences among the women in each country in their degree of malnutrition, distinguished by quintile level of wealth. For the countries as a group, the highest wealth quintile of women had one-half the percentage of malnutrition as the lowest wealth quintile of women (Population Reference Bureau/World Bank 2004). The two percentages varied widely from country to country. The figures were, for example, 15 and 50 for India, 5 and 9 for Brazil, 6 and 18 for Kenya, and 25 and 32 for Ethiopia.

Undernutrition is the leading dietary problem of women in parts of South/Southeast Asia (e.g., Bangladesh, India) and in eastern Africa (e.g., Eritrea). Undernutrition is characteristic of the LDC, but overnutrition is common there as well. Overnutrition has become the principal dietary problem of women in North

Table 11.13 Percentage of population who are undernourished, for world regions and selected countries in sub-Saharan Africa: 2002–2004

Region and country	
World	14
More Developed Countries	<2.5
Less Developed Countries	17
Least Developed	35
Northern Africa	8
Morocco	6
Sudan	26
Sub-Saharan Africa	31
Western Africa	15
Liberia	50
Guinea	24
Togo	24
Eastern Africa	40
Burundi	66
Ethiopia	45
Malawi	35
Middle Africa	55
Angola	35
Congo, Dem. Rep.	74
Gabon	5
Southern Africa	4
Namibia	24
South Africa	<2.5
Northern America	<2.5
Latin America/Caribbean	10
Northern/Western Europe	<2.5
Eastern Europe	3
Southern Europe	<2.5
Western Asia	8
South Central Asia	21
Southeast and East Asia	11
Oceania	<2.5

Source: Population Reference Bureau (2008). Reprinted with permission of the Population Reference Bureau. Primary source: Statistical Division of the United Nations Food and Agriculture Organization

Note: Undernourishment is defined here as a condition where dietary energy consumption is continuously below the minimum dietary energy requirements for maintaining a healthy life and carrying out light physical activity

Africa/West Asia/Europe, Central Asia, and Latin America and the Caribbean. Both under- and overnutrition exist side by side in western Africa (e.g., Guinea, Senegal, Togo). Currently, in fact, there are more overweight people than underweight people in the world; some 1.8 billion people in the world are overweight (Levinson and

Bassett 2007) while there are less than 1.0 billion underweight people. In the MDC overnutrition is far more common than undernutrition.

Stunting. Stunting prevalence is high in most regions of the world. Stunting is commonly found in poor countries and in the poorest population segments of these countries. The regions with the highest prevalence ratios of stunting for children under 5 are eastern Africa and middle Africa, with 50% and 40%, respectively, and South-Central Asia, with a ratio of almost 41% (Population Reference Bureau/World Bank 2004). In the LDC as a whole roughly 30% of all children are stunted. The prevalence of stunting among children has been rising in sub-Saharan Africa while it has been falling in Asia and Latin America and the Caribbean. Stunting is not unknown in the MDC as well, particularly among the lowest-income segment of each country's population. The prevalence ratio for stunting among children under 5 in the United States was 2.4% in 2006.

Stunting varies with the economic status of the mothers within countries. The analysis of the Demographic and Health Surveys data for wealth quintiles in the LDC by the World Bank cited earlier found that the children of the highest wealth quintile of mothers had one-third the percentage of stunting as the children of the lowest wealth quintile of mothers. The figures for India were 27% and 58%, for Brazil 2% and 23%, for Kenya 17 and 44, and for Ethiopia 43% and 53% (Population Reference Bureau/World Bank 2004). A strong covariate of stunting, in addition to mother's wealth, is mother's education. Whether the mother has no education or a high school education makes a considerable difference in the nutritional status of her children in all the LDC surveyed.

Causes of Undernutrition

Numerous causes of undernutrition in the LDC can be enumerated, accounting for it at different logical levels – environmental, demographic, socioeconomic, medical, and political. Opinions differ greatly as to the underlying reasons for the persistence of undernutrition in the LDC. Rapid population growth, limited natural resources (i.e., scarcity of water and arable land), extremes of weather in the form of drought and flooding, and deforestation have been cited as largely responsible for the deplorable food situation in sub-Saharan Africa. Some reject the view that there is a direct causal relationship between rapid population growth, on the one hand, and famine, on the other, and maintain that high vulnerability to food shortages in the countries of sub-Saharan Africa is situational and chronic. In this view, major changes in government development policy are needed to increase agricultural growth and food supply.

It can be reasonably argued that every country has sufficient agricultural capacity to feed its own population and that "hunger" is not typically the result of a lack of food. Rather, it results from inadequacies in the food distribution network and from inept government policies. It is hard to move food around the world and to distribute it in landlocked countries lacking a good road system. The World Bank

maintains that the best way for countries to overcome the problem of hunger is to develop economies based on exports that will provide the means to buy food on the international market. It can also be reasonably argued that famine, particularly in sub-Saharan Africa, is primarily a result of a lack of purchasing power of the people, i.e., a case of too little income to buy a minimum supply of food. In Niger, an extremely poor, landlocked country in the Sahel region of sub-Saharan Africa, many people could not afford to buy sufficient food when a food crisis resulting from drought faced the country in 2005. The apparent solution was to give people enough money to buy the available food or to create jobs for them. A fair amount of international aid had been made available to Niger, but it was applied to debt relief and other national expenses, and not to increasing agricultural productivity or, more generally, improving the economy.

To reduce chronic undernutrition in the affected areas, it is necessary, in a short-term emergency situation, to overcome the obstacles in getting food from the distribution centers to individual families and to favor giving it directly to the families themselves, not to governments. In the longer term, it is necessary to pursue a wide-ranging program of socioeconomic development: To reduce civil strife, to improve the social status of women, to reduce the prevalence of communicable diseases, to support the establishment of small businesses and other community programs by providing access to credit and loans, to expand family planning services, and to support local educational programs on nutrition and healthy behavior. In the long term it also seems prudent to encourage the production, consumption, and improvement of native crops, and to curtail the direct provision of food from overseas, which reduces the incentives of native farmers to produce food for sale and reduces their income.

In the LDC, nutritional deficiencies are almost always associated with famine. In the event of famine and other such emergency situations, measures for the assessment of the nutritional status of individuals in the affected populations are needed because of operational exigencies. During famine-relief operations, workers increasingly confront and treat severe undernutrition. This requires a more formal and extended definition of the terms for nutritional deficiencies than we have given above. Yet, there is, at present, no universally accepted definition of undernutrition and no specific treatment guidelines for the condition (UN/ACC/SCN 2000). Thus, the screening and selection of admissions into therapeutic feeding centers and the dietary treatment of those admitted become problematic.

Formal Measures of Undernutrition

The terms relating to the consequences of famine require more precise definition not only for statistical/analytic purposes but also for the purpose of evaluating a claim by a national government for international food aid. The Demographic and Health Surveys define under- and overnutrition of adults solely in terms of the Body Mass Index (BMI). Undernutrition is defined as a BMI of less than 18.5. Overnutrition is defined as a BMI of 25.0 or more, with the separate categories of 25.0–29.9

representing overweight and 30.0 or more representing obese. By this definition at least one-fifth of the women aged 15 to 49 are undernourished in 10 of 50 countries surveyed, and at least one-fifth are overnourished in 25 countries.

Some of the terms and possible measures for them have been formally assessed by the United Nations System's Standing Committee on Nutrition (UN/ACC/SCN). Other statistical initiatives to improve the collection and compilation of relevant data are the Health Metrics Network (HMN) and Standardized Monitoring and Assessment of Relief and Transitions (SMART).[10] The UN/ACC/SCN and WHO have developed a common set of concepts and measures for purposes of distribution of food aid. They recognize a chain of conditions reflecting vulnerability based on the shortage of food, namely, food insecurity, malnutrition, famine, and premature mortality.[11] The leading international nongovernmental service organizations adhere to the same common set of concepts and measures for purposes of distribution of food aid as the UN and the WHO. The formal definitions of undernutrition involve both biometric components and nonbiometric components. Biometric measures include, for example, height, weight, and circumference of the upper arm, taken individually or in combination. The nonbiometric measures include, for example, lack of sufficient food through the year and consumption of less than two meals a day.

The UN/ACC/SCN (2000) has not reached a consensus on a method of assessing undernutrition, but has announced its preliminary recommendations. The key measurement devices are the Body Mass Index (BMI), the Mid-Upper Arm Circumference (MUAC), and various clinical criteria. UN/ACC/SCN has assessed these

[10]United Nations, Administrative Committee on Coordination, Standing Committee on Nutrition (UN/ACC/SCN), Geneva, Switzerland, is composed of a number of UN member agencies (including UN, FAO, UNICEF, UNESCO, WHO, WFP, UNHCR, and UNAIDS), the World Bank, bilateral donor agencies, and other NGOs. Its function is to raise awareness of nutrition problems, mobilize commitment to solve them, increase the scale of efforts to reduce malnutrition, and promote cooperation among UN agencies and partner organizations in support of national efforts to end malnutrition.

The Health Metrics Network (HMN) is a global partnership comprised of countries, multilateral and bilateral development agencies, foundations, global health initiatives, and technical experts, with the goal of increasing the availability and use of timely, reliable health information by catalyzing the funding and development of core health information systems in developing countries. In this way the partnership works to improve decision-making in public health through better health information.

Standardized Monitoring and Assessment of Relief and Transitions (SMART) is a global effort to coordinate the assessment, monitoring, and evaluation of the interventions of the organizations providing humanitarian assistance. USAID and the Department of State are spearheading this effort, which is seeking to establish a consensus on a methodology for assessing the mortality and nutritional status of populations in crisis. SMART plans to pilot a program to routinely collect, analyze, and disseminate information on the nutrition and mortality experience of the populations served by humanitarian organizations.

[11]Hunger is not commonly used in the formal measurement of the consequences of famine. It is the psychological response to the physiological condition of nutritional deficiency. It is often used as an informal substitute for undernutrition, however. The word is no longer used to describe the official extent of food insecurity in the United States by the U.S. Department of Agriculture.

methods for their usefulness in determining the prevalence of undernutrition among both adults and children in a population and for screening severely undernourished persons for entrance into feeding clinics. According to the SCN, BMI may be used to measure the prevalence of chronic undernutrition in a population although this measure has several limitations when comparing different populations. Adults 20–60 years of age should be classified according to the following categories of the BMI (weight in kg.÷ height in m^2):

Normal	18.5–24.9
Underweight	
Grade I	17.0–18.4
Grade II	16.0–16.9
Grade III	<= 15.9

The BMI is known to vary with age and body shape. In order to account for differences in body shape when comparing different populations, the Cormic index, the ratio of leg length to trunk length or the ratio of sitting height to standing height, should be taken into account.[12]

The Mid-Upper Arm Circumference (MUAC) is another biometric measure that may be used to assess the prevalence of chronic undernutrition in a population. MUAC measures the circumference of the upper arm about midway between the shoulder and the elbow. The MUAC classification is used differently as a criterion for screening a population for acutely undernourished persons and as a criterion for admitting persons into adult therapeutic feeding centers. For the latter purpose, the recommended criteria combine MUAC with various clinical measures, called the CHANCES model:

MUAC	<160 mm irrespective of clinical signs
MUAC	161–185 mm plus one of the following clinical signs:
	Bilateral pitting edema (Beattie grade 3 or worse)[13]
	Inability to stand
	Apparent dehydration
Famine edema (Beattie grade 3 or worse) alone, as assessed by a clinician to exclude other causes.	

[12]The Cormic index corrects the BMI by a factor derived from a linear regression model. For an illustration of its use, see UN/ACC/SCN (2000).

[13]The Beattie classification of famine edema is a scale of the severity of edema from 0 to 5. Grade 0 represents the absence of edema, grade 1 minimal edema on the foot or ankle that is not obvious, grade 2 obvious edema on the foot or ankle, grade 3 edema demonstrable up to the knee, grade 4 edema demonstrable up to the inguinal ligament, and grade 5 total body edema.

In addition to the biometric measures and the clinical measures, certain social factors may be included in the model. The relative weighting of these social factors must be determined locally. Social factors could include the following: Access to food (quantity and quality), distance from feeding centers, presence/absence of carers, availability of cooking utensils, dependents, and so on. The social factors can aid in determining whether the individual should get supplementary or therapeutic care. According to UN/ACC/SCN, admission to adult supplementary feeding centers should be based on the following criteria: MUAC 161–185 mm and no relevant signs or few relevant social factors. UN/ACC/SCN recommends that workers should take these standards as the starting point and adapt them according to the situation.

Limitations of BMI and MUAC. The use of weight alone to assess nutritional status should be limited to monitoring purposes because it is confounded by height. The BMI was intended to allow for the effect of height on weight. However, the BMI has certain limitations as a measure of nutritional status secured in health surveys. Many factors other than nutritional status determine BMI. Most important of these, as mentioned, is body shape, represented by the Cormic index, i.e., the ratio of sitting-height to standing-height (SH/S). This index varies both within populations and between populations. These differences show world-wide variation in the SH/S ratio from 0.48 in Australian aborigines up to 0.55 in the Japanese. This variation has considerable influence on the BMI. During an emergency, however, few people would know how, or would be free, to perform this special adjustment.

Even height and weight measurements may be difficult to obtain during a famine. Skeletal problems are common and many persons cannot stand straight. Furthermore, the frequent concurrence of undernutrition and edema results in an overstatement of weight. In addition, adult body size, shape, and composition vary with age; for example, adults tend to put on fat mass with age. These features may alter the significance of BMI at different ages. However, many adults in the LDC do not know their exact age and it may, therefore, be difficult to differentiate the diagnosis of nutritional status according to age, especially in emergency settings.

MUAC also has a number of limitations. Data are lacking so far to relate mortality and other functional measures to specific levels of MUAC in emergency famine-relief operations. Age affects the distribution of subcutaneous fat and so different levels of MUAC may be required for broad age classes. MUAC is subject to measurement error and a lack of reproducibility both by the same observer and between observers. On the other hand, MUAC is an appropriate indicator for the assessment of acute adult undernutrition. It requires no equipment other than a tape measure, is simple to determine, and can be assessed easily.

Children. Two main consequences of undernutrition found in children, stunting and wasting, were mentioned. Different processes produce these two conditions and they are assessed using different anthropometric indices. Wasting is caused by an acute nutritional deficit and a disease such as diarrhea, and is characterized by a reduction in weight-for-height (BMI) or arm circumference (MUAC), or both. Stunting results

from prolonged nutritional deficit and/or disease and is characterized by a reduction in height-for-age. Wasting and stunting are associated with different functional consequences; the former is a strong predictor of short-term mortality and the latter predicts mortality in the longer term.

MUAC alone is indicated as a measure for undernutrition for children between 1 and 5 years of age during emergencies (UN/ACC/SCN 2000). The use of MUAC for children is associated with such problems as the lack of reproducibility of the measurements and measurement errors by observers in sequential tests and between observers. Undernutrition, however measured, can be confounded by infection, metabolic dysfunction, and disease-induced edema. On the other hand, in some studies MUAC predicted deaths in children better than any other anthropometric indicator.

Elderly persons. The measurement of undernutrition for elderly persons in the LDC is a more complex task than for younger persons (UN/ACC/SCN 1999). The effect of undernutrition is confounded with the effect of disease, with the result that the levels of the biometric indicators for undernutrition are overstated. Changes in body size, shape, and composition are great in older age. These changes greatly alter the functional significance of the BMI at these ages. The measurement of height and weight becomes more problematic because of musculoskeletal problems and lack of measurement equipment in the field. Proxy measures may have to be used for height. Researchers have found a useful relation between height and various other skeletal measures, such as femur length and knee height, from which estimates of height can be made. When height has to be estimated in this way, however, the error of the estimate can be substantial. Weight can be greatly increased by edema, accounting for up to 10% increase in body weight, with a serious effect on BMI measurement. For patients with severe famine edema a high BMI is more an indicator of severe edema than undernutrition. Hence, the BMI is not a good measure of undernutrition of older people.

Measuring functional ability rather than anthropometric characteristics may serve as a useful screening tool. The measurement is usually done by using scores derived from the answers to a set of questions similar to the Activities of Daily Living. An indicator of nutritional status would still have to be included in the package of questions. Such an indicator is needed to differentiate clinical illness, which can be treated best in medical units, from undernutrition, which can be treated best in special feeding centers.

Measuring mortality and malnutrition due to famine. In emergency relief situations, the form of the assessment of health problems and the implementation of relief programs must be tailored to match the resources (i.e., personnel and equipment) available in each particular situation. Each situation is rather different and unpredictable since famines involve great social upheavals and in many cases have been accompanied by armed conflict. These conditions make it difficult to apply measurement devices, assess the extent of health problems, and implement effective relief programs.

In spite of their complexity, sample surveys have been generally used to gather information regarding mortality, the causes of death, and malnutrition in famine-affected regions. For example, Salama et al. (2001) employed a two-stage cluster sample survey over a 5-day period in July 2000 to secure such information for the Godi district in the Somali region of Ethiopia, the epicenter of the famine in that region in the year 2000. Anthropometric measures were taken along with retrospective information on deaths over the 8-month period preceding the survey date. The crude mortality rate (deaths per day per 10,000 population), the corresponding rates for children under 5 and 5–14 years of age, death rates before and after the major relief intervention, the death rates and prevalence ratios for children due to wasting, and the prevalence ratios of undernutrition among adults (body mass index adjusted for body shape) were calculated. (For further discussion of this subject, see Chap. 10, section on Refugee Populations.)

Financial, Geographic, and Cultural Barriers to Accessing Health Care

Poor/nonpoor and socioeconomic class differences within countries apply on a global scale and are associated with health differences. The poor in the Less Developed Countries (LDC), like the poor in the More Developed Countries (MDC), have less access to health care and are more likely to receive untimely and less effective treatment than the nonpoor. They are more likely to live in an unhealthful environment, work at a hazardous job, and follow an unhealthful lifestyle, e.g., smoke and eat a nonnutritious diet. Similarly, worldwide the link between education/literacy and health prevails. Lutz (2009/2010) uses this link to project the possibility of a world with fewer people (2–6 billion) sometime in the twenty-second century who are much more educated than the world's present population and therefore much more wealthy and healthy. His conclusion is based on projections of population, disaggregated by age, sex, and four levels of educational attainment. These projections allow for different mortality and fertility rates for people with different educational statuses, and take account of the rapid increase to date in the share of the world's population that has received a secondary or higher education and the moderating effect of education on both fertility and mortality.

A focus on access to health care in the LDC, particularly on the part of women, rather than risk factors, identifies a combination of factors – financial, geographic, and cultural – that act as barriers to accessing health care. In short, if residents in the LDC, especially women, get sick and need medical aid, they are less likely to obtain it than the residents of the MDC, because of lack of money, difficulty in traveling to the health facility, or cultural barriers stemming from women's subordinate status. For the 27 countries surveyed with respect to access to health care by ORC MACRO in the Demographic and Health Surveys (DHS) program, lack of money was the barrier to receiving needed health care most often cited by married women in 20

countries (ORC MACRO 2006). At least three-quarters of the married women in Armenia, Cambodia, Haiti, and Rwanda considered lack of funds an obstacle. The shortage of money could have resulted from their poverty per se or their lack of control over household resources.

According to the DHS, more than half of the married women in eight countries reported a problem accessing health care as a result of geographic difficulties related to distance and lack of transportation (ORC MACRO 2006). These problems were especially common in mountainous countries, such as Bolivia and Nepal, but existed also in other countries such as Namibia and Zimbabwe. Cultural concerns arising from women's roles and status were the barrier that married women cited most frequently in five countries. Illustrative of these barriers are the need for women to secure a spouse's permission to go for treatment, their inability to travel alone to a health facility, or their discomfort in consulting a male health provider. The "cultural" problem was common in Armenia, Bolivia, Cambodia, Nicaragua, Nepal, and Peru; it was not a common problem in sub-Saharan Africa.

References and Suggested Readings

Sources and Quality of Data

Arriaga, E. (1999). Causes of death in developing countries and in countries with economies in transition: An overview of current knowledge, data availability, and novel methods of data collection. In United Nations (Eds.), *Health and mortality issues of global concern* (pp. 84–109). New York: United Nations.

Manton, K., Myers, G., & Andrews, G. R. (1987). Morbidity and disability patterns in four developing nations: Their implications for social and economic integration of the elderly. *Journal of Cross-Cultural Gerontology*, *2*, 115.

Mosley, W. H., & Chen, L. C. (1984a). An analytical framework for the study of child survival in developing countries. *Population and Development Review*, *10*(Suppl.), 25–45.

Mosley, W. H., & Chen, L. C. (Eds.). (1984b). *Child survival: Strategies for research*. New York: Population Council.

Oosterbaan, M. M. (1995). Guinea Bissau: Maternal mortality assessment. *World Health Statistics Quarterly*, *48*(1), 34–38.

Pan American Health Organization (PAHO). (2002). *Health in the Americas*. (2002 ed.). Washington, DC: Pan American Health Organization.

Ruzicka, L. T. (1984). Mortality transition in Asia: Technology confronts poverty. In G. W. Jones (Ed.), *Demographic Transition in Asia* (pp. 31–56). Singapore: Maruzen Asia.

Snow, R. W., & Marsh, K. (1992). How useful are verbal autopsies to estimate childhood causes of death? *Health Policy and Planning*, *7*(1), 21–29.

Snow, R. W., Armstrong, J. R., Forster, D., et al. (1992). Childhood deaths in Africa: Uses and limitations of verbal autopsies. *Lancet*, *340*(8, 815), 351–355.

World Health Organization. (1999). *Ageing and health. A global challenge for the 21st century*. In Proceedings of a WHO Symposium, Kobe, Japan, 10–13 November 1998.

Socioeconomic Status, Lifestyle, Environmental Factors, and Nutrition

Brown, L. R. (2008). *Plan B 3.0: Mobilizing to save civilization*. Earth Policy Institute. New York: W.W. Norton.

Brown, L. R. (2009). Could food shortages bring down civilization? *Scientific American*, May 2009, 50–57.

Caldwell, J. (1986). Routes to low mortality in poor countries. *Population and Development Review*, *12*, 171–219.

Dyson, T. (1991). On the demography of South Asian famines. *Population Studies*, *45*, 5–25.

Galloway, P. (1984). *Long term fluctuations in climate and population in the pre-industrial Era*. Berkeley: University of California.

Gwatkin, D. R., Rutstein, S., Johnson, K., Suliman, E. A., & Wagstaff, A. (2003). *Initial country-level information about socioeconomic differences in health, nutrition, and population* (Vols. I and II). Washington, DC: World Bank.

Levinson, F. J., & Bassett, L. (2007). Malnutrition is still a major contributor to child deaths. In *Policy Brief*. Washington, DC: Population Reference Bureau.

Lutz, W. (2009/2010a, Winter). What can demographers contribute to understanding the link between population and climate change. *Popnet: Population Network Newsletter*, *41*, 1–2. IIASA.

Lutz, W. (2009/2010b, Winter). Toward's a world of 2–6 billion well-educated and therefore healthy and wealthy people. Popnet: *Population Network Newsletter*, *41*, 5–7. IIASA.

ORC Macro. (2006). *Women's lives and experiences: Changes in the past ten years*. By D. Vadnais, A. Kols, & A. Abderrahim. Calverton, MD: ORC Macro. Under contract from U.S. Agency for International Development (US AID).

Population Reference Bureau/World Bank. (2004). Improving the health of the world's poorest people. In *Policy Brief*, and *The Wealth Gap in Health*, Wall Chart. Washington, DC: Population Reference Bureau. Compiled by the World Bank from the Demographic and Health Surveys.

Population Reference Bureau. (2008). *2008 World Population Data Sheet*. By C. Haub & M. M. Kent. Undernourishment data from United Nations Food and Agriculture Organization. Washington, DC: Population Reference Bureau.

Prüss, A., Kay, D., Fewtrell, L., & Bartram, J. (2002). Estimating the burden of disease from water, sanitation, and hygiene at a global level. *Environmental Health Perspectives*, *110*, 537–542.

Prüss-Üstün, A., & Corvallán, C. (2006). *Preventing disease through healthy environments: Towards an estimate of the environmental burden of disease*. Geneva, Switzerland: World Health Organization.

Salama, P., Assefa, A., Talley, L., Spiegel, P., et al. (2001). Malnutrition, measles, mortality, and the humanitarian response during a famine in Ethiopia. *Journal of the American Medical Association*, *286*(5), 563–571.

UN/ACC/SCN. (1999). Nutrition and healthy ageing. *SCN News*, *19*. Geneva, Switzerland: UN/ACC/SCN.

UN/ACC/SCN. (2000, July). *Assessment of nutritional status in emergency-affected populations: Adults*. Geneva, Switzerland: UN/ACC/SCN (Unofficial).

UNFPA. (2002). Report of an expert group meeting, 29–31 October, 2001, Valletta, Malta. Social, health, and gender issues with a focus on the poor in old age. *Population Ageing and Development*, *3*. New York: UNFPA.

United Nations, Development Programme. (2008). *Fighting climate change. Human solidarity in a divided world* (Human Development Report 2007–2008). New York: United Nations.

United Nations, Intergovernmental Panel on Climate Change. (2007). *Climate change, 2007* (Fourth Assessment Report (AR4) Working Group II Report, *Impacts, Adaptation, and Vulnerability*, Chapter 8, "Human health"). Geneva, Switzerland: United Nations.

Waldman, R. J. (2001). Public health in times of war and famine: What can be done? What should be done? *Journal of American Medical Association*, *286*(5), 588–590.
Warner, K. E., & Mackay, J. (2006). The global tobacco disease pandemic: Nature, causes, and cures. *Global Public Health*, *1*(1), 65–86.
Weeks, J. R. (1999). *Population: An introduction to concepts and issues* (7th ed.). Belmont, CA: Wadworth Publishing Company. Esp. pp. 500–501.
World Bank. (1999). *Curbing the epidemic: Governments and the economics of tobacco control*. Washington, DC: World Bank.
Zimmer, Z. (2006). *Poverty, wealth inequality, and health among older adults in rural Cambodia* (Working Papers. No. 217). New York: Population Council.

Diseases, Injuries, Disabilities, and Deaths

Bawah, A. A., & Binka, F. N. (2005). *How many years of life could be saved if malaria were eliminated from a hyperendemic area of northern Ghana? (Working Papers*. No. 203). New York: Population Council.
Dye, C., Scheele, S., Dolin, P., Pathania, V., & Raviglione, M. C. (1999). Global burden of tuberculosis: Estimated incidence, prevalence, and mortality by Country. *Journal of the American Medical Association*, *282,* 677–686.
Fontaine, O., & Boschi-Pinto, C. (2006). Special section: Diarrheal diseases. In M. M. Kent & S. Yin (Eds.), Controlling infectious diseases (pp. 10–13). *Population Bulletin*, *61*(2), 1–52. Washington, DC: Population Reference Bureau.
Gallup, J. L., & Sachs, J. D. (2001). The economic burden of malaria. *American Journal of Tropical Medicine and Hygiene*, *64*(1, 2 Suppl.), 85–96.
Garrett, L. (1994). *The coming plague: New emerging diseases in a world out of balance*. New York: Penguin.
Gribble, J. N., & Preston, S. H. (Eds.). (1993). *The epidemiological transition: Policy and planning implications for developing countries* (Workshop Proceedings). Washington, DC: National Academy Press.
Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., et al. (2007). Control of neglected tropical diseases. *New England Journal of Medicine*, *357*(10), 1018–1027.
Jamison, D. T. (1993). Disease control priorities in developing countries: An overview. In D. T. Jamison, W. H. Mosley, A. R. Measham, J. L. Bobadilla (Eds.), *Disease control priorities in developing countries* (pp. 3–34). Oxford: Oxford University Press.
Janzen, J. M., & Feierman, S. (Eds.). (1979). *The social history of disease and medicine in Africa.* Special issue of *Social Science and Medicine 13B*(4).
Kent, M. M., & Yin, S. (2006). Controlling infectious diseases. *Population Bulletin*, *61*(2), 1–52. Washington, DC: Population Reference Bureau.
Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., & Murray, C. J. L. (Eds.). (2006). *Global burden of disease and risk factors*, *1990–2001.* New York: Oxford University Press and World Bank.
Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease: 2002 to 2030. *PLoS Medicine*, 3(11), e442.
Mathers, C. D., Lopez, A. D., & Murray, C. J. L. (2006). The burden of disease and mortality by condition: Data, methods, and results for 2001. In A. D. Lopez et al. (Eds.), *Global burden of disease and risk factors*, 1990–2001. New York: Oxford University Press and World Bank.
Molyneaux, L. (1985). The impact of parasitic diseases and their control, with emphasis on malaria and Africa. In J. Vallin & A. Lopez (Eds.). *Health policy, social policy and mortality prospects* (pp. 13–44). Liege, Belgium: Ordina Editions.

Murray, C. J. L., & Lopez, A. D. (Eds.). (1996). *Summary: The global burden of disease*. Cambridge, MA: Harvard University School of Public Health, World Health Organization, and World Bank.

Murray, C. J. L., & Lopez, A. D. (Eds.). (1996). *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors, and projected to 2020.* Cambridge, MA: Harvard University School of Public Health, World Health Organization, and the World Bank.

National Academy of Sciences. (1993). *The epidemiological transition: Policy and planning implications for developing countries*. Washington, DC: National Academy Press.

National Academy of Sciences, Institute of Medicine. (2001). *Neurological, psychiatric, and developmental disorders: Meeting the challenge in the developing world* (Brief Report). Washington, DC: National Academy Press.

Olshansky, S. J., Carnes, B., Rogers, R. G., & Smith, L. (1997). Infectious diseases: New and ancient threats to world health. *Population Bulletin,* 52(2), 1–52. Washington, DC: Population Reference Bureau.

Parris, B. (2004). False economies: A global health crisis. *OECD Observer, 243.*

Ratzan, S. C., Filerman, G. L., & LeSar, J. W. (2000). Attaining global health: Challenges and opportunities. *Population Bulletin,* 55(1), 3–48. Washington, DC: Population Reference Bureau.

Reddy, K. S. & Yusuf, S. (1998). Emerging epidemic of cardiovascular diseases in developing countries. *Circulation*, *97*(6), 59–601.

Reichman, L. B., & Tanne, J. H. (2002). *Time bomb: The global epidemic of multi-drug resistant tuberculosis*. New York: McGraw Hill.

Rheingans, R., Dreibelbis, R., & Freeman, M. C. (2006). Beyond the millennium development goals: Public health challenges in water and sanitation. *Global Public Health*, *1*(1), 31–48.

Timaeus, I., Chackiel, J., & Ruzicka, L. (Eds.). (1996). *Adult mortality in Latin America*. Oxford: Clarendon Press.

World Health Organization. (1993). *Implementation of the global malaria control strategy* (Report of a WHO study group on the implementation of the global plan of action for malaria control, 1993–2000). Geneva, Switzerland: World Health Organization.

World Health Organization. (2000). *Economic costs of malaria are many times higher than previously estimated* (Press release: African Summit on Roll Back Malaria, Abuja, Nigeria." WHO/28). Geneva, Switzerland: World Health Organization.

World Health Organization. (2004). *World health report 2004*. Geneva, Switzerland: World Health Organization.

Yin, S. (2006). Special section: Malaria. In M. M. Kent & S. Yin (Eds.). *Controlling infectious diseases* (pp. 14–16). *Population Bulletin*, *61*(2), 1–52. Washington, DC: Population Reference Bureau.

HIV/AIDS

Bongaarts, J., Buettner, T., Heilig, G., & Pelletier, F. (2008). Has the HIV epidemic peaked? *Population and Development Review*, *34*(2), 199–224.

Degrées du Loû, A. (1999). Reproductive health and AIDS in sub-Saharan Africa: Problems and prospects. *Population* (Paris) 4, 1998. *Population: An English Selection*, 11, 61–88.

Lamptey, P. R., Johnson, J. L., & Khan, M. (2006). The global challenge of HIV and AIDS. *Population Bulletin, 61*(1). Washington, DC: Population Reference Bureau.

National Academy of Sciences. (1996). *Preventing and mitigating AIDS in Sub-Saharan Africa: Research and data priorities for the social and behavioral sciences*. Washington, DC: National Academy Press.

Ostrow, D. G., & Kessler, R. C. (Eds.) (1993). *Methodological issues in AIDS behavioral research*. New York: Kluwer Academic Publishers.

Terceira, N., Gregson, S., Zaba, B., & Masu, P. (2003, July). Contribution of HIV to fertility decline in rural Zimbabwe, 1885–2000. *Population Studies*, *57*(2), 149–164.

UNAIDS/WHO. (2007). *2007 AIDS Epidemic update*. Geneva, Switzerland: UNAIDS/WHO.

United Nations. (2007). *World population prospects: The 2006 revision. Extended data set*. CD-ROM. New York: United Nations.

Chapter 12
Interrelations of Health, Mortality, and Age Structure

In the twentieth century demographers focused on the tremendous growth of the world's population, particularly the high fertility rates in poor regions of the globe that accounted for burgeoning shares of children living in seriously depressed circumstances and the rapid increases in the number of elderly persons. In the twenty-first century they may be expected to turn their attention mainly to the rapid aging of the world's population and to the persistently low fertility and the prospective decline in national numbers of the more developed countries. The age structure of a population has an important effect not only on its levels of mortality and health but also on the characteristics of persons who are ill and on the illnesses that beset the population. At the same time the levels of mortality and health affect the age distribution of a population and the health characteristics of persons in the various ages. Some of these relations have already been mentioned in earlier chapters, as in our explanation of the rationale for age-adjusted death rates (Chap. 3), our use of life table measures to compare the mortality of different populations (Chap. 4), and our recognition of the impact of the death rate of a population on its growth rate and, in turn, on its age structure. This chapter will explore the interrelations of mortality/health and age structure further.

I first consider the concepts and measures of aging, recent trends in population age structure and aging, and the demographic factors accounting for these changes. For the More Developed Countries (MDC) the data are sufficiently accurate to permit a direct use of the published materials in making such analyses. For the Less Developed Countries (LDC) the published data are usually defective and the mortality rates and the age structure have to be estimated by indirect means in order to make a satisfactory evaluation of the effect of age structure on mortality and health and the reverse. There are numerous procedures for ascertaining the demographic parameters of the populations of the LDC. I discuss some of these methods, including the concept, uses, and design of model life tables and model age distributions.

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_12, © Springer Science+Business Media B.V. 2012

Concepts and Measures of Aging

Concepts of Aging

Individual Aging

It is important at the outset to distinguish two demographic concepts of aging, individual aging and population aging, and to distinguish these from the concept(s) of biological aging. Individual aging refers to the chronological aging of individuals or groups of individuals such as birth cohorts. It is the process by which individuals gain a year of age with the passage of a calendar year, or gain x years of age with the passage of x years of time. Demographers simply call this process aging. The demographic factor mainly responsible for changes in the numbers in an aging cohort is mortality or the age-specific risk of dying, although net migration may also be involved. Mortality has its greatest direct effect in early childhood and old age. Net migration normally makes its largest direct impact at the young adult ages where mortality tends to be low. Fertility is not a factor in the changes in the size of any birth cohort as it grows older, since fertility's contribution to the size of the cohort is completely determined at its start.

Population aging in brief refers to the fact of a population getting older as compared with the aging of individuals. A variety of measures have been devised to measure population aging. Inasmuch as much of our current discussion involves population aging, we postpone this material until after a brief note on biological aging, which is a principal item of discussion in the next chapter.

Biological Aging

Biological aging, a term often used by molecular biologists and biodemographers, has a different focus. It refers to the increasing vulnerability of living things to deleterious influences both from within the body and from the environment as the individual grows older. The term is also used to refer to the accumulated damage to the molecules, cells, tissues, and organs of the body that occurs as an individual get older, particularly after the start of the reproductive period. In the latter meaning, biological aging is also called senescence. Like the demographer's concept of individual aging, biological aging is a characteristic of individuals, but it concerns the somatic and psychological concomitants of chronological aging processes, especially the molecular/cellular declines besetting the individual as she or he gets older. This concept of aging is troubling to some analysts because it defines aging entirely, or almost entirely, in negative terms. A modified definition for aging proposed as an antidote to the conventional biological concept is one that encompasses both the positive and negative changes that individuals experience with advancing age. This modified concept allows for the gains in wisdom, judgment, creativity, freedom to pursue largely discretionary activities, and other positive traits and feelings that individuals may experience as they grow older.

Population Aging

The concepts of aging designated as individual aging and biological aging are of interest in relation to the study of population age structure mainly because of their role as determinants of population aging. Population aging in contrast to individual aging and biological aging is a characteristic of a population. Measures of population aging function as measures of change in population age structure.

Measures of Population Aging

Conventional Measures

Conventionally population aging is measured by various demographic indicators reflecting a rise in the "age" of the population, such as an increase in the proportion of persons 60 or 65 years old and over, a rise in the median or mean age of the population, or an increase in the ratio of persons 65 years and over to persons under 18 years. The most commonly used measure of population aging is the proportion of the population 65 years and over although the United Nations prefers using the population 60 and over, presumably because of the lesser longevity and much smaller proportion of older persons in the LDC than in the MDC. Table 12.1 presents values for these measures at decennial intervals for the period from 1900 to 2050 for the United States, and Fig. 12.1 depicts the trend in U.S. aging in terms of aging indexes for the period, 1950–2050, with 1950 as the base year. The various measures indicate the same general aging pattern over the 150-year period from 1900 to 2050, but they differ greatly in their slope.

It is possible for a population to grow younger although this has been very uncommon historically. Population "younging" is indicated by a decline in the "age" of a population, such as a decline in the proportion of persons 65 years and over or the ratio of persons 65 years and over to persons under 18 years (e.g., United States, 1990–2000). It is also possible for a population to grow older and younger at the same time, as when the proportion of persons 65 and over and the proportion of children under 18 both increase and the proportion in the intermediate ages decreases (e.g., United States, 1950–1960).

Aging status of countries. A population's age structure is determined by all three factors of population change – fertility, mortality, and migration – but a country's recent fertility history largely explains its general age structure and hence its classification as young or old. The various countries of the world may be classified with reference to their aging status on the basis of the percent of the population 65 years of age and over or the mean age of the population, but the two measures are highly correlated and so the first measure can conveniently be employed to distinguish the countries in this regard.

A proposed classification scheme for the aging status of countries is: 20.0% and over, very old; 15.0–19.9%, old; 10.0–14.9%, middle aged; 5.0–9.9%, young; and

Table 12.1 Comparison of various measures of population aging: United States, 1900–2050

Year	Percent 65 years and over	Median age	Mean age	$\frac{65+}{0-17} * 100$[a]	$\frac{65+}{0-64} * 100$[b]	Slope of age distribution[c]
Census						
1900 (June 1)	4.1	22.9	26.3	10.0	4.2	8.0
1910 (April 10)	4.3	24.1	27.2	11.6	4.5	7.9
1920 (January 1)	4.7	25.3	28.1	12.4	4.9	8.7
1930	5.4	26.5	29.4	15.6	5.7	14.5
1940	6.8	29.0	31.6	22.4	7.4	9.2
1950	8.1	30.1	32.1	26.2	8.9	10.5
1960	9.2	29.5	31.7	25.8	10.2	14.4
1970	9.9	28.1	32.0	28.7	10.9	9.7
1980	11.3	30.0	34.0	40.1	12.7	9.2
1990	12.6	32.8	35.3	48.6	14.3	12.3
2000	12.4	35.3	36.3	48.4	14.2	15.1
Estimate						
2005	12.4	36.2	37.3	50.0	14.2	13.9
Projections[d]						
2010	13.0	36.0	37.8	54.1	15.0	14.4
2020	16.3	37.0	39.1	68.0	19.4	16.6
2030	19.6	38.0	40.2	83.4	24.5	19.6
2040	20.4	38.1	40.8	87.3	25.7	23.0
2050	20.7	38.1	41.0	88.0	26.0	22.1
Amount of change, selected periods						
1900–1950	4.0	7.3	5.8	16.2	4.7	2.6
1950–2000	4.3	5.1	4.2	22.2	5.3	4.5
2000–2030	7.2	2.7	3.9	35.0	10.3	4.5
2030–2050	1.1	0.1	0.8	4.6	1.5	2.5

Source: U.S. Census Bureau, various decennial census reports; estimates and projections; www.census.gov

Census figures as of census date, April 1, unless shown otherwise. Figures refer to U.S. resident population. Estimates and projections as of July 1

[a]Ratio of the population 65 years and over to the population under 18 years per 100

[b]Ratio of the population 65 years and over to the population under 65 years of age per 100

[c]Number of ages corresponding to one-quarter of one percent change in the age distribution from the average population aged 22–24 to the average population aged 82–84

$$\text{Formula}: .0025 \div (\text{Ave.}\, P_{22} \ldots P_{24} - \text{Ave.}\, P_{82} \ldots P_{84}) * 60$$

where P = Percent in single-year-of-age group

[d]Middle series. Base date of projections is July 1, 2004

under 5.0%, very young. The population of most, less developed countries falls in the "very young" or "young" categories, and the population of most, more developed countries falls in the "old" category (Table 12.2). Several, more developed countries

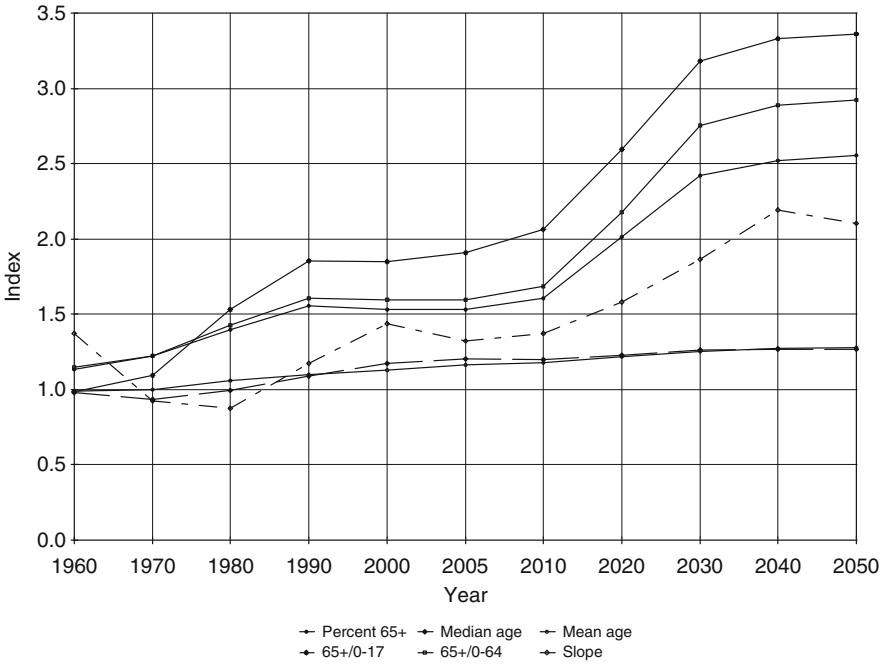


Fig. 12.1 Comparison of indexes of aging of the United States population: 1950 (base) to 2050 (Note: In interpreting the figure, note that the x-scale showing "year" exaggerates the trend in the period 2000–2010; Source: See Table 12.1. Primary source: U.S. Census Bureau, decennial census data and population projections published in 2004)

(e.g., United States) are now "middle-aged," but they will move rapidly into the "old" category and then into the "very old" category in the next few decades. Nearly all countries of Africa are "very young," and most countries of Latin America and Asia are "very young" or "young." With 22% of its population 65 or over in 2008, Japan is a special case; it is already "very old." China is still "young" with 8% of its population 65 or older, but it is destined to age rapidly in the next few decades and by 2025 will be "very old." With the exception of most of Eastern Europe and a few western European countries (e.g., Iceland, Ireland), the countries of Europe fall into the "old" category and are destined to become "very old" within a decade or two.

Other Measures of Population Aging

Several other measures of population aging have been proposed, but they have not gained general acceptance. Some are based on the age distribution of the observed population, some are based on the age distribution of observed deaths, and some are based on current life tables. One measure based on the age distribution of the observed population is the slope of the age distribution, either the entire

Table 12.2 Age structure and societal age-dependency ratios, for world regions and selected countries: 2005

	Population (%)		Dependency ratio (per 100)		
Region and country	Under 15	65 and over	Total[a]	Child[b]	Elderly[c]
More developed regions	17	15	47	25	22
Less developed regions	32	5	59	51	7.9
Africa	42	3	82	76	5.5
Sub-Saharan	44	3	89	83	5.7
Northern	36	4	67	60	6.7
Western	44	3	89	83	5.7
Nigeria	43	3	85	80	5.6
Eastern	44	3	89	83	5.7
Middle	46	3	96	90	5.9
Southern	34	4	61	55	6.5
Northern America	21	12	49	31	18
United States	21	12	49	31	18
Mexico	31	5	56	48	7.8
Central America	33	5	61	53	8.1
Caribbean	29	7	56	45	11
South America	30	6	56	46	9.4
Brazil	29	6	54	45	9.2
Asia	29	6	54	45	9.2
Western	34	5	64	56	8.2
South Central	36	4	67	60	6.7
Southeast	30	5	54	46	7.7
East	21	9	43	30	13
China	22	8	43	31	11
Japan	14	20	52	21	30
Europe	16	16	47	24	24
Northern	18	16	52	27	24
Western	17	17	53	26	26
Eastern	16	14	43	23	20
Russia	16	13	41	23	18
Southern	15	17	47	22	25
Oceania	25	10	54	38	15
Australia	20	13	49	30	19

Source: Basic population data: Population Reference Bureau (2005). *2005 World Population Data Sheet*. Washington, DC: Population Reference Bureau. Used with permission of the Population Reference Bureau. Dependency ratios computed by the author from data in this report

[a]Ratio of children (under 15) and elderly (65 and over) to population of principal working ages (15–64 years).

[b] Ratio of children (under 15) to population of principal working ages (15–64 years).

[c]Ratio of elderly (65 and over) to population of principal working ages (15–64 years).

age distribution or the age distribution after some adult age, say age 50 or 60. The slope can be determined, for example, by fitting a straight line to the grouped age data from age 50 on. An increasingly downward slope in the curves at successive

dates would suggest aging of the population. Alternatively, a parabola can be fit and increasing rectangularization of the curve would suggest aging. For a simple measure of the slope, I calculated the average number of ages covered by one percent of the population between age 23 and age 82 at successive decennial dates (Table 12.1).

A number of indirect measures of population aging are based on actual death statistics or life table deaths. This group of measures includes the median age at death in the observed population 25 years and over, the median age at death at ages 25 years and over in the current life table, or total life expectancy at birth or at age 25. As we saw, the median age at death of a population is a function of both the age distribution of the population and the pattern of age-specific death rates. In low mortality-low fertility populations with declining old-age mortality, the domination of the age-pattern of mortality by the high age-specific death rates at the older ages, in combination with the high and increasing proportion of elderly persons, means that most deaths occur at the older ages and that the average age of deaths has been rising. Because of this combination of causes, the indication of the aging of the population is exaggerated by changes in the median age at death as compared with other measures. Changes in life expectancy, particularly at the adult ages, may be used as another indirect measure of population aging. Rises in life expectancy at these ages are consistent with the aging of the observed population, but the measure takes no direct account of the actual population age distribution and at low levels of mortality tends to understate the rate of population aging.

Years until death and prospective proportion of elderly. There has been a tremendous "delay" in the experience of mortality in the last half century. An example will serve to remind the reader of this shift. In 1949–51, life expectancy at age 60 was 17.0 years while in 2000 life expectancy did not fall to this level until age 66.2. Because of the considerable declines in mortality in the last century, the associated massive shift in the age composition of the populations in this period, particularly in the industrialized countries, and the historical change in the concept of "old age," some demographers have proposed the introduction of new measures of population aging that specifically take the rise in longevity into account.

Accordingly, the age in the life table with a specified number of years until death, that is, the age corresponding to a specified life expectancy, has been proposed as the basis for a measure of defining old age (Siegel 1993). This method of measuring old age has been referred to as "counting backward from death." The period of 10 or 15 years is arbitrarily selected as the duration of old age and the age corresponding to a life expectancy of 10 or 15 years is taken as the lower bound of old age. As life expectancy at the older ages rises, the period of old age begins at a progressively later age. Accordingly, the age of onset of old age fluctuates in accordance with the level of death rates at the older ages. Old age would begin at higher ages for populations that have comparatively low death rates at the older ages and at lower ages for populations that have comparatively high death rates at the older ages.

Table 12.3 sets forth a time series of the ages at which the survivors in a series of U.S. life tables from 1950 have a life expectation of 10 years and15 years. Historical

Table 12.3 Age at which average remaining lifetime equals 10 or 15 years, and the percentage of the total population above this age, for the United States: 1950–2004 and projections, 2010–2050

	10 years of average remaining life		15 years of average remaining life		
	Age at which average remaining life equals 10 years	Percentage of total population above this age	Age at which average remaining life equals 15 years	Percentage of total population above this age	e_{65}
Year					
1935					12.5[a]
1949–1951[b]	71.7	4.0	63.1	9.6	
1959–1961	72.5	4.3	64.0	10.0	
1965					14.6
1969–1971	73.7	4.4	65.0	9.9	
1979–1981	75.9	4.4	67.3	9.4	
1989–1991	76.9	4.7	68.4	9.9	
1999–2001	77.0[c]	4.8	68.9	9.8	
2004	78.4	4.4	70.2	9.0	18.7
Projections[d]					
2010	77.1	5.2	69.3	9.6	
2030	78.4	6.6	70.8	13.4	
2050	79.7	8.2	72.2	15.8	
Increase					
1950–2000	5.3	0.8	5.8	0.2	
2000–2050	2.7	3.8	3.3	6.0	
1935–2004					6.2
1965–2004					4.1

Sources: Based on various decennial and annual life tables published by the U.S. NCHS, and decennial life tables for 2010, 2030, 2050 published by the Actuary's Office, U.S. Social Security Administration (2005). Population data from U.S. Bureau of the Census, decennial census reports, and projections, www.census.gov

[a]Estimated

[b]Conterminous United States (excluding Alaska and Hawaii)

[c]U.S. Actuary' Office figure for 2000 in its historical and projected series of life tables is 76.8

[d]Intermediate series

figures for 1950–2004 are derived from reports of the National Center for Health Statistics and projections from 2010 to 2050 are derived from material published by the Actuary's Office, U.S. Social Security Administration (2005). The age at which average remaining lifetime equals only 10 years rose steadily from 71.7 years in 1950 to 77.0 years in 2000, and is expected to reach 79.7 years by 2050. Thus, an increase of 5.3 years in the age at which an average person had 10 years of life remaining was recorded in the half century from 1950 to 2000 and an increase of 2.7 years is expected in the half century from 2000 to 2050.

A choice more concordant with the public's current view of old age would designate the last 15 years of life as the period of old age; thereby, old age begins earlier and lasts longer. The age corresponding to a life expectancy of 15 years rose from 63.1 years in 1950 to 68.9 years in 2000 and is expected to reach 72.2 years by

2050. According to this series, an increase of 5.8 years in the age at which old age starts was recorded for the half century from 1950 to 2000 and an additional increase of 3.3 years in this age is expected for the first half of this century. The series of ages initiating old age rises because age-specific death rates have been declining and are expected to decline further at the higher ages.

The concept of years until death could serve as the basis for a new measure of population aging. Specifically, the demarcation line for old age could be a variable line that moves upward as life expectancy, or age at "15 years until death," rises. While such a measure has limitations as a general measure of old age, it has some potential administrative and legal uses. Such a linkage of the definition of old age to changing longevity may, for example, be the basis for defining old age in programs where the financial viability of the program is affected by the length of life (e.g., a Social Security retirement program).

"Years until death" is not a measure of population aging per se, however, since it is based on only a limited age segment of the life table and does not take account of the observed age distribution. To convert this measure of old age into a measure of population aging, that is, one linked to the actual age distribution, we need to determine the proportion of the actual population above the life-table age corresponding to a life expectancy of 10 or 15 years. The results for 1950–2050 are shown in Table 12.3. From 1950 to 2000 the proportion of the population above the age corresponding to a life expectancy of 10 years moved up slowly between four and five percent, even as this age moved up from 72 to 77. From 2000 to 2050, however, the proportion is expected to rise sharply to eight percent. A roughly similar pattern is observed when 15 years of remaining life is chosen as the threshold of old age.

As an alternative measure of population aging, this proportion is problematic. The proportion does not necessarily rise in the past years because the age at which people have 10 or 15 years of remaining life is also rising. While this measure of population age structure links shifts in mortality to the actual age structure, it is still heavily dependent on changes in mortality, as is the underlying measure, "years until death." This population proportion may fluctuate erratically. Since a social security system seeks some stability and continuity in the balance of age groups (e.g., contributors vs. beneficiaries), having an aging measure that is relatively stable and continuous when the "age until death" rises cannot always be assured.

To describe a complement of measures of population aging involving "years until death," Sanderson and Scherbov (2008) introduced the concept of "prospective age" and built on it in order to develop modified measures of the proportion of elderly in the population, the old-age dependency ratio, and the median age of the population. These measures also shift in relation to changes in life expectancy in later life. Sanderson and Scherbov assume that the changes in life expectancy are accompanied by general changes in health and that persons are as functional at the prospective age as at the conventional age. They have produced estimates of these measures for all countries from 1955 to 2045. (See www.prb.org/Publications/PopulationBulletin/2008/globalaging.aspx).

The measures are called prospective proportion of elderly in the population, prospective old-age dependency ratio, and prospective median age. The calculation of the first measure is carried out in the same fashion as I described above for the series in Table 12.3. The elderly population is that population with ages above the age with a life expectation of 15 years in the current life table. The prospective aged dependency ratio is computed in a fashion analogous to the proportion of the population elderly; the aged dependency ratio is discussed later below. The prospective median age is explained further in footnote.[1]

Changes in Age Groups

The population at any age can be viewed as the survivors of a particular birth cohort. The size of the original birth cohort is modified, as it reaches older and older ages with the passage of time, by the effect of mortality and the flows of immigrants and emigrants from the time of birth. To measure the demographic changes in the cohort, we are concerned only with the mortality rates and migration flows affecting this particular cohort as its age reference changes with the passage of time, not with the rates or flows affecting other cohorts or with any fertility changes. Change in every age group can be interpreted in this way, up through the oldest age of interest. Compare the situation in the standard life table. The radix, a birth cohort with an assumed size of 100,000, is gradually reduced by age-specific mortality rates until it is entirely extinguished by mortality a hundred or so years later, migration being excluded.

A change in the same population age group (i.e., with the same age identification) over time between two dates requires a more elaborate explanation. This change depends on the shift in the sizes of two birth cohorts born in different periods and the shifts in the survival rates affecting the two cohorts from birth up to the current ages of interest. The Lexis diagram given as Fig. 12.2 depicts how two populations in the same age group at different dates, ages 15–19 in 1980 and 2000, arise from two cohorts born at different prior periods, 1960–1965 and 1980–1985. Because of the unique origins of the two populations and the dependence of the two cohorts

[1]The prospective median age is calculated in a special way, using a standard life table for all populations being compared. Sanderson and Scherbov take the United States life table for 2000 as the standard life table. They define prospective age as the age of a person in a standard life table who has the same remaining life expectancy as the person of interest, and the prospective median age of a population in a particular year as the prospective age of median-aged persons in the population in that year. A few examples will help clarify these definitions. The median age in the United States in 1980 was 30.0 years and a life expectation of 30 years in the life table for 2000 corresponds to an age of 50.0 years. We then select the age corresponding to a life expectation of 50.0 years as the prospective median age, or 28.2 years. In the year 2000, the median age of 35.3 years is the life expectancy at age 44.0 years in the life table of that year, and a life expectancy of 44.0 years corresponds to age 34.0, the prospective median age. Accordingly, the prospective median age has risen by 5.8 years in this half century, as compared with the observed change of 5.3 years.

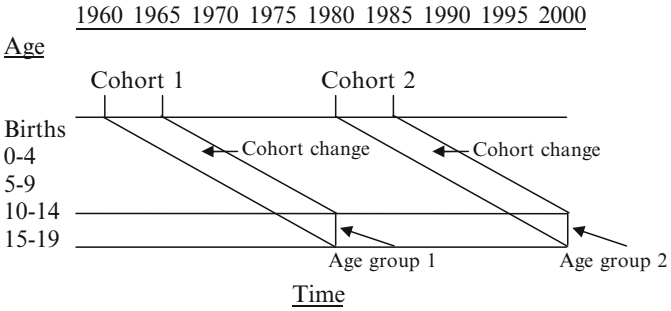


Fig. 12.2 Lexis diagram depicting the relation of age and time for two cohorts aged 15–19 years at dates 20 years apart (Note: Relevant birth cohorts and corresponding age groups are bounded by the vertical lines)

on their own sets of survival and migration rates, the percent change in the size of an age group between two dates is largely independent of the percent change in any other age group between these dates. Corollaries of this generalization are that the percent change in the total population between the two dates is essentially independent of the percent change in any age group between the same two dates and that the percent change in the total population cannot be inferred with confidence from the change in any age group during the same period. A similar type of analysis would apply to differences between two age groups in the same year.

Determinants of Population Age Structure and Its Changes

As mentioned, with rare exception the fertility level of a population and its recent history determine the general contour of the age distribution of the population. This is illustrated in Fig. 12.3, where the series of age pyramids for the world's population are reshaped by the historical and prospective decline in fertility, modified somewhat by the historical and prospective decline in mortality.

Typically, mortality levels and migration flows (for national areas) are secondary influences. In both the MDC and the LDC, whether mortality levels are relatively low or high, the effect of mortality on the age distribution of the population is tempered by the fact that mortality takes a toll of persons at all ages. Migration also tends to affect the numbers at all ages. Fertility, on the other hand, directly affects the youngest ages only. Even where migration has been a major factor contributing to population growth, as in the United States, the general shape of the age distribution mainly reflects the fertility history of the country. In the MDC low mortality tends to bolster the effect of low fertility. In the LDC, where higher mortality prevails, mortality is an influential secondary factor in its impact on the age distribution, especially where fertility has fallen sharply, as it has in many countries.

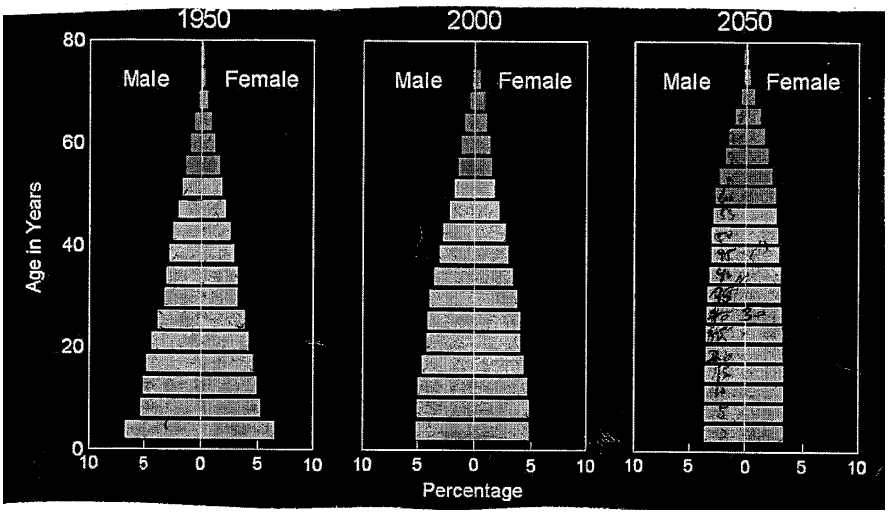


Fig. 12.3 Age-sex structure of the population of the world: 1950, 2000, and 2050 (Population pyramids, 5-year age groups) (Source: United Nations, Population Division (2002). *Population prospects. The 2002 revision, Vol. II. The sex and age distribution of the world population*. E.03.XIII.7. New York: United Nations)

The interpretation is different if we are trying to explain a shift in the proportion at some age group as compared to the general contour of an age distribution and a shift in the numbers in the age group. Suppose we inquire why a population ages, that is, why it experiences, for example, an increase in the proportion of persons 65 years and over. Leaving immigration aside, we can identify four factors that account for population aging in the United States in particular and the MDC in general at the present time: Relatively greater increases in survival rates at the later adult ages than at the younger ages in recent decades, increased survival from birth to the older ages, increases in the annual number of births 65 or more years earlier, and recent declines in birth rates. The considerable reductions in death rates in childhood and youth in the last century, combined with the more recent reductions of death rates in the middle and older ages, have resulted in an increase in the number and proportion of persons reaching old age and surviving well into old age. Just as important have been the gradual rise in the number of births 65 years ago, which adds to the potential survivors, and the long-term decline in the birth rate, which directly causes an upward shift in the share of the population of older age. As a result of these changes, the younger population has tended to grow more slowly than the population in the older ages and the population has become older.

The above discussion suggests that the relative roles of fertility, mortality, and migration can shift when we consider their impact on changes in the age distribution over time as compared with their role in determining the general shape of the age distribution at some specific date. The demographic history of the United

States in the twentieth century provides an illustration of how the relative role of the demographic components of change can shift under different circumstances. Throughout that century (with a brief exception) the population was aging; that is, the proportion of the population 65 years and over and the mean age were rising (Table 12.1 and Fig. 12.1). Up to about 1970 the principal demographic factor accounting for this change was the decline in fertility. Because fertility changes are concentrated at the beginning of the age distribution, declines in fertility always tend to make the population grow older. Mortality declines were concentrated in the ages below the mean age of the population even though they were dispersed to some extent over all ages. (See Table 4.4) Net immigration was also concentrated among youth. When mortality declines and immigration influxes are concentrated among the younger ages, they tend to make the population younger. Mortality changes and immigration tended to offset the effect of fertility changes only in part, however, and the population grew steadily older.

After about 1970, as described in Chap. 6, the mortality declines shifted to the older ages. To a lesser extent this was true of migration also, i.e., the ages of migrants were more dispersed. As a result, mortality began to contribute to the aging of the population rather than to its younging and in fact became the primary factor causing the population of the United States to age (Hermalin 1966). A similar drop in death rates at the older ages occurred in the other Western countries during the same period and this development has contributed to the aging of their populations also (Preston et al. 1989). An operative factor in this process is the change from one date to another in survival rates for particular age ranges. Comparison of changes in sets of survival rates for 1900–1950 and 1950–2000 shows larger increases at the ages above 50 than at the ages below 50 in the second period as compared with the first period (Table 4.4). An analysis of the separate contribution of fertility and mortality to the aging of the population of the United States between 1940 and 1970 and between 1970 and 2000 shows that the relative importance of fertility and mortality was shifting between these periods.[2]

[2]There are alternative ways of decomposing the total change in the proportion of persons 65 and over between two dates into the contributions of the components of change. They generally employ the simulation technique of holding each of the components constant at their base-year value for the test period and recomputing the proportion over 65 in the terminal year. For example, the age-specific birth rates and death rates of the initial year can be sequentially substituted for the actual rates from the initial year to the terminal year of the test period. For net immigration, it may be assumed that, after the initial year, the flow of net immigration stops until the end of the period. Another technique of discounting the effect of net immigration is to remove the net contribution of the foreign-born to the change in the percent 65 years and over during the test period. The latter procedure recognizes that immigration has an indirect effect as well as a direct effect. Each component can be expressed in two forms, as follows:

Net immigration (direct effect only), births, and deaths, or
Net immigration and its natural increase, births excluding births of immigrants and their children, and deaths excluding deaths of immigrants and their children.

Trends and Consequences of Population Aging

Trends in Population Aging

United States

Figure 12.4a, b depict the historical and prospective trends in the number of persons 65 years and over and the proportion of persons 65 years and over in the United States from 1950 to 2050. The number has shown a steady and rapid rise since 1900. It was then about 3 million, by 1950 it had grown to about 12 million, and in 2000 it was over 35 million. Under the middle series of projections the number is expected to double between 2000 and 2030 – the period when the baby-boom cohorts will arrive at age 65 – and to reach 89 million by 2050. The rate of growth in the number was very rapid in the first half of the twentieth century, fell somewhat in the second half, and is expected to continue at the reduced, level during the first half of this century. The percent of elderly persons grew on a roughly linear course during most of the twentieth century, and is now experiencing a brief two-decades pause (1990–2010) before it moves sharply upward for the next few decades (2010–2030). By 2050, the percent is expected to be about three-fifths greater than in 2010, rising from 13% to 21% in merely three decades.

The historical decline in the birth rate has been the principal factor in the long-term aging of the U.S. population. Declines in the birth rate since 1965 have reinforced this historical trend. In the decades since 1965, however, sharp declines in death rates at the older ages have occurred and have become a dominant factor in the aging of the population. The effect of these declines has been reinforced by the declines in mortality rates at the middle ages; they have brought more survivors to old age. With the advent of the huge baby-boom cohorts of 1946–1964 to age 65 and the enhanced effect of the mortality factor, we can expect a massive increase in the number of elderly persons in the two decades from 2010 to 2030. These factors will combine with the decline of the birth rate to produce a sharp rise in the proportion of elderly persons.

Most women in the United States, as in other industrial countries, are choosing to have only one or two children in their lifetime, and although some couples are having more children, the average number – i.e., the total fertility rate – is not enough or barely enough for replacement. Many factors have contributed to the fall in fertility. In addition to the increase in the level of education of the population and the decline in child mortality, the increase in the equality between the sexes, the growing general secularism, the pursuit of "self-fulfillment," the proportion of couples cohabiting, the proportion of women working outside the home, and the proportion of single parents have all also played a role. There is little prospect that these trends will be substantially reversed, and although women report that they consider that more children would be ideal, they are not acting on this notion.

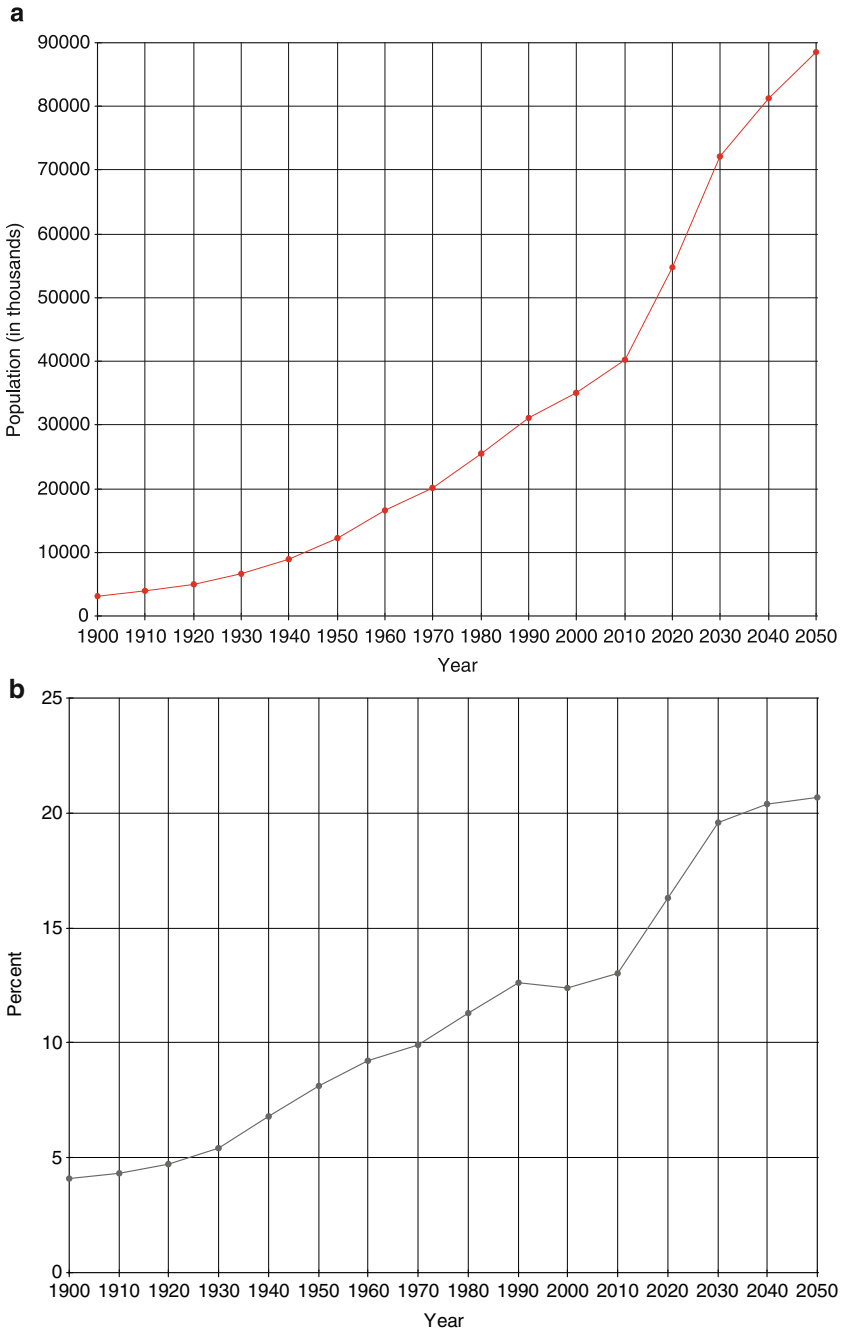


Fig. 12.4 (**a**) Trend in the number of persons 65 years and over: United States, 1900–2050. (**b**) Trend in percent of population 65 years old and over: United States, 1900–2050 (Source: U.S. Bureau of the Census, census data and middle series of projections published in 2008. See www.census.gov)

International Trends

As measured by the change in the proportion of elderly persons, in the last decades of the last century, the MDC were aging rapidly while the LDC were aging very slowly or not at all. Because of high fertility and high mortality, the share of elderly persons in the LDC has tended to remain very low – 5% as compared with the MDC figure of 15% in 2005 – except in some low-fertility countries such as Cuba and China (where government policy calls for one-child families). (See Table 12.2) However, the numbers of elderly persons have been increasing rapidly for many decades both in the LDC and the MDC. This is because, with a rising population and declining mortality, birth cohorts have been growing increasingly larger, especially where fertility rates were constant or declined only a little in decades past.

Western Europe's population, in particular, will become considerably older over the next half century. Lutz et al.'s (2001) median forecast shows that the proportion of the population above age 60 is expected to increase from its current level of 20–35% by 2050 and to 45% by 2100; the confidence range for the projection for year 2100 is 32–58, with a probability of roughly 80%. Sanderson et al. (forthcoming, quoted by Lutz and Scherbov (2003)) offer a similar prognosis:

> … if life expectancy continues to increase markedly and fertility stays at a low level or declines further (and there is no significant migration), then Europe may well develop into a society with more than half of the population above age 60. The proportion above age 80 is likely to increase even more rapidly. Toward the end of the century Europe may have more inhabitants above age 80 than young people below age 20.

As noted earlier, the combined influences of fertility and mortality on the shape of, and trends in, the age distribution of the world's population is shown in Fig. 12.3. We see how the long-term trends in the declines of the birth rate and the death rate are reshaping the age pyramids, from a nearly perfect triangular form in 1950 to a roughly rectangular tower by 2050.

Consequences of Rapid Population Aging

It is not feasible to present a detailed exposition of the consequences of population aging. They encompass changes in the demographic, social, economic, and biological characteristics of populations. Hence, only selected topics are discussed here. Some of the economic implications of aging are explored in Chap. 15 in connection with the discussion of issues in health policy. Most important of the consequences of aging are the outlook for increased demand for health care and services, the requirements for additional professional resources as we move into the elderly "baby-boom" era, and the need to restore the viability of Social Security and Medicare. Even if mortality and disability rates continue to fall,we may

encounter an expansion of demand for health services because of a new emphasis on comprehensive preventive care and greater use of expensive medical diagnostic and therapeutic technology.

From one point of view, the aging of the population reflects a triumph of society, inasmuch as it means lower death rates and fewer unwanted births, i.e., survival of more persons to higher ages and closer correspondence between the numbers of births that women actually have and their fertility preferences. However, for many developed countries the rapid aging of the population is creating, or is about to create, severe problems for the solvency of their age entitlement programs, i.e., their social security and health insurance systems. They face the possibility of having to reduce benefits for the participants (e.g., raise the retirement age) and/or increase worker contributions (e.g., raise the tax rate on income subject to tax). For most less developed countries, rapid population aging could become an obstacle to implementing or expanding social programs in health and retirement. As described later, the current transitional period in age structure provides an opportunity for these countries to restructure their social insurance systems, with the possibility of improving their prospective financial condition.

The Dependency/Support Problem

The proximate demographic basis for the emerging problem of financial stresses on the entitlement programs is the rapid rise in the age-dependency ratio, that is, the ratio of the number of persons in the principal nonworking ages – children and elderly – to the number of persons in the principal working ages. The ratio may be represented as follows:

$$\text{Total age-dependency ratio} = (P_{<18+65+} \div P_{18-64}) * 100 \tag{12.1}$$

The age-dependency ratio may be partitioned into the child dependency ratio and the elderly dependency ratio:

$$\text{Child age-dependency ratio} = (P_{<18} \div P_{18-64}) * 100 \tag{12.2}$$

$$\text{Elderly age-dependency ratio} = (P_{65+} \div P_{18-64}) * 100 \tag{12.3}$$

The age bands selected may vary. For example, the child population may be represented by the ages under 15 or under 20 and the working-age population by the ages 15–64 or 20–64, respectively. The reciprocal measure, the support ratio, is also used to describe these relations and is the way much of the public tends to conceptualize it. The support ratio expresses the ratio of persons in the principal working ages to persons in the principal dependent ages (although it should be

Table 12.4 Historical and projected aged dependency ratios for the United States: 1950–2050

Historical series		Middle series projections	
Year	Ratio	Year	Ratio
1950	13.4	2010	20.7
1960	16.9	2020	27.2
1970	17.6	2030	34.6
1980	18.7	2040	36.4
1990	20.3	2050	37.0
2000	20.1		
2005	19.8		

Source: U.S. Census Bureau, population estimates for July 1 of year indicated from *Current Population Reports*, Series P-25, and www.census.gov; projections are those released March 2004, www.census.gov
Resident population of the United States as of July 1. Aged dependency ratios calculated as: Population 65 + ÷ Population 18–64, per 100

recognized that notions of support and dependency have changed rapidly in recent decades). For example, the formula for the elderly support ratio is

$$\text{Elderly support ratio} = (P_{18-64} \div P_{65+}) * 100 \tag{12.4}$$

The ratio of the number of children (under 15) and elderly (65 and over) to the number of persons 15–64 in the United States is currently 49 (per 100). In the MDC as a group, the ratio is only a little more favorable (47). In the MDC, the total dependency ratio has already reached a problematic stage because of the long-term rapid aging of the population. The dependency problem in the United States is suggested by the data in Table 12.4, which shows an historical and projected series of elderly dependency ratios, that is, ratios of the population 65 years and over to the population 18–64 years of age. These age bands were selected to represent the principal non-working and working ages in the United States. From 13.4 per 100 in 1950, this ratio has moved irregularly upward to 20.1 per 100 in the year 2000 and, according to the middle projections of the U.S. Census Bureau, it is expected to reach 37.0 per 100 in the year 2050. The increase anticipated in the ratio in the next half century is more than twice that in the past half century. Alternatively viewed, in terms of support ratios, in 1950 there were 7.5 "workers" for each elderly "dependent," today there are 5, and in 2050 we can expect 2.7.

One striking consequence of the aging of the populations of the countries of Europe is the experience of deaths outnumbering births, or of deaths being expected to outnumber births soon. Even now, Germany, several countries of Northern and Eastern Europe, including Bulgaria, Latvia, Lithuania, Estonia, Romania, and Hungary, and some countries of Southern Europe, including Italy and Serbia, are recording natural decreases. In fact, the European Union as a whole will show a

natural decrease by 2015. Then migration will be its only source of growth. Most European countries and the European Union will soon begin to lose population. By 2035 all the countries named above are expected to experience rapid population decline.

In many less developed countries, population aging has been associated with a decrease in the child dependency ratio as fertility fell rapidly and only a modest increase occurred in the elderly dependency ratio. In the first few decades of this century, working-age adults in these countries are expected to have only a moderate dependency burden as fertility falls and aging progresses slowly. The current total dependency ratio is relatively high at 59, largely because of the very high proportion of children. The elderly dependency ratio is quite low at 5 (per 100 persons of working age) because of the very low proportion of elderly persons. This state of demographic affairs provides an opportunity for designing appropriate social policies and programs to deal with the potential crisis in social security and health costs that will develop when the share of elderly persons moves upward more rapidly as a result of the recent declines in fertility. After a further period of declining fertility, the reduced birth cohorts will enter the working ages, the share of elderly persons will rise very rapidly, and the elderly age dependency ratio will begin to rise. With the expected rise in elderly dependency, the elderly will become as numerous as the children. The MDC have nearly reached that point already.

Age-Structural Transitions

Transitions in the age structure of a population occur continuously as fertility, mortality, and migration steadily reshape its age structure. The size of new-born cohorts steadily change, mortality rates change, and the volume and age composition of migration shift. The various birth cohorts move in predictable fashion from a younger age group to an older one as time passes, modifying the proportion of the population in each older age group as it reaches the next stage of the life cycle. In recent decades fertility declines in many less developed countries have been dramatic and caused population growth to slow greatly. Child dependency ratios are falling fast as fertility declines and elderly-dependency ratios are beginning to rise. For these countries, the balance between dependent-age persons and working-age persons has shifted considerably – and in a favorable direction. As a result, the increase in the share of elderly persons in the LDC is expected to remain modest for the first few decades of this century but then to turn rapidly upward.

In most less developed countries today overall dependency is largely determined by child dependency (Table 12.2). In time, aged dependency will become the dominant component of total dependency. This may not occur for many years in some countries, but after a while this transition will have occurred nearly universally. In the LDC, after a grace period with relatively favorable dependency ratios, total and elderly dependency ratios will tend to rise as the share of elderly persons rises.

In sum, the situation in the MDC and the LDC with respect to aging and dependency is quite different currently. As we have seen, the MDC are well on

Table 12.5 Aging of the elderly population of the United States: 1950–2050

Historical series		Middle series projections			
Year	Percent	Year	Percent	Net increase	Percentage points
1950	4.7	2010	15.2	1950–2000	7.4
1960	5.6	2020	13.3	2000–2010	3.1
1970	7.5	2030	13.4	2010–2030	−1.8
1980	8.9	2040	19.2	2030–2050	10.7
1990	9.7	2050	24.1		
2000	12.1				
2005	13.9				

Source: Historical series, 1950–2000, U.S. Census Bureau, census data; 2005 estimate, www.census.gov; and projections, 2010–2050, released March 2004, www.census.gov
Resident population of the United States as of April 1 for historical series; July 1 for projections. Population 85 years old and over as percent of the population 65 years and over

the way to becoming aged populations with very high elderly-dependency ratios. Up to the end of the last century, the LDC not only had extremely low percentages of persons aged 65 and over but their populations were not aging. The absolute numbers of elderly were increasing rapidly, however. Recent sharp declines in fertility and declines in mortality in many of these countries have produced declining percentages of children, low percentages of elderly, and continuing low elderly-dependency ratios (CICRED 2005). However, the current "window of opportunity" will come to an end in a few decades. Impending age-structural transitions will raise elderly-dependency ratios sharply and lead to great increases in the costs of supporting the elderly and providing health care for them.

Other age-structural transitions. At present, in the MDC the elderly population is itself aging. For example, the share of persons 65 years and over who are 85 years and over is increasing. This is a temporary phenomenon, however, and is expected to reverse itself when the baby-boom population begins to reach age 65 in 2011 (Table 12.5). Between 2011 and 2030 the elderly population will be getting younger. Then, when the baby-boom cohorts begin to reach age 85, about 2030, the share of the elderly population which is 85 and older will rise rapidly again. At present also, the ratio of middle-aged adults, say 45–54 years, to persons 75–84 years is rising. This is significant because these middle-aged adults are the persons who may be pressed into supporting their parents at the higher ages, while possibly still having children of their own who need their support (Table 12.6). In the years to come, this ratio will be reversed as the number of persons of advanced age rises in relation to the number at the younger ages.

Increase in Average Number of Generations per Family

A notable effect of the decline in mortality has been an increase in the number of surviving family generations. This change can be represented by the average number

Table 12.6 Three-generation familial dependency ratios for the United States: 1950–2050 (Ratio of persons 15–24 years and 75–84 years to persons 45–54 years)

Year	$\frac{15-24}{45-54}$	$\frac{75-84}{45-54}$	$\frac{(15-24)+(75-84)}{45-54}$
1950	1.277	0.189	1.466
1960	1.173	0.226	1.399
1970	1.526	0.264	1.789
1980	1.683	0.339	2.202
1990	1.477	0.400	1.877
2000	1.040	0.328	1.368
2005	0.990	0.307	1.298
2010	0.974	0.291	1.265
2020	1.072	0.387	1.459
2030	1.111	0.558	1.669
2040	1.085	0.618	1.703
2050	1.123	0.583	1.707

Source: 1950–2000, based on U.S. Census Bureau, decennial census figures; 2005–2050, estimates and projections released in 2008, www.census.gov

of generations per "family". In this conceptualization the family is not confined to the members of the same household but includes members of the extended family in other households. The number of generations is based on the number of tiers in a parent-child line of ascendancy or descendancy. For example, a grandparent-parent-child line equals three generations.

In the United States the average number of living generations in a family has been undergoing a gradual increase in the past several decades, mainly as a result of increasing longevity. We do not know the average number of generations for the United States with any precision because information on family composition is usuallyconfined to the members of the same household in censuses and surveys.[3] Reports regarding generational family size would have to be compiled for every (sample) household or, rather, all individuals in every (sample) household to obtain an unbiased result.

I surmise that families contained about 2.5 generations in 1950 and a little over 3 generations in 2000 – perhaps 3.2 generations. Thus, the average family gained a little over one-half generation in the half century from 1950 to 2000. The trend is

[3] Shanas reported in 1980, on the basis of a national sample survey taken in 1975, that 75% of the elderly were grandparents and that about 38% of the elderly belonged to four-generation families. Shanas was cited in G. O. Hagestad (1986, Winter). The aging society as a context for family life. *Daedelus, 115*(1), 119–139. If the respondents in such surveys are the grandparents, as seems to be the case here, the average number of generations over all families tends to be overstated. This is because only families with elderly members are included in the reporting universe and families without elderly members are omitted.

expected to continue and by 2050 the average family is expected to contain about 3.5 generations. These rough figures assume that the Census Bureau's middle mortality projections prevail. The expected slower growth of the generational size of families results from the assumption of a slower rise in longevity in the future than in the last several decades and a continuation of the practice of delayed marriage and parenthood.

The mean number of generations per family in a population depends mainly on the survival record of the population, including survival from birth to early adulthood and survival from early adulthood to advanced old age. The measure is also dependent on a number of other factors, including the proportion of women who marry, the mean age at first marriage, the mean age at birth of first child, and similar factors. While we are continuing to experience an increase in longevity, which adds to the number of generations, the trends of most of the other factors have undergone reversals. For example, the mean age of women at first marriage and the mean age of women at childbearing have risen well above their levels at mid-century. These changes tend to retard the rise in the mean number of generations. Longevity is the only factor that tends to move in an essentially monotonic course; changes in the other factors are much more likely to fluctuate since they are affected heavily by changes in attitudes, fashions, and socioeconomic conditions.

The decline in mortality rates has been associated with low replacement-level fertility and the reshaping of the family network. The number of generations in a direct line has grown but the number of relatives among the different generations has been contracting. Each generation now has fewer siblings and collateral relatives such as uncles, aunts, and cousins. The "length" of multigenerational families has been expanding while its "width" has been contracting. The combination of the decline in fertility, and the rise in survival in both childhood and the higher ages, have made families into more vertical and less horizontal structures. The relative frequency of ascendant (i.e., parents, grandparents and great grandparents) and descendant (i.e., children, grandchildren, great grandchildren) kin is growing, but the number of brothers, sisters, collateral kin (i.e., uncles, aunts, cousins), and affinal kin (i.e., kin by marriage) is decreasing and is expected to continue decreasing. It is likely that, in spite of increasing longevity, elderly persons of the future will have a smaller total number of closely-related relatives, including surviving children, who may contribute to their support.

Effect on Health Characteristics

The aging of the population in the MDC has affected the demographic characteristics of patients, the types of illnesses and causes of death that are most prevalent, and the types of medical services needed. Here are some illustrations of these shifts using U.S. data. The proportion of deaths at ages 65 years and over among all deaths and the proportion of elderly patients among all patients have been rising. Deaths at ages 65 and over made up 73% of all deaths in 2005 (U.S. NCHS 2008) but only

62% of all deaths in 1970, even though death rates at these ages were higher at the earlier date (U.S. NCHS 1974). Of the persons reporting fair or poor health or an activity limitation due to a chronic condition in 2005, over one-third (34%) were 65 or over according to the National Health Interview Survey (U.S. NCHS 2007a), but in 1995 the percent 65 or over reporting an activity limitation due to a chronic condition was only 30% (U.S. NCHS 1998). In 2005, persons 65 and over accounted for about 38% of hospital discharges and 44% of days of hospital care as compared with 20% and 33%, respectively, in 1970 (U.S. NCHS 2007b).

Earlier chapters referred to the epidemiological transition in the causes of death and disease, particularly in the United States. In a sense the epidemiological transition is both cause and consequence of the aging of the population. The indications of the transition with respect to the leading causes of death and disease have been growing in the second half of the last century although the transition began in the first half. Nevertheless, the number of persons reporting several chronic degenerative illnesses (e.g., kidney disease, diabetes, and arthritis) is increasing at a faster pace than the general adult population. Some of these diseases were already very common decades ago, especially among women. With further aging, the prevalence of such diseases as well as the number of disabled persons and the demand for long-term care have increased and are very likely to increase further, barring unexpectedly great progress in the treatment of the chronic diseases of later life. As these changes occur, so does the need for additional health personnel, equipment, and facilities to serve the increased general and specialized demand.

Health Dependency Ratios

Going beyond the measures of general familial and societal dependency described above, we can consider a more specialized type of dependency ratio, a health dependency ratio. Such a ratio can be defined as the ratio of health-dependent persons to persons who may be expected to provide support to them. The principal idea behind such a measure is to relate persons with a health condition to the specific familial age groups or societal age groups that normally offer the support that health-dependent persons need.

We can define a few such ratios in which the dependents are represented largely by persons in the ages where disability and frailty are extremely common. For example, we can relate the number of persons 85 years and over to the number in the younger "support" group, such as persons 50–64 years of age. This is a type of familial health dependency ratio. The corresponding societal dependency ratio is the ratio of persons 85 years and over to the population 18–64 year of age. The trend of such ratios for the United States is set forth in Table 12.7.

Another type of health dependency ratio is defined by health status rather than by age. The decennial census and a number of national sample surveys provide data

Table 12.7 Illustrative health dependency ratios for the United States: 1960–2050

	Ratio[a]	Ratio[b]	Percent with a work disability[c]	
			${}_{wd}P_{16–64} \div P_{16–64}$ per 100	
Year	$P_{85+} \div P_{50–64}$ per 100	$P_{85+} \div P_{18–64}$ per 100	Year	Percent
1960	3.69	0.94	1987	8.6
1970	5.09	1.33	1992	9.3
1980	6.70	1.63	1997	10.2
1990	9.32	1.97	2002	9.9
2000	10.13	2.43	2007	10.2
2010	9.85	2.95		
2020	10.33	3.22		
2030	14.25	4.09		
2040	21.12	6.16		
2050	26.24	7.15		

Source: 1960–2000, U.S. Census Bureau decennial censuses; 2010–2050, U.S. Census Bureau population projections published in 2008 at www.census.gov. Work disability data, various *U.S. Statistical Abstracts* and www.census.gov. Primary source, Current Population Survey

[a]A type of familial health dependency ratio

[b]A type of societal health dependency ratio

[c]A type of societal health dependency ratio; work disability for the civilian noninstitutional population 16–64 years old as percent of the civilian noninstitutional population 16–64 years old. The corresponding ratios with the total resident population would be slightly higher

to compute such measures. The census, for example, provides figures on disability of various types for computing measures of health dependency ratios over several recent censuses while some national surveys provide data on work disability for the principal working ages. A few illustrative figures of the latter kind are given in Table 12.7. They show the ratio of persons 16–64 with a work disability to the number of persons 18–64 years of age.

Health as a Component in Measuring Human Capital

Health status is an element commonly included in measures of human capital along with educational status. Human capital refers to the capabilities and potentialities of humans as workers, homemakers, caregivers, and family members. Accordingly, it includes the physical and mental health of the population, allowing for disabilities and chronic illnesses, and skills and knowledge as acquired through experience and formal education, particularly the latter. Education enters into measures of human capital in the form of literacy status or school enrollment. Because the necessary data are not available for many countries, refined measures of health status, such as health expectancy, and of education, such as educational attainment, cannot be employed in constructing human capital measures for the countries of the world and less elaborate components are used.

The Human Development Index (HDI) published by the United Nations Development Program employs life expectancy at birth to represent health, and literacy and school enrollment ratios to represent education. In addition to the measures representing health and education, this index incorporates a third type of measure, the material standard of living. This third element encompasses the economic state of a population, a presumed product of the two basic elements of human capital – health and education. The HDI is viewed as a standard means of measuring human well-being, especially child welfare. It varies from 0 to 1, and it is used to distinguish the countries according to low, medium, and high development and to measure the effect of economic policies on human well-being. The 2008 Statistical Update of the HDI (UNDP 2008) calculated HDIs for all except 15 member countries of the United Nations and included data up through 2006. Iceland and Norway tied for first place and United States ranked 15th. All 22 countries in the low development category (HDI < 0.5) are in Africa.

The HDI has been subject to a number of criticisms and alternative indexes have been proposed, although the HDI is the most widely used index of human capital. The Satisfaction-with-Life Index, created by A.G. White (2007), measures subjective life satisfaction in different nations. It does not take health directly into account, but the results are highly correlated with health, wealth, and access to basic education. The Physical-Quality-of-Life Index was developed in the mid-1970s by M.D. Morris for the Overseas Development Council. It averages the literacy ratio, the infant mortality rate (indexed), and life expectancy at age 1 (indexed) equally. Lutz et al. (2008/2009) have suggested a combination of the measures of longevity, health, and education into a single human capital indicator, with a measure called literate life expectancy. They are now planning to incorporate elements of quality of education and disability status into human capital measures.

Models of Mortality, Morbidity, and Age Structure

Demographers and actuaries have long tried to make generalizations about the age pattern of human mortality and to determine its parameters, as discussed in Chap. 3. They have also used the observed mortality data for the populations with inadequate or defective mortality data in the Less Developed Countries (LDC) to construct estimates of their "true" mortality levels and patterns. Accordingly, sets of model life tables were developed as generalized paradigms for representing classes of mortality schedules in terms of levels and patterns for populations with statistically inadequate data. Sets of model stable populations were also developed to represent generalized paradigms for the age-sex distributions of populations with statistically defective data on the assumption that these populations typically had stable or unchanging age distributions, growth rates, birth rates, and death rates. Such model tables have been used mainly as substitutes for observed data on mortality and age distributions considered to be defective, but they have also proved useful as ideal-typical constructs in the analysis of observed mortality schedules that are relatively

accurate. In the last half century considerable effort has been devoted to the construction of sets of model life tables by the United Nations, academic researchers such as A.S. Coale/P. Demeny and W. Brass, and other private researchers, primarily to replace the defective mortality data in the LDC. These tools of estimation and analysis have proved extremely useful in understanding the population dynamics of the countries of the world, particularly the less developed countries.

For several reasons, this "corrective" purpose has become less important and the use of model life tables for analytic purposes in the more developed world has assumed greater importance. The quality of vital registration systems has improved and national health surveys have been implemented in many parts of the less developed world. Fertility declines have rendered application of stable population models less applicable to many less developed countries. Even in areas where the model tables had been applied effectively for many years, such as sub-Saharan Africa, other influences, such as the ravages of AIDS, widespread civil wars, massive refugee movements, and other population transfers, limit the applicability of the existing model tables.

On the other hand, in some less developed countries fertility has changed little and stable population analysis remains useful for evaluation of age distributions and rough approximations of birth and death rates. In countries where fertility decline is recent and mortality decline has not greatly modified the adult age distribution, stable population analysis remains useful for evaluating age distributions and studying population dynamics among adults. Moreover, stable population analysis is always useful for simulation of the effects of holding constant the current rates of mortality.

Other types of model tables setting forth generalized age patterns, such as model fertility tables and model marriage tables have been prepared in addition to model life tables and model population age distributions. We have a special interest in model health tables such as model chronic disease tables or model disability tables, but to my knowledge these have not been prepared. The variation in health levels and patterns may be too great to permit the generation of useful paradigms for health that could be linked to sets of model life tables and model stable population tables. Such tables may have sufficient utility in the analysis of the health of populations in many parts of the world, however, to justify a preliminary effort at their construction. Moreover, they could serve to systematize what is known from the various national health surveys and provide theoretical constructs for the analysis of morbidity in a variety of socioeconomic and developmental settings.

Construction of Model Life Tables

Empirically Based Systems

As we saw, the first modern life tables were published in the early part of the nineteenth century and by the beginning of the twentieth century life tables began

to be published regularly for many of the more developed countries. The first set of model life model tables was published by the United Nations in the 1950s. Revised sets of model life tables were published at various subsequent dates on the basis of empirical life tables constructed by different methods as additional life tables of acceptable quality for many more developed countries and some less developed countries became available.

UN model tables, 1955. The United Nations published the first set of model life tables in 1955. They were based on 158 empirical life tables and consisted of 48 tables (24 for each sex). The tables were constructed on the assumption that the level of mortality in any age group is closely correlated with the level of mortality in an adjacent age group. Specifically, parabolic (i.e., second degree) regression equations representing the relation between adjacent ${}_nq_x$ values were fitted to the data from the actual life tables selected. The value of ${}_4q_1$ was estimated with the regression equation from the level of the infant mortality rate, ${}_1q_0$, and the value of ${}_5q_5$ was estimated from the resulting estimate of ${}_4q_1$, and so on.

$$ {}_nq_x = \beta_0(x) + \beta_1(x)q_{x-5} + \beta_2(x)q_{x-5}^2 \tag{12.5} $$

In this fashion an entire schedule of ${}_nq_x$ values was determined. Hence, by specifying various infant mortality rates, an entire set of life tables could be constructed with regression equations.

The UN system of model life tables was pioneering in its conception and design but suffered from various limitations. The model tables were based on a relatively small number of national life tables relating to a relatively short period and were hardly representative of the range of human mortality experience with the age patterns shown. Their design is rather inflexible; once an infant mortality level is selected, it allows no variation in mortality level or pattern in spite of the great variation observed or possible. Moreover, the results could be seriously biased as a result of the use of a series of regression equations in chain fashion.

Ledermann system. To reduce the inflexibility of the UN model life tables and to allow for differences in mortality patterns in different populations, Ledermann (1969) devised a more flexible and variegated system of constructing model life tables. His system used a linear (i.e., first degree) regression equation with two independent variables to estimate the value of the logarithm of ${}_nq_x$. The regression equation has the following form:

$$ \ln {}_nq_x = \beta_0(x) + \beta_1(x)\ln Z_1 + \beta_2(x)\ln Z_2 \tag{12.6} $$

The two independent variables could be the pairs of mortality rates, ${}_5q_0$ and ${}_{20}q_{45}$, ${}_{15}q_0$ and ${}_{20}q_{30}$, or ${}_{15}q_0$ and m_{50+}. This system is more flexible than the UN system because there are two independent variables and there are several general choices for the independent variables. However, the parameters were developed from the same limited set of life tables as the United Nations used in 1955, and therefore the tables did not cover the range of mortality patterns experienced globally.

Coale and Demeny model tables 1966. The systems described above are not much used currently in the countries with defective or limited mortality data. They have been replaced by the Coale/Demeny system of model tables (1966) and a later set of UN tables (1982). In 1966 Coale and Demeny published sets of so-called regional model life tables, designated North, South, East, and West. Although the Coale/Demeny tables were mostly derived from the mortality experience of MDC, they were developed from a larger number of empirical life tables than the original UN tables that covered the experience of the last century and a half. From this search, 192 tables were chosen from 326 male and 326 female empirical life tables. The final set of 192 life tables used in deriving the model life tables included only 16 tables from Africa and Asia; the others came from Europe, North America, Australia, and New Zealand.

As suggested, analysis of these life tables revealed four mortality patterns, called North, South, East, and West. For example, the East pattern shows relatively high infant mortality and high old-age mortality, while the North pattern shows relatively low infant mortality, high child mortality, and low mortality after age 50. The West pattern was derived from the largest set of observed patterns (130) and represented the most general mortality pattern. It is recommended for use when reliable information is lacking to suggest the choice of another pattern.

Coale and Demeny also used regression equations to construct their system of model life tables. The regression equations linked the logarithm of ${}_nq_x$ in a linear relation with a single independent variable, e_{10}, for the four mortality patterns (separately for males and females). The regression equation had the following form:

$$ {}_nq_x = A_x * B_x^{e10} \tag{12.7a} $$

$$ \ln {}_nq_x = \ln A_x + e_{10} \ln B_x \tag{12.7b} $$

Various life table ${}_nq_x$ values were derived by inserting various values of the independent variable, e_{10}, into the equations. Both of these equations were employed in deriving the model life tables. The equation in natural numbers was used for one segment of the age distribution, the logarithmic formulation was used in a second segment, and the mean of the two for the remaining age segment. The age scale of the model tables went up to age 80. (In 1983 Coale and Demeny republished their model tables, extending them to age 100.)

Life tables for females were first constructed on the basis of a series of model values of e_{10}. Specifically, a set of life tables for females for each region, corresponding to a series of model e_{10} values ranging from 20.0 years to 77.5 years at intervals of 2.5 years, was generated. Accordingly, 24 life tables were produced and identified as levels 1, 2, 3,... 24. (The 1983 version added a level 25, with an assigned life expectation at birth of 80 years.) For the model tables for males, the e_{10}'s for males were paired with e_{10}'s for females on the basis of the male-female relation of e_{10}'s in the particular life tables in each region. Comparative analysis of the results indicates that the mortality rates can differ sharply from region to region. See, for example, the comparison of infant mortality rates and mortality at age 65 for selected model mortality levels under the four patterns in Table 12.8.

Table 12.8 Infant mortality rates (q_0) and mortality rates at age 65 (q_{65}) for females at selected model mortality levels, by regional pattern, according to the United Nations model life tables

	Mortality level			
Regional pattern	Level 11($e_0 = 45$)	Level 15 ($e_0 = 55$)	Level 19 ($e_0 = 65$)	Level 23 ($e_0 = 75$)
Infancy (q_0)				
Latin American	.1297	.0945	.0618	.0321
Chilean	.1670	.1201	.0769	.0390
South Asian	.1564	.1195	.0732	.0366
Far East Asian	.1078	.0749	.0456	.0211
General Pattern	.1253	.0898	.0571	.0280
Age 65 (q_{65})				
Latin American	.0408	.0324	.0238	.0148
Chilean	.0455	.0354	.0254	.0153
South Asian	.0436	.0341	.0245	.0145
Far East Asian	.0570	.0442	.0311	.0178
General Pattern	.0471	.0370	.0267	.0159

Source: United Nations (1982)

UN model life tables, 1982. To deal with the problem of the inflexibility of the UN model tables of 1955, and the dependence of the Coale/Demeny tables of 1966 on mostly developed countries for generating their four regional patterns, the United Nations devised a new set of model life tables in 1982 based on the experience of a group of less developed countries. From its review of the available life tables for the LDC, it found 72 life tables (36 for each sex) of acceptable quality representing 22 less developed countries. Tables for five regions were then constructed to represent world mortality patterns.

The models were named by geographic region according to the region that predominated within each group: Latin American, Chilean, South Asian, Far Eastern, and a general pattern. The mortality patterns of the regions differed on the basis of the relative levels of mortality in the principal parts of the age distribution. For example, the pattern of the Latin American region shows relatively high infant, child, and young adult mortality, and relatively low mortality at the older ages; and the pattern of the Far Eastern region shows relatively low mortality at the early ages and relatively high mortality at the older ages. This was the largest study of the mortality patterns in the LDC conducted to date and the results have been widely used by demographers. In the UN report (1982) model life tables are presented for the five "regions" for males and females at 1-year levels of life expectation at birth from level 35.0 years to 75.0 years. Table 12.9 illustrates a UN model abridged life table for females of the South Asian region with a life expectation at birth of 55 years.

The UN model life tables of 1982 were constructed by the method of principal component analysis. The age pattern of each original life table was analyzed in terms of its ${}_nq_x$ values expressed in logits of the form:

$$ {}_nY_x^{ij} = \text{logit}\,({}_nq_x) = 1/2 \ln ({}_nq_x \div 1 - {}_n q_x) \tag{12.8} $$

Table 12.9 Illustrative UN model abridged life table, with life expectation at birth of 55 years: South Asian pattern, females

Age	${}_nm_x$	${}_nq_x$	${}_nd_x$	l_x	${}_nL_x$	T_x	e_x
0	.12280	.11372	11372	100000	92608	5500000	55.000
1	.01994	.07578	6716	88628	336786	5407392	61.012
5	.00337	.01673	1370	81911	406130	5070606	61.904
10	.00144	.00720	580	80541	401255	4664476	57.914
15	.00204	.01015	812	79961	397864	4263221	53.316
20	.00247	.01228	972	79149	393371	3865378	48.836
25	.00274	.01359	1063	78177	388287	3471987	44.412
30	.00330	.01636	1261	77115	382510	3083699	39.988
35	.00399	.01974	1497	75853	375646	2701189	35.611
40	.00509	.02514	1869	74356	367302	2325543	31.276
45	.00691	.03401	2465	72487	356624	1958241	27.015
50	.01084	.05285	3701	70022	341475	1601608	22.873
55	.01720	.08266	5482	66321	318772	1260142	19.001
60	.02763	.12967	7889	60839	285496	941371	15.473
65	.04427	.19183	10157	52950	240283	655875	12.387
70	.06534	.28149	12046	42793	184352	415591	9.712
75	.09930	.39631	12185	30747	122707	231239	7.521
80	.14125	.51475	9555	18561	67644	108533	5.847
85	.22028	1.00000	9007	9007	40889	40889	4.540

Source: United Nations (1982) *Model life tables for developing countries* (Population Studies No. 77). New York: United Nations, p. 149

The model construction can be carried out with a first component or can be extended to include the second and third components. After the estimated logits are obtained from the model, the model estimated logits must be converted back to ${}_nq_x$ values as follows:

$$ {}_nq_x = 1 \div [1 + e^{-2nyx}] \tag{12.9}$$

A discussion of the application of the principal components method in the development of mortality patterns is beyond the scope of this book. An explanation and an illustration of the fitting of a pattern to the rates for a particular country are given in Siegel and Swanson (eds.)/Suchindran (2004) and U.N. (1982). Table 12.10, adapted from the Suchindran description, shows the observed ${}_nq_x$ values and the values fitted to the several U.N. mortality patterns by a one-component factor model for Tunisia females in 1995.

Limitations of model fitting. Two sets of age-specific death rates describing the mortality in two populations or the same population at different dates are likely to differ in their level and pattern. If a country is experiencing rapid change in the level of mortality, it is almost certain that the pattern of mortality is changing as well and that the mortality conditions calls for a different mortality model during the different periods. Coale and Guo (1989) noted that "after mortality had fallen to a relatively low level, there was a tendency for populations that once conformed

Table 12.10 Observed ${}_nq_x$ and ${}_nq_x$ fitted to UN model mortality patterns by a one-component factor model, for Tunisian females, 1995

		Fitted ${}_nq_x$			
Age	Observed ${}_nq_x$	Latin American	Chilean	South Asian	Far Eastern
0	.02715	.03488	.04763	.04924	.02588
1	.00657	.01201	.00720	.01697	.00528
5	.00295	.00305	.00176	.00351	.00156
10	.00220	.00164	.00144	.00156	.00126
15	.00280	.00207	.00225	.00206	.00239
20	.00310	.00288	.00321	.00245	.00332
25	.00374	.00389	.00428	.00298	.00461
30	.00479	.00515	.00568	.00392	.00591
35	.00683	.00740	.00805	.00553	.00860
40	.01055	.01053	.01192	.00850	.01308
45	.01391	.01350	.01773	.01337	.02062
50	.02031	.02231	.02609	.02317	.03223
55	.03589	.03477	.04060	.03966	.05005
60	.05112	.05577	.06382	.06835	.07717
65	.09872	.09376	.10677	.11218	.11985
70	.14614	.14835	.16180	.17900	.17950
75	.28289	.21900	.23621	.26977	.25737
80	.57791	.34571	.35291	.39516	.39164
$SS = \frac{1}{18}\sum_x({}_nq_x - \text{fitted } {}_nq_x)^2$					
	–	.00323	.00300	.00198	.00211

Source: Suchindran (2004)

Note: *SS* Sum of squared differences over all ages between the observed ${}_nq_x$ and the fitted ${}_nq_x$ under each model. Note that the SS is smallest for the South Asian model

most closely to the West, South, or East patterns to move in the most recent years to close conformity with the North pattern…" Contributing to the tendency toward shifting of regional membership is the fact that child and adult mortality may change independently (Woods and Rees 1986; Wrigley et al. 1997). Changes in mortality patterns in Australia, England, Sweden, and the United States have been examined by means of the Heligman-Pollard mortality schedule (Forfar and Smith 1987; Rogers and Gard 1991), and the mortality patterns of the populations of England, France, Sweden, and the United States have been compared over more than a century. The comparisons show that mortality patterns in these populations have gone through substantial changes and none has conformed to only a single mortality model.

Furthermore, assignment of a given regional name to a set of model tables does not mean that all the countries in the region conform to the model. Some countries in a region may have mortality patterns that conform to the model designated by another regional name. Moreover, similarity of mortality patterns may not follow regional lines because mortality patterns may not be closely related to the level of economic development, as suggested by the regional groupings. Mortality patterns

in the LDC may even resemble those in the MDC. Accordingly, it is useful to examine the socioeconomic and other determinants of mortality in order to interpret the differences in patterns between populations. Globalization and the international efforts at eliminating disease may have reduced differences in patterns of mortality characterizing regions. Hence, regions may no longer represent a good proxy for the socioeconomic and other determinants of mortality. It may now be important to examine the role of these determinants in affecting the mortality situation in particular populations in order to discern differences in mortality patterns.

Other Methods of Life Table Construction

The variations in mortality patterns around the world are so great that analysts have sought to devise even more flexible systems for constructing model life tables than the UN and Coale/Demeny designs. Even though the later UN system was based in part on life tables from the LDC, no tables from sub-Saharan Africa were included and the patterns of these countries do not fit well into the four or five patterns that the UN and Coale/Demeny recognized. To fill this need, Brass (1975) developed a so-called "relational model" system.

Brass' relational model life tables. Brass' system is a two-parameter logit system in which one parameter represents the mortality level and the other represents the mortality pattern. In this system, the model equation expresses a linear relation between life table survivors at age x in logits for the study population (i.e., the dependent variable) and life table survivors at age x in logits for a standard table (i.e., the independent variable). If $\lambda(l_x)$ is the logit transformation of $1 - l_x$ from the study population and $\lambda(l_x^s)$ is the logit transformation of $1 - l_x^s$ from the standard population, the model equation can be expressed as

$$\lambda(l_x) = \alpha + \beta\, \lambda(l_x^s) \tag{12.10}$$

As shown, the model equation expresses a linear relationship with two parameters, α and β. Once a standard life table is selected and values for the parameters, α and β, are calculated, a set of life tables can be generated.

The principal steps in constructing model life tables on the basis of a standard life table are as follows:

1. Calculate logits, $\lambda(l_x^s)$, from the standard life table:

$$\lambda(l_x^s) = 1/2 \ln [(1 - l_x^s) \div l_x^s] \tag{12.11}$$

2. Calculate logits, λ_x, from the life table for the study population:

$$\lambda(l_x) = 1/2 \ln[(1 - l_x) \div l_x] \tag{12.12}$$

3. Determine the values of α and β in Eq. 12.10 by inserting values of $\lambda(l_x^s)$ and $\lambda(l_x)$ from the standard table and the various life tables in the set and using a least squares criterion.

Table 12.11 Life table function (l_x) calculated by the Brass relational two-parameter logit-system, for Tunisian females, 1995

Age	l_x	$\lambda(l_x)$	$\lambda(l_{xs})$	Predicted l_x
0	1.0000			1.0000
1	.9728	−1.78848	−.9972	.9847
5	.9664	−1.67953	−.6514	.9705
10	.9636	−1.63805	−.5498	.9644
15	.9615	−1.60892	−.5131	.9619
20	.9588	−1.57632	−.4551	.9576
25	.9588	−1.53691	−.3829	.9515
30	.9522	−1.49587	−.3150	.9451
35	.9477	−1.44852	−.2496	.9382
40	.9412	−1.38651	−.1817	.9301
45	.9312	−1.30264	−.1073	.9203
50	.9183	−1.20974	−.0212	.9072
55	.8997	−1.09695	.0832	.8888
60	.8674	−0.93908	.2100	.8622
65	.8230	−0.76840	.3746	.8200
70	.7417	−0.52741	.5818	.7533
75	.6334	−0.27342	.8611	.6406
80	.4542	.09816	1.2433	.4602
85	.1918	.71918	1.7810	.2320

Source: Suchindran (2004)

Note: l_x = original values for Tunisian females 1995; $\lambda(l_x) = \text{logit}(1 - l_x) = 1/2 \ln[(1 - l_x) \div (l_x)]$ = logit of complement of original Tunisian values; $\lambda(l_{xs}) = \text{logit}(1 - l_{xs})$ = logit of complement of African standard l_x values

4. Compute values of $\lambda(l_x)$ with the estimated values of α and β in the model equation by inserting values of $l^s{}_x$.
5. Transform the computed values of $\lambda(l_x)$ into l(x) values using the following equation:

$$l(x) = 1 \div [1 + e^{2a+2b\lambda(lxs)}]. \qquad \text{(Same as 12.7)}$$

One can develop a set of life tables from a given standard life table by varying the parameters α and β. The standard life table may be any appropriate life table, but Brass selected an African standard and a general standard in the form of the West-model mortality pattern from the Coale/Demeny system. Suchindran illustrates how the Brass model can be fit to an observed life table (Siegel and Swanson (eds.)/Suchindran 2004). Using the Brass African pattern as a standard, he fit the model equation to the 1995 female life table of Tunisia. The model parameters estimated by the least squares criterion are $\alpha = -1.1194$ (standard error = 0.0225) and $\beta = 0.9644$ (standard error = 0.0327). The survival probabilities predicted by the model are given in Table 12.11. The predicted l_x probabilities show that the model overestimates the l_x values at the very young and very old ages and underestimates the l_x values at the intermediate ages.

Other parametric model designs. With the view that the two-parameter model does not adequately fit several life table patterns, others have developed extensions of it. Namboodiri (1991) proposed a five-parameter model, elaborating the two-parameter linear logit model of Brass. Heligman and Pollard (1980) proposed an additive (three-term) multicomponent nonlinear model with eight parameters; and Rogers and Little (1994) proposed a (five-term) multicomponent exponential model with five parameters. Note that modern computer technology has made it possible to fit complex mathematical models to mortality schedules – a solution not practical in the precomputer era.

Models of Age Structure Linked to Mortality Models

Model life tables can be combined with model growth rates or model fertility patterns to derive model stable populations. Such populations maintain the same percentage age distributions indefinitely over time. Model stable populations are used as tools to estimate and analyze growth rates, mortality and fertility levels, and age structures, to show how the population would grow under specified vital conditions, to estimate or reestimate vital rates, and to correct defective population age distributions of statistically underdeveloped countries. A particular model stable population may result from any one of a range of growth rates, model life tables, or model fertility levels and patterns, but any two of these parameters determine the others and in particular determine a particular model stable population.

Under actual conditions a stable population emerges when a particular mortality schedule and a particular growth rate or fertility schedule remain unchanged over several decades. The period must be long enough for the mortality rates and the growth rate to restructure the age composition of the initial population in their own image – say 75 years or more. There is the further assumption that the population is closed; that is, it is not subject to net immigration. As a result of these processes, the population converges to a stable form; that is, its age composition stops changing. The final age composition is independent of the initial population's composition. This property of a population's forgetting its past age structure is known as ergodicity. These relationships offer the possibility of inferring the level of mortality and fertility of a population which is shown or hypothesized to be stable and whose growth rate can be estimated. In the sections that follow I shall briefly elaborate on the relationships just described.

The type of stable population whose size as well as age distribution does not change is called a stationary population. It results when a schedule of age-specific death rates are assumed to prevail, or prevails, without change over the lifetime of a birth cohort and when the population growth rate is zero. In this case the birth rate and the death rate of the population agree with one another, and the total population and the numbers at each age do not change, as is the case with the stationary population of a life table.

The theory of the stable population linking mortality rates and growth rates with the stable age distribution was first set forth in a mathematical formula over a century ago by A. Lotka (1907). The Lotka equation for the stable population model is

$$\int_0^\infty e^{-rx} f(x)\, p(x)dx = 1 \tag{12.13}$$

where r is the rate of (natural) increase per person per year in the stable population (also called the intrinsic rate of natural increase), $f(x)$ is the number of live female births per year to each women of age x, and $p(x)$ is the probability of surviving from birth to age x. Alternatively,

$$b = \frac{1}{\int_0^\infty e^{-rx} p(x)dx} \tag{12.14a}$$

where b is the birth rate per person per year in the stable population (also called the intrinsic birth rate). The integral in Eq. 12.14a represents the total female stable population and 1.00 represents the annual number of births per person. Equation 12.14a can be restated in discrete notation for purposes of calculating the stable age distribution, its birth rate, and its death rate:

$$b = \frac{1}{\sum e^{-r(x+1/2)} * (L_x \div l_0)} \tag{12.14b}$$

Note that $p(x)$ in the continuous formulation is $(L_x \div l_0)$ of the life table in the discrete formulation. Applying Eq. 12.14b calls for prior knowledge of r and the L_x's (i.e., the rate of natural increase and the mortality schedule, respectively).

Unchanging mortality and fertility schedules in many less developed countries over many decades have led to stable, or nearly stable, populations in these countries. Inasmuch as most of these countries had deficient demographic statistics, they often only appeared to have the characteristics of a stable population. Under these conditions, the Lotka formula has proved to be an effective tool in establishing the demographic parameters of these populations. To apply it, we must have either approximate estimates of the mortality level and the growth rate to establish the "true" age distribution, or an estimate of the age pattern of the population and the growth rate in order to establish the levels of mortality and fertility.

With approximate values of two or more parameters of a stable population, for example the growth rate and age distribution, and by referring to the sets of model stable populations, model life tables, and model fertility rates, it is possible to determine the other parameters, for example, life expectancy and the death rate. The model tables make it convenient to determine the mortality and fertility characteristics as well as the age-sex structure of stable populations with limited and defective data. It is only necessary to refer to the model tables with the data on the parameters available and read off the remaining data desired. Instructions for applying stable population theory and equations to the determination of the basic demographic parameters of a population are set forth in Chap. 22 of Siegel and

Table 12.12 Illustrative U.N. model stable populations for South Asian region females, with life expectation at birth of 55 years, at selected growth rates (Proportion of population in indicated age group)

	Growth rate				
Age group	0.0	0.010	0.020	0.030	0.040
0	.01719	.02344	.03063	.03858	.04709
1–4	.06575	.08741	.11142	.13688	.16295
5–9	.08073	.10260	.12503	.14683	.16711
10–14	.07996	.09667	.11205	.12518	.13551
15–19	.07901	.09086	.10019	.10647	.10963
20–24	.07739	.08466	.08879	.08975	.08792
25–29	.07532	.07838	.07820	.07519	.07006
30–34	.07294	.07220	.06852	.06267	.05555
35–39	.07024	.06614	.05971	.05195	.04380
40–44	.06713	.06012	.05163	.04273	.03427
45–49	.06340	.05401	.04412	.03473	.02650
50–54	.05868	.04755	.03695	.02767	.02008
55–59	.05265	.04058	.03000	.02137	.01475
60–64	.04517	.03312	.02328	.01578	.01036
65–69	.03636	.02536	.01696	.01093	.00683
70–74	.02670	.01772	.01127	.00691	.00411
75–79	.01718	.01085	.00656	.00383	.00216
80–84	.00913	.00548	.00315	.00175	.00094
85+	.00507	.00285	.00154	.00080	.00040
Birth rate	.01818	.02490	.03268	.04133	.05065
Death rate	.01818	.01490	.01268	.01133	.01065

Source: United Nations (1982). Stable populations corresponding to the new United Nations model life tables for developing countries (Series R, No. 44). New York: United Nations, 1982

Swanson (Eds.)/Popoff and Judson 2004. Model stable populations are illustrated for the South Asia region females, at various growth rates, for mortality level $e_0 = 55$ in Table 12.12.

Relationships Between Mortality and Age Structure of Western Populations

For populations whose age distributions have been changing, the various measures of the stable population describe what would happen if the fertility and mortality schedules of a given period continued unchanged sufficiently long under closed (i.e., no immigration) conditions. These measures do not represent a description of what is or has been happening, or forecasts of what will eventually happen to this population. In those less developed countries where the conditions of unchanging levels of fertility and mortality and no immigration appear to have applied, stable

population theory has provided us with improved estimates of the age distribution of these populations and their vital rates. However, inasmuch as fertility and mortality levels have been changing in many of them and some have been greatly affected by disasters of various kinds, such as civil wars, refugee movements, famine, and the AIDS epidemic, stable population theory has become less applicable to them. At the same time, many Western populations have been experiencing such persistently low levels of fertility that they have attained an approximately stationary condition, that is, a stable condition with a zero population growth. This has raised interest in the application of classical stable population analysis to the study of the age structure of the MDC and their vital rates.

Demographers have long been interested in the relationship between changes in the components of change of populations (i.e., fertility, mortality, and migration) and changes in their age-sex structure. Empirical historical analysis has relied on sensitivity analysis and simulation to isolate the contribution of each component to changes in age structure. In this type of analysis one component is "held constant" at a time while the actual historical values of the other components are applied.

As we saw in Chap. 3, the effect of mortality improvement on the growth rate varies with the initial level of mortality as well as the magnitude of the change. A change in expectation of life at birth from 30 to 40 years, for example, will produce a higher growth rate than a change from 60 to 70 years, for a given level of fertility. Applying simulation analysis to the U.S. data for the first six decades of the twentieth century, Hermalin (1966) concluded that:

- Immigration has had only a small effect on the age composition of the U.S. population and this effect was to make the population younger.
- Fertility not only determines the general shape of the age distribution but, in the period considered, changes in fertility were the dominant influence in the age composition of the population.
- Fertility declined greatly during this period and, as a result, there has been a marked aging of the population.
- Declining mortality has led to a somewhat younger population, although alternative measures give somewhat different indications as to the degree to which declining mortality has resulted in a younging of the population.

Coale (1956) had shown earlier that the reason that declining mortality had so little effect on age structure was that the age pattern of improvements in mortality was J-shaped or U-shaped. In other words, the improvements were spread out over the age distribution and therefore had little effect on it. The reduction of mortality has had a considerable effect on population growth, however. It has been estimated that the population of the United States in 1960 was about one-third larger as a result of improvements in mortality since 1900. Further improvements in mortality after 1960 have had a smaller effect on population growth; fertility and migration have been the chief determinants of post-1960 growth. The effect of mortality on the aging of the population, however, appears to have exceeded that of fertility in this later period (Preston et al. 1989).

Reflecting analytically on the role of mortality in relation to age structure, Hermalin (1966) and others have noted that (1) the measure of mortality that directly determines changes in age composition is the relative change in survival rates, not the relative change in mortality rates, and (2) a uniform percentage increase in survival rates during a period would leave the age distribution of the population unaffected. It would reduce the crude death rate and increase the growth rate, however.

Models of Morbidity and Disability

Much model-building research was carried out before the intensive concern of demographers with health and the interrelations of health and mortality. Moreover, the model-builders were interested in developing tools for estimating the basic population parameters for areas lacking adequate data on population age structure, fertility, and mortality. By the current date many of the statistically less developed countries have compiled adequate data on age-structure, mortality, and fertility, but their data on morbidity and disability are still quite limited. The Demographic and Health Surveys (DHS) have filled some of the gaps, but basic data are still missing for many countries and there is a serious question about much of the health data collected in the DHS. Moreover, the DHS have been principally concerned with child and maternal health. Accordingly, the construction of sets of model health tables designed to describe various levels and patterns of age-specific chronic disease and disability under various mortality conditions would be useful. Standard age-patterns of disability ratios could be developed from actual data for those MDC and LDC for which usable health data are available. Alternatively, the Brass relational system could be used to derive estimated and improved values based on the available sets of age-specific disability or chronic illness ratios. The age-specific disability data for each age group for the countries of interest could be paired with the age-specific disability ratios of the standard model table in the form of logits, and linear regression equations could then be evaluated to determine the parameters of this relationship. Then, new age-specific disability ratios and morbidity ratios for the countries being studied could be estimated.

Appendix 12.1: MORTPAK

The United Nations Population Division has developed MORTPAK, for Windows, Software Package for Demographic Measurement, revised in 2003, with special emphasis on mortality measurement. It was originally published as MortPak: The UN Software Package for Mortality Measurement, in 1988. The package is no longer available in printed form but is available on CD-ROM for use with Windows in Adobe PDF format. MORTPAK includes 17 applications covering the following areas:

Population projections
Construction of life tables and stable populations
Graduation of mortality data
Estimation of mortality levels and patterns by indirect methods
Estimation of fertility levels and patterns by indirect methods
Evaluation and adjustment of census counts and age distributions
Other applications focus mainly on mortality measurement, including the measurement of the completeness of adult death registration and estimation of infant and child mortality.

This package takes advantage of the United Nations model life tables and stable population models. These model tables were published in *Stable Populations Corresponding to the New United Nations Model Life Tables for Developing Countries, Model Life Tables for Developing Countries*, and *Unabridged Model Life Tables Corresponding to the New United Nations Model Life Tables for Developing Countries*, by the Population Division of the United Nations in 1981 and 1982. Indirect techniques of fertility and mortality measurement are described in detail in Manual X, *Indirect Techniques for Demographic Estimation*, published in 1983 by the UN Population Division.

MORTPAK for Windows requires a personal computer with Windows 95 operating system or above and Internet Explorer 4.0 or above. See www.unpopulation.org for further details.

This package has been widely used in research institutions of both the more and less developed countries since its release in 1988.

References and Suggested Readings

Sources

Lutz, W. (2008/2009). Measuring human capital. *Popnet: Population Network Newsletter*, (International Institute for Applied Systems Analysis), *40*, 2.

United Nations. (1993). *Demographic yearbook: Special issue on population ageing and the situation of elderly persons*. New York: United Nations.

United Nations Human Development Program. (2008). Human development indices: A statistical update 2008. *Statistics of the Human Development Report*. Available at: http://hdr.undp.org/en/statistics/

U.S. National Center for Health Statistics. (1974). *Vital statistics of the United States, 1970: Vol. II, Mortality, Part A*. Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (1998). Current estimates from the National Health Interview Survey, 1995. By V. Benson & M. A. Marano. *Vital and Health Statistics, 10*(199).

U.S. National Center for Health Statistics. (2007a). Summary health statistics for the U.S. population: National Health Interview Survey, 2005. By P. F. Adams, A. N. Dey, & J. L Vickerie. *Vital and Health Statistics, 10*(233).

U.S. National Center for Health Statistics. (2007b). National hospital discharge survey: 2005 Annual summary with detailed diagnosis and procedure data, By C. J. De Frances, K. A. Cullen, & L. J. Kozak. *Vital and Health Statistics, 13*(165).

U.S. National Center for Health Statistics. (2008). Deaths: Final data for 2005. By H-C. Kung, D. L. Hoyert, J. Xu, & S. L. Murphy. *National Vital Statistics Reports, 56*(10).

U.S. Social Security Administration, Actuary's Office. (2005). Short-range actuarial projections of the old-age, survivors, and disability insurance program, 2005. By C. Motsiopoulos & R. R. Tucker. *Actuarial Study, 119*.

White, A. G. (2007). A global projection of subjective well-being: A challenge to positive psychology? *Psychtalk, 56*, 17–20.

Measurement of, Trends and Factors in, and Consequences of Population Aging

Committee for International Cooperation in National Research in Demography (CICRED). (2005). *Policy implications of age-structural changes* (CICRED Policy Paper No. 1). Paris: CICRED.

Davis, K., Bernstam, M. S., & Ricardo-Campbell, R. (Eds.). (1987). *Below-replacement fertility in industrial societies: Causes, consequences, policies*. NewYork: Cambridge University Press/Population Council.

Foster, C. (2000). The limits to low fertility: A biosocial approach. *Population and Development Review, 26*, 209–234.

Goldman, N. (1986). Effect of mortality levels on kinship. In United Nations (Ed.), *Consequences of mortality trends and differentials* (Population Studies No. 95, pp. 79–87). New York: United Nations.

Hermalin, A. I. (1966). The effect of changes in mortality rates on population growth and age distribution in the United States [Part I]. *Milbank Memorial Fund Quarterly 44*(4), 451–469.

Lutz, W., & Scherbov, S. (2003). *Will population ageing necessarily lead to an increase in the number of persons with disabilities? Alternative scenarios for the European Union* (European Demographic Research Papers. No. 3). Vienna, Austria: Vienna Institute of Demography.

Lutz, W., Sanderson, W., & Scherbov, S. (2001). The end of world population growth. *Nature, 412*, 543–545.

Ogawa, N. (1986). Consequences of mortality change on aging. In United Nations (Ed.), *Consequences of mortality trends and differentials* (Population Studies No. 95) (pp. 175–184). New York: United Nations.

Okolski, M. (1986). Relationships between mortality and morbidity levels according to age and sex and their implications for organizing health care systems in developed countries. In United Nations (Ed.), *Consequences of mortality trends and differentials* (Population Studies No. 95, pp. 150–164). New York: United Nations.

Preston, S. H., Himes, C., & Eggers, M. (1989). Demographic conditions responsible for population aging. *Demography, 26*, 696–704.

Sanderson, W., & Scherbov, S. (2008). Rethinking age and aging. *Population Bulletin, 63*(4). Washington, DC: Population Reference Bureau.

Siegel, J. S. (1993). *A generation of change: A profile of America's older population*. New York: Russell Sage Foundation.

Sonnenschein, E., & Brody, J. A. (2005). Effect of population aging on proportionate mortality from heart disease and cancer, U.S. 2000–2050. *Journals of Gerontology: Social Sciences, 60B*(2), S110–S112.

United Nations. (1986). *Consequences of mortality trends and differentials* (Population Studies No. 95). New York: United Nations.

Young, C. M. (1986). The residential life cycle: Mortality and morbidity effects on living arrangements. In United Nations (Ed.), *Consequences of mortality trends and differentials*. (Population Studies No. 95, pp. 113–125). New York: United Nations.

Models of Mortality and Age Structure; Model Life Tables and Model Stable Populations

Brass, W. (1975). *Methods for estimating fertility and mortality from limited and defective data*. Chapel Hill, NC: University of North Carolina.

Cambois, E., Robine, J.-M., & Brouard, N. (1999). Life expectancies applied to specific statuses: A history of the indicators and the methods of calculation. *Population*, 3, 1998. *Population: An English Selection* (Vol. 11). Paris: INED.

Coale, A. J., & Demeny, P. (1966). *Regional model life tables and stable populations*. Princeton, NJ: Princeton University Press.

Coale, A. & Guo, G. (1989). Revised regional model life tables at very low levels of mortality. *Population Index, 55*, 613–643.

Coale, A. J., & Demeny, P., with Vaughn, B. (1983). *Regional model life tables and stable populations* (2nd ed.). New York: Academic.

Forfar, D. O & Smith, D. M. (1987). The changing shape of English life tables. *Transactions of the Faculty of Actuaries, 40*, 98–134.

Goldman, N. (1980). Far Eastern patterns of mortality. *Population Studies, 34*, 5–19.

Graggnolati, M., Elo, I. T., & Goldman, N. (1999). New insights into the Far Eastern pattern of mortality. *Population Studies, 53*, 81–95.

Heligman, L., & Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries, 10*, 49–80.

Ledermann, S. (1969). Nouvelles tables-type de mortalité. *Travaux et documents* (Cahier no. 53). Paris: INED.

Lotka, A. J. (1907). Relation between birth rates and death rates. *Science, 26*, 21–22.

Manton, K. G., Stallard, E., & Corder, L. (1997). Changes in the age dependence of mortality and disability: Cohort and other determinants. *Demography, 34*(1), 135–157.

Namboodiri, K. (1991). *Demographic analysis: A stochastic approach*. Chapter 7. San Diego, CA: Academic.

Olshansky, S. J., Carnes, B. A. (1997). Ever since Gompertz. *Demography, 34*(1): 1–15.

Popoff, C., & Judson, D. H. (2004). Some methods of estimation for statistically underdeveloped areas (Chap. 22). In J. S. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed.). San Diego, CA: Elsevier/Academic.

Rogers, A., & Gard, K. (1991). Applications of the Heligman/Pollard mortality schedule. *Population Bulletin of the United Nations, 30*, 79–105.

Rogers, A., & Little, J. S. (1994). Parameterizing age patterns of demographic rates with the multiexponential model schedules. *Mathematical Population Studies, 4*, 175–195.

Suchindran, C. M. (2004). Model life tables. Part II in Appendix B, In J. S. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed.). San Diego, CA: Elsevier/Academic.

United Nations. (1955). *Age and sex patterns of mortality: Model life tables for underdeveloped countries*. New York: United Nations.

United Nations. (1982). *Model life tables for developing countries*. New York: United Nations.

United Nations. (1988). *MortPak: The United Nations software package for mortality measurement*. New York: United Nations.

Williamson, P., & Woods, R. I. (2003). A note on the fetal-infant mortality problem. *Journal of Biosocial Science, 35*(2), 201–212.

Woods, R., & Rees, P. (Eds.). (1986). *Population structures and models: Developments in spatial demography*. London: Allen & Unwin.

Wrigley, E. A., Davies, R. S., Oeppen, J. E. & Schofield, R. S. (1997). *English population history from family reconstruction, 1580–1837*. Cambridge, UK: Cambridge University Press.

Part IV
Theories of Aging and Longevity and Projections of Mortality and Health

Chapter 13
Concepts and Theories of Longevity

Concepts of Longevity

This chapter is concerned with the concepts, statistics, theories, and future levels of human longevity, particularly extreme longevity. The term, extreme longevity or superlongevity, is used here to refer to centenarians, that is, persons with a verified age of 100 years or more. Among these we identify for special attention a subgroup known as supercentenarians, that is, persons with a verified age of 110 years or more. In both popular and scientific reporting, people aged 85 years and over have often been considered as extreme aged, but the rapid increase in the number of people in these older age groups suggests a redefinition of the notion of extreme aged. Persons in the broad age group 85 years and over are also referred to as persons of advanced age and the oldest-old. The former description now seems appropriate but the latter one does not. Preferences vary among demographers and gerontologists as to the choice of designations for these groups.

Until recently, the age of 100 has been used by demographers, epidemiologists, and others to define the life span of humans. This was considered to be the age to which a substantial number of humans as a species could survive under optimum health and living conditions, but beyond which only a negligible number of persons could be expected to live. The number defining the life span of humans was not based on any empirical investigation but on general observation, the indications of current life tables, and more importantly, conventional wisdom. Although human life span is a theoretical concept and measure for which no precise value can be given, it has been useful to have such a concept for tracking the gap between the longevity achieved by the average person in a given population and a hypothetical maximum longevity achievable by humans under the most favorable conditions. It also serves a useful purpose in comparative analysis of longevity with other species, to which similar measures have been assigned. In the later discussion, I elaborate on the concept of human life span, seeking to give it a more precise meaning.

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_13, © Springer Science+Business Media B.V. 2012

Individual Measures

Age 120 has often been cited as the maximum life span for humans. This number does not originate in science, but rather in a passage that appears in the Old Testament, "My spirit will not contend in man forever, for he is mortal; his days will be a hundred and twenty years" (*Genesis* 6:3). The record for the maximum recorded human life span, or the highest verified age of a member of the human species, living or deceased, is held by Jeanne Calment of France, who died in 1997 at the age of 122 years, 164 days (Exhibit 13.1). There is no validated record of anyone ever surviving to this age or beyond except Mme. Calment. Discounting the case of S. Izumi (who is reputed to have lived 120 years, 237 days) as a case of incorrect documentation, Sarah Knauss appears next in rank for having lived to the age of 119 years, 97 days. Disregarding K. Hongo also (who had an alleged final age of 116 years, 45 days, but whose records were inconsistent), C. Mortenson (115 years, 252 days) is the oldest male ever to have lived. Survival of many persons even to age 100 is conjectural inasmuch as only about one percent of any large population survives to this age.

I assign the term maximum observed life span to the highest verified age for a living person. The identity of this person changes from year to year because of the high risk of death for people who have reached such advanced ages. As of July 1, 2008, the position is held by Edna Parker of USA at age 115 years, 1 month. The oldest living male is T. Tanabe of Japan, who is 112 years, 7 months, old.

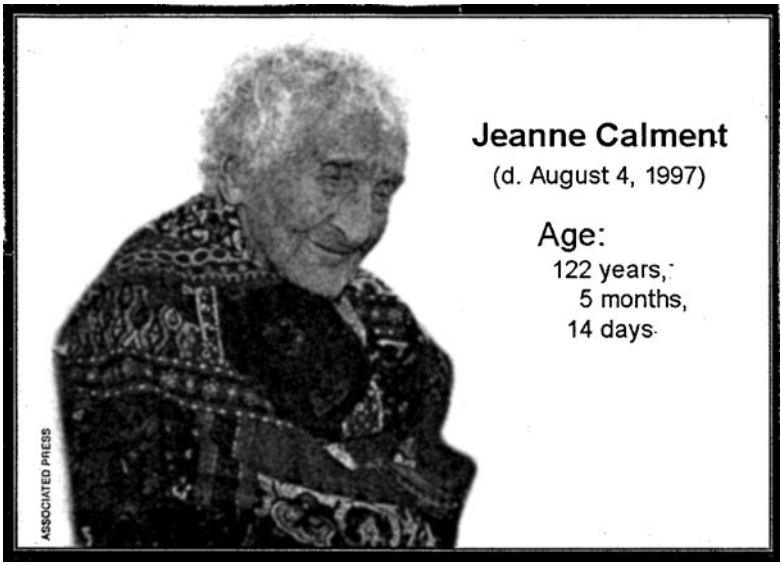


Exhibit 13.1 Portrait of Jeanne Calment, oldest person who ever lived, shown at age 122 years, 5 months, 14 days, just before her death in August 1997

The maximum age at death in a population in a given year is the age of the oldest decedent in that year. For 2007 the oldest authenticated decedent was E. del Toro of Puerto Rico, who died at age 115 years, 156 days. According to the Gerontology Research Group, as of spring 2008 there were an authenticated 77 living supercentenarians (66 females, 11 males). It is quite possible that there are many supercentenarians in the parts of the world other than in the Western countries, especially in the most populous countries such as India and China, but the records in these countries are inadequate for evaluating and authenticating any claims of superlongevity.

Population Measures

The measures of longevity given above relate to individuals. Several measures of longevity refer to population groups. Most simply, there are the absolute numbers of aged persons and the percent changes in these numbers. Table 13.1 shows the number of persons 85 years and over in the United States, distributed by age groups for three recent censuses and 2050 and percent changes in the intervening periods. Next, there is a group of measures showing different percentages of broader totals for the oldest populations. Table 13.2 shows, for example, the percent of the total population that is 85, 95, or 100 and over, the percent 100 and over of those 85 and over, and the percent 85 and over of those 65 and over. The table gives historical and projected series of such measures for 1900–2050, for the United States.

Several other measures focus on the population 100 or more and are based on either the actual population or life tables. One is the average or mean age of all persons 100 years or more. I designate this measure the average maximum life span – survivors; it represents the average age of persons over the age of 100 who are alive on a designated date. Depending on the interest of the analyst, either the data for the actual population or the data in the current life table may be employed to compute it. The corresponding measure for decedents 100 or more in a year is

Table 13.1 U.S. population 85 years and over, by age: 1980–2050

	Population (000)				Percent increase		
Age	1980	1990	2000	2050[a]	1980–1990	1990–2000	2000–2050[b]
Total, 85 and over	2240	3049	4240	20,861	36.1	39.1	392.0
85–89	1532	2047	2790	10,253	33.6	36.3	267.5
90–94	562	760	1113	6,473	35.2	46.4	481.6
95–99	132	205	287	2,984	55.3	40.0	939.7
100 and over	14[c]	37[d]	50[d]	1,150	164.3	35.1	2200.0

Source: U.S. Census Bureau: Internet, www.census.gov;decennial census data
[a]U.S. Census Bureau middle series projections, based on 2000 census
[b]Average decennial increases are 32 percent for ages 85 years and over, and 26, 35, 47, and 63 percent for the component age groups
[c]Records of the Social Security Administration; the census count, 32,000, was reestimated
[d]Reestimated as 22,000 for 1990 and 33,000 for 2000 by Kestenbaum and Ferguson (2005)

Table 13.2 Various measures of extreme population aging based on population data, 1900–2005, and projections, 2010–2050, by decades

Year	Percent 85+ of total pop.	Percent 85+ of 65+	Percent 100+ of total pop.	Percent 100+ of 85+
1900	0.2	4.0	NA	NA
1910	0.2	4.2		
1920	0.2	4.3	|	|
1930	0.2	4.1		
1940	0.3	4.0	|	|
1950	0.4	4.7	|	|
1960	0.5	5.6		
1970	0.7	7.5	|	|
1980	1.0	8.8	0.01	1.0
1990	1.2	9.9	0.01	1.2
2000	1.5	12.1	0.02	1.2
2005	1.7	13.8	0.02	1.3
Projections[a]				
2010	2.0	15.2	0.04	1.9
2020	2.2	13.3	0.07	3.3
2030	2.6	13.4	0.11	4.2
2040	3.9	19.2	0.15	3.8
2050	5.0	24.1	0.27	5.5

Source: U.S. Census Bureau: Internet, www.census.gov; various Census Bureau reports
NA reliable figures not available for years from 1900 to 1970
[a]U.S. Census Bureau middle series projections, based on 2000 census

called the average maximum life span – decedents; it represents the average age at death of those who die at age 100 or beyond in a year. These concepts of life span and their changing values clearly suggest that (maximum) life span is not a fixed number, as has so often been assumed in the past, but may change. Increases in average maximum life span can provide evidence as to whether the members of a given population are tending to live longer or not. Another group measure of extreme longevity, called life endurancy, is a life-table measure only. Life endurancy refers to the age to which 0.005%, 0.001%, or some other specified percent of the initial birth cohort in a life table survives. The shifts in this age for a given level of survivorship from birth may provide evidence that members of a population are living to increasingly advanced ages. Data for these and other measures of extreme longevity are shown in Table 13.3; they are based on the life tables for 1950, 1980, and 2000 prepared by the Office of the Chief Actuary of the Social Security Administration.

The various measures of life span discussed so far are to be distinguished from life expectancy at birth, which, as we have seen, represents the average years of life remaining for a child born in a given year for a particular population, as indicated by a life table. There is currently a gap of roughly 25 years between life expectancy at birth in the United States and average maximum life span (77 years vs. about 102 in the year 2000). This difference suggests how much progress in mortality reduction must be made at this time for the average person in the United States to reach the

Table 13.3 Selected measures of superlongevity of males and females based on period life tables for the United States Social Security Area: 1950, 1980, and 2000

	Males		
Measure	1950	1980	2000
Average maximum life span–survivors[a]	101.7	101.9	101.8
Average maximum life span–decedents[a]	101.9	102.2	101.5
Percent survivors 100 years and over	0.14	0.38	0.46
Percent survivors 100 and over of survivors 85 and over	0.5	0.9	0.7
Life expectation at age 100	1.92	2.20	1.98
Total life expectation at age 100[b]	101.9	102.2	102.0
Life endurancy (Pr = 0.01)[c]	95.1	97.4	98.2
Life endurancy (Pr = 0.001)[c]	100.7	103.0	103.0
	Females		
Measure	1950	1980	2000
Average maximum life span–survivors[a]	102.1	102.1	102.0
Average maximum life span–decedents[a]	101.6	102.5	102.2
Percent survivors 100 years and over	0.33	1.72	1.74
Percent survivors 100 and over of survivors 85 and over	0.6	1.8	1.5
Life expectation at age 100	1.92	2.42	2.26
Total life expectation at age 100[b]	101.9	102.4	102.3
Life endurancy (Pr = 0.01)[c]	97.5	101.5	101.4
Life endurancy (Pr = 0.001)[c]	102.4	106.4	105.9

Source: Based on U.S. Social Security Administration (2005, August)
[a]Mean age of life table survivors or decedents 100 years of age and over
[b]Life expectation at age 100 plus 100
[c]Age to which the proportion shown of the original birth cohort survives

average age of very long-lived persons. Assuming uniform reductions in death rates at every age, age-specific death rates for the United States in the year 2000 would have to decline by over 99% to achieve a life expectancy at birth of 100 for the population of the United States.

Sources of and Quality of Data

Data on the numbers, age distribution, and survival probabilities of the extreme aged may be obtained from censuses and various administrative records (e.g., the files of Social Security beneficiaries and Medicare enrollments). They may also be obtained from baptismal and genealogical records, and estimated from vital statistics or a combination of vital statistics and population data. For studying the issues of extended longevity, use can be made of aggregate data as well as microdata (i.e., individual-level data) drawn from the general sources just cited. Particularly valuable are historical data compiled in mortality and population

databases. Among the databases that are especially useful for studies of longevity are the Human Mortality Database (Univ. of Calif., Berkeley), the Utah Population Database (University of Utah), the Umeå Demographic Database (Umeå University, Sweden), the Iceland Demographic Database (deCODE genetics), the International Database on Longevity (Montpelier University and Max Planck Institute), and the Kannisto-Thatcher database on Old-Age Mortality (Odense University). For example, the Iceland database links genealogical data to medical information for the country's entire population (Amundadottir et al. 2004) and the Utah database provides genealogical, genetic, epidemiological, and demographic information for over 6.5 million individuals in the United States.

In general, the U.S. data on the extreme aged, both population and death statistics, have tended to be of too uncertain quality for researchers to be able to determine definitively what the true numbers at the advanced ages are. Often the people at these ages are unable to respond to census takers, so others, such as family members, neighbors, or nursing-home staff, report on their behalf with regard to their residence, age, and other demographic characteristics. Population coverage at particular ages in the census may be grossly misstated and ages are often misreported. In one classic example of age misstatement, analysts at the Social Security Administration determined that a significant number of people who were listed as centenarians in the 1990 Census were, in fact, under the age of 10 (Kestenbaum 1992). The reason for the error was that the wrong century was recorded as the year of birth. Census Bureau analysts discovered a similar problem in the 1970 census.

The data on deaths at the oldest ages from the vital registration system may also suffer from serious errors of age reporting. For example, the tabulations of deaths over age 100 for the United States have typically included some deaths at ages over 120. No validation of these ages has been conducted and it may be assumed that the persons have been assigned erroneous ages.

Because of these errors in population and death statistics, death rates at the highest ages, based on official collection systems and published in official publications, are often inaccurate and should be disregarded when the numbers of persons or deaths, and the resulting death rates, appear too erratic. Accordingly, the official death rates at advanced ages have at times been adjusted before use in life tables, such as by mathematical extrapolation of the rates recorded at the earlier ages or by borrowing rates from alternative sources.

Numerous devices have been employed to improve the quality of census population data. First, it is important to edit the microdata for obvious errors of reporting or processing. This procedure may eliminate only some of the erroneous entries but it is used largely to adjust the data only for gross errors. Sometimes more drastic adjustments or modifications are required. In one unusual case, the census of Taiwan in 2000, the elderly were reenumerated in face-to-face interviews after the initial general census had been conducted by mail-back methods (Yue 2005). Alternatively, the census data may be replaced by estimates when the original data are thought to be unreliable. One method of doing this is to extend census data from younger ages forward to the next census date by methods of demographic analysis.

For example, the population 80 years and over may be "survived" 10 years with the use of current survival rates in order to reestimate the population 90 years and over:

$$ {}_{as}P_{90+} = {}_{as}P_{80+} * {}_{as}S_{a,a+5}{}^{a+10,a+15} \tag{13.1} $$

where *as* refers to age/sex detail and $S_{a,a+5}{}^{a+10,a+15}$ is a survival rate for a 5-year age group for a 10-year time period. This is a simple application of what is known as the cohort-survival method, involving the application of survival rates to the initial population distributed by age groups. Substitutes for census data may be employed for the base population, as by "projecting" Medicare enrollments for the population 90 years and over at one date forward for 10 years by survival rates to derive the population 100 years and over 10 years later (Kestenbaum and Ferguson 2005).

Another device that can provide alternative estimates of population at the extreme ages is "la méthode de générations extinctes," the method of population reconstruction by reverse cumulation of cohort deaths (Vincent 1951). In this method, for example, the population 90 years and over at a given past date is estimated by combining the deaths that have occurred or will occur to this cohort from the date of estimate forward until its complete extinction by death, that is, by adding, approximately, deaths 90 and over in year 1, deaths 91 and over in year 2, and so on.[1]

$$ {}^{y}P_{90+} = \sum ({}^{y}D_{90+} + {}^{y+1}D_{91+} + {}^{y+2}D_{92+} + \cdots\cdots + {}^{y+15}D_{105+} + \cdots) \tag{13.2} $$

The method assumes that ages of deaths are more accurately registered than the ages of the population are recorded and that reasonably good "guesses" can be made of the number of future deaths for the most extreme ages. A time lag is required for enough deaths to be registered after the reference date to complete the population estimate. This method is not useful for preparing a reliable current estimate of the population 95 or 100 years and over.

The age distribution of the older population for some countries, if considered to be reliably enumerated, may serve as model distributions for other countries whose data are not as accurate (Coale and Kisker 1990). For example, data on the age distribution of people at the most advanced ages for the countries of northern and western Europe are generally recognized as of superior quality. These data may serve as models of the distribution of the population at the extreme ages for other countries of Europe and the United States.

Another source of information on the identity of long-lived individuals in the United States is computerized online genealogies (Gavrilova and Gavrilov 2005). The names of the persons can be checked against early census data and the Social Security Administration's (SSA) Death Master File. Such match studies by the Gavrilovs demonstrated the value of computerized genealogies as a basis for developing a family-linked database on extreme human longevity.

[1] In the actual computations, care must be taken to combine deaths that fall in the precise cohort for which the population is being estimated.

To derive mortality rates at the extreme ages, mathematical models may be fit to the mortality rates for the younger ages of the same cohort and then extended, but this procedure incurs the risk of failing to allow for any inflections in the series. Data on deaths from the SSA Death Master File may be substituted for the vital statistics from the registration system because of the presumed superior quality of the former. The mortality data for the extreme ages may also be taken from administrative records such as Medicare records, which contain both the numerator and denominator for computing such rates and are more thoroughly evaluated and edited than census data or vital registration data. Alternatively, the year-to-year changes in the death rates from the Medicare enrollments at the advanced ages (e.g., 85 years and over) may be linked to the death rates from the vital registration system at a younger age (e.g., 84 years). (See U.S. NCHS 2002a.)

The considerable uncertainty surrounding the reported figures on the number and ages of alleged supercentenarians and the timing of their deaths is changing, at least for some countries. Increasingly, methods of verification are being applied to the reports of supercentenarians and deaths at the very extreme ages, and a validated file of the extreme aged is being developed for several countries of northern and western Europe, Japan, the United States, and Canada. In recent decades two international collaborative efforts have been undertaken to assemble international databases on supercentenarians. One source of data on this special group is the International Database on Longevity, maintained by Montpelier University and the Max Planck Institute (Robine and Vaupel 2002), and the other is the Gerontology Research Group, which is located at the University of California at Los Angeles. Records in more than 10 countries of alleged supercentenarians are being collected from national statistical offices, health departments, other government agencies, and private sources, and validated by teams of researchers. They apply a variety of rigorous tests of assessing age, and maintain a running tally of all verified supercentenarians currently alive and of all previously authenticated supercentenarians. At the end of 2007 more than 500 validated records of supercentenarians have been assembled by these groups. We can now identify with some confidence the supercentenarians in these countries over the last several decades.

Because of the valuable information that living centenarians may provide on the factors contributing to longevity, several studies are being conducted that involve subject populations consisting wholly of centenarians. They are being interviewed periodically and, in some cases, physically examined. Among these centenarian studies are the Boston study (Perls et al. 1999), the Sardinian study (Poulain et al. 2004), the Okinawan study (Willcox et al. 2008), and the Georgia study (Poon 2008). Two of these are taking place in areas where there are known concentrations of centenarians – Sardinia and Okinawa. The latter are so-called founder populations; these are geographically isolated and hence are characterized by considerable endogamy and no immigration. The reader will note that I have not included the once-celebrated, but now notorious and debunked, centers of mythical superlongevity in Ecuador (Vilcabambans), Pakistan (Hunzas), and Georgia (Abkhazians), where the reports of superlongevity were grossly exaggerated (Leaf 1982).

Analysis of Changes at Very Advanced Ages

The number of persons reaching age 100 in any year may be viewed as the survivors of the births occurring 100 years earlier, augmented by immigrants and reduced by emigrants over this period with ages corresponding to the same cohort as the births.

$$ {}^{y}l_{100} = ({}^{y-100}B * S_0^{100}) + (I_x - E_x) * S_x^{100} \tag{13.3} $$

where S represents a generation survival rate either from birth to age 100 (S_0^{100}) or from the median age of net arrival of immigrants to age 100 (S_x^{100}). For simplicity in this theoretical model, we are assuming that the level of mortality of the births and the migrants are the same, differing only for the length of the survival period and the ages of exposure, and that "net immigrants" are exposed to the risk of dying for the average period of their residence in the country. For countries not affected by immigration or, for the case disregarding immigration, the number of persons surviving to 100 is simply the product of the number of births 100 years earlier and a cumulative survival rate from birth to age 100:

$$ {}^{y}l_{100} = {}^{y-100}B * S_0^{100} \tag{13.4} $$

The number of births and survival rates change continuously because of changing population size, age-sex composition, fertility and mortality levels, and migration. The annual changes in the current numbers of 100-year-olds reflect past changes in numbers of births, survival rates, and net migration. The numbers of current and future centenarians tend to rise if either past births or survival rates rose, if both rose, or if net migration led to an increase in population numbers. Similarly, the numbers of centenarians tend to fall if either the number of births or survival rates fell, if both fell, or if net migration was negative. For a heavily immigrant country with improving mortality, such as the United States was over the last century, the numbers of births would tend to rise from year to year a century earlier and the current numbers of centenarians would tend to increase from year to year. Concomitantly, the maximum age at death would tend to rise from year to year because, with each passing year, a greater number of survivors will reach any particular extreme age and will be available to survive to a higher age. The maximum age of survivors and of deaths will also tend to rise if the number of births remains unchanged and survival rates rise.[2] On the other hand, if either the number of births or survival rates fall, the number of centenarians, the maximum age at death, and maximum life span would tend to fall in later years. In sum, there is a long-term relation between the change in birth-cohort size and mortality trends, on the one hand, and maximum age at death and maximum life span, on the other.[3]

---

[2]Consider a standard life table with a radix of 100,000 births and one remaining survivor at age 113. If the radix were 200,000, there would be two survivors at age 113 and one would likely survive to age 114.

[3]The average maximum life span, that is, the average age of persons 100 and over, would not necessarily rise if the number of births was increasing a century earlier or if survival rates over

Since the maximum age at death can be "driven up" in the way described, some analysts have concluded that maximum life span has no natural limit (Wilmoth 1998). Others have suggested that, even if population size rises and survival prospects for successive birth cohorts are improved, there are biological and biomechanical constraints on maximum life span that can only be overcome with developments in biogerontology that slow the aging process (Olshansky et al. 1990).

On the basis of population size alone, we would expect to find the greatest number of centenarians and supercentenarians in China and India, where the numbers of births have historically been huge in comparison to the numbers in the more industrial countries. These countries should also show an increasing maximum age of decedents because of presumed increases in the numbers of births a century earlier. We cannot be sure that these propositions are true, however, because historical survival rates in these countries have been so much lower than those of the more developed countries that the likelihood of survival of the larger numbers of births to the very advanced ages may have been sharply reduced. The census data and other source data on persons of extreme age for these countries have not been evaluated sufficiently to incorporate their records into a supercentenarian databank.

On the other hand, data that have been thoroughly evaluated for some countries of northern and western Europe provide the demographic evidence needed to demonstrate that the numbers of persons of extreme age and the maximum ages at death have been rising in these countries. However, the numbers of births have been falling in these same countries in recent decades. Many of the newborn children will survive beyond their 100th birthday, but the numbers of persons reaching such extreme ages as well as the maximum ages of deaths in these countries may tend to fall a century from now, unless sharply declining mortality at the advanced ages helps to maintain their numbers.

Historical and Theoretical Evidence for Increased Longevity

How long can humans live? Are there limits to life expectancy and life span? These questions have long intrigued not only scholars but the "average" person. The evidence, arguments, and theories relating to these questions are being debated by biodemographers, epidemiologists, molecular biologists, and other scientists. I consider these questions here under five headings, labeled (1) demographic perspective, (2) epidemiological perspective, (3) biological and biodemographic perspectives, (4) evolutionary perspective, and (5) engineering reliability theory. The dividing lines between these various areas of analysis are not clear-cut as all of them have interdisciplinary aspects, and hence the arguments overlap.

---

the century have been rising. Such changes could make the extreme aged population "younger" as larger numbers of survivors reached the younger ages in the centenarian age range.

Demographic Perspective

To evaluate the possibilities for a rise in maximum human life span as well as a continuing increase in life expectancy, demographers have analyzed data mainly for the countries of northern and western Europe, Japan, and North America. As suggested earlier, the records for these countries, particularly those of northern Europe, are believed to be sufficiently accurate to serve as probative material for these questions. As evidence, demographers point to (1) the long, essentially uninterrupted historical rise in life expectancy, (2) the acceleration in the rise of the numbers of older aged persons (85 years and over), (3) the increase in the numbers and percents of centenarians and supercentenarians, (4) the steady rise in the maximum recorded life span and the maximum age at death, (5) the deceleration in the course of death rates at the advanced ages of the life cycle, (6) the lack of a strong positive correlation between levels of mortality and rates of improvement in mortality, and (7) the momentum of cohort succession.

Historical Trends in Longevity

In the United States life expectancy at birth grew at a rapid but diminishing pace over the century from 1900 to 2005. As the reader may recall, in 1900–1902 it was only 49; it reached 68 by 1949–1951; and in 2005 it stood at 78 (Table 6.3).

By fitting a straight line to the logarithms of the highest recorded annual life expectancy for any country for the 1900–2000 period (in effect, fitting an exponential curve to the original data for these countries), Oeppen and Vaupel (2002) observed that the national world record for life expectancy at birth has increased at a steady rate for more than a century. They then predicted that life expectancy would continue to rise in the United States at the same average historical pace as during the twentieth century, and reach 100 years by the year 2065. The past trend is a matter of record, but its use to "determine" the future in this way is a matter of debate. Lee and Carter (1992) applied time series analysis to age-sex-specific death rates from 1933 to 1987, with very different results from the prediction of Oeppen and Vaupel. Their projection of life expectancy for 2065 is well below 100 years – about 86 years – but substantially higher than the projection of the Social Security Administration's Office of the Actuary for that year – about 83 years.

As discussed in Chap. 6, several dozen countries around the world have the same or. greater life expectancies at birth than the United States. The U.S figures in 2005 were 75.2 years for males and 80.4 years for females. Japanese women are setting the international record for females with a life expectancy at birth of 85 years, and Icelandic men and Japanese men are setting the international record for males, with life expectancies of 79 and 78 years, respectively (Population Reference Bureau 2005). The rise in life expectancy at birth shows no signs of stopping at this date. The United States ranked higher among the countries with respect to life expectation at age 65 and even higher at age 85, particularly for females, in spite

Table 13.4 Life expectancy at age 85 and age 65 for males and females, for selected countries: 2003–2005

	e_{85}		e_{65}		Ratio	
Country (year)	Male	Female	Male	Female	Male	Female
Australia (2003–2005)	5.9	7.1	18.1	21.4	.33	.33
Austria (2005)	5.3	6.1	17.0	20.3	.31	.30
Canada (2004)	5.9	7.2	17.7	21.0	.30	.31
Denmark (2004–05)	5.0	6.3	16.0	19.0	.33	.34
France (2004)	5.8	7.2	17.7	22.1	.31	.33
Germany (2003–2005)	5.3	6.1	16.5	19.9	.33	.33
Israel (2005)	6.5	6.7	18.0	20.2	.32	.31
Japan (2005)	5.9	8.0	18.1	23.2	.36	.33
Mexico (2005)	6.4	6.8	16.4	18.4	.33	.34
Netherlands (2005)	4.9	6.2	16.4	20.0	.39	.37
Spain (2003–2004)	5.4	6.2	17.0	20.9	.30	.31
Sweden (2005)	5.2	6.5	17.4	20.6	.30	.32
Switzerland (2004–2005)	6.7	6.8	18.1	21.5	.37	.32
United States (2005)	6.1	7.2	17.2	20.0	.36	.36
Uruguay (2004)	4.7	6.0	14.5	19.2	.32	.31

Source: UN *Demographic Yearbook*, *2005*, Table 22; U.S. NCHS/Kung, H.-C., et al. (2008). *National Vital Statistics Reports*, 56(10)

of its relatively poor record for life expectancy at birth (Table 13.4). As a result, the ratio of life expectancy at age 85 to life expectancy at age 65 is near the highest among the industrialized countries at 36.

As a result in large part of the increased survival to age 65 and the reduction of death rates above age 65, in the United States in recent decades, the numbers of persons 85 years and over and 100 years and over in this period have increased sharply (Table 13.1). In fact, there appears to have been an acceleration in the rise in the numbers of persons 85 and over and 100 and over. The number of centenarians appears to have increased 22-fold between 1950 and 2000, from 2,300 to 50,000, although even now centenarians are still relatively rare.

A definitive study of the relative contribution of fertility, mortality, and immigration to the recent rise in the number of centenarians in the United States has not been conducted. It is likely that all three factors contributed to the rise since the numbers of births were increasing during the last decades of the nineteenth century and the first decade of the twentieth century, mortality rates were falling steadily through the twentieth century, and the flow of migrants was strong at least until the First World War. However, I am postulating that most of the rise in the number of centenarians today can be attributed to the increased survival of the birth cohorts of a century earlier.

Whether we consider census data, vital statistics, Medicare data, or Social Security records, we cannot accurately assess the numbers and trends of supercentenarians in the United States because the U.S. figures are too unreliable for that purpose. Data from western Europe tell us that the number of supercentenarians is increasing rapidly in these countries (Robine and Vaupel 2002). The first

confirmed supercentenarians go back to the 1960s but their numbers have increased exponentially since then. It is uncertain how many supercentenarians might have [lived] prior to the twentieth century.

All of the measures of advanced longevity for the United States show net increases in the half century after 1950, but closer examination reveals that those measures reflecting change within the 100-and-over age group hardly increased between 1950 and 2000. According to the U.S. life table for 1950 the mean maximum life span of survivors aged 100 or more was 101.9 (for both sexes combined), and according to the 2000 life table, the corresponding figure was 102.0 years. In 1950 total life expectation at age 100 (life expectation at age 100 plus 100) was 101.9, and in 2000 it was 102.2. In 1950 the age corresponding to a life endurancy with a probability of 0.01 was 96.3 years, and based on death rates at that time, 0.24% of the initial birth cohort would survive to age 100.0. The corresponding figures in 2000 were age 100 and 1.22% of the initial cohort. From these figures and the others in Tables 13.2 and 13.3, we can reasonably conclude that survival to the century mark has been increasing, but that the measures of progress within the centenarian population do not show this trend. We cannot say with confidence, therefore, that the centenarian population in the United States in 2000 is older than it was a half century earlier.

Maximum Age at Death

The evidence from Europe gives a different impression. The maximum age at death or the maximum recorded life span has been steadily increasing at least for 140 years in Sweden (Wilmoth and Robine 2003). The "rate" of increase has also been rising. It was 0.4 year per decade before 1969 and 1.1 years per decade thereafter. Maximum age of death in several countries of Western Europe has risen even more rapidly. It rose linearly since mid- nineteenth century at an average "rate" of three months per year (2.5 years per decade) for females and at an average "rate" of 2 1/2 months per year (2.0 years per decade) for males (Oeppen and Vaupel 2002). For every year since 1977 the oldest validated age at death was 110 years or higher. For the 7 countries with the most thoroughly validated data, maximum age at death rose 1.4 year per decade from 1977 to 2000 (Wilmoth and Robine 2003). Wilmoth and Robine attribute the rise in the maximum age at death in Sweden largely to declines in mortality over age 70 and secondarily to increased numbers of survivors to old age (reflecting both larger birth cohorts and increased survivorship from birth to age 70).

Age-Pattern of Mortality Rates

In Chap. 3, I described various efforts to model the age-specific curve of mortality mathematically. Recall that the rate of increase in death rates according to the Gompertz model approximates constancy at most adult ages but then tends to

slacken off at the very advanced ages. At the more extreme ages the rate of increase may possibly become zero or even negative, although one cannot be certain of the pattern because of the sparseness of the data above age 105. As stated in Chap. 4, the phenomenon of decelerating old-age mortality was actually observed and reported by Gompertz using English data (Gompertz 1825). It has been known for more than 175 years, therefore. Other analysts have noted it in the years since 1825 and several of them have described it recently: Robine and Vaupel (2002); Horiuchi and Wilmoth (1998); Wilmoth (1998); Olshansky et al. 1998; and Robine and Saito (2003). The phenomenon has been observed to occur in several countries, and it has also been observed in a number of subhuman species. Robine and Vaupel (2002) reported, on the basis of the experience of a list of validated supercentenarians constructed from the International Database on Longevity (IDL; n = 159), a probability of dying at age 110 of 0.52 and then virtual stability of the rate just below this level until age 114. However, the sparse numbers used to generate these probabilities make it difficult to assess their reliability.

The deceleration in age-specific death rates appears to occur after about age 85 in the United States data. This pattern appears in the official life tables for the United States in 1979–1981 and 2000. These official U.S. life tables show rising rates of mortality between adjacent 5-year age groups until about age 85, after which the increase in the rates begins to drop off until about age 105. Table 4.2 shows rates of mortality change for single ages from age 84 to age 98 for 1997 based largely on Medicare data. There is a steady and substantial decline in the rate of increase in death rates over the whole range of these ages.

Some demographers maintain that the historical trends described and the pattern of decelerating rises in death rates at the later phases of the life cycle are inconsistent with the assumption of limits to life expectation and life span. They contend that, based on such evidence, life expectancy and life span can continue to increase without any specifiable limit and that there are no biological or demographic constraints on the trajectory of death rates preventing their rate of increase from declining to zero (Wilmoth 1998). According to this view, the maximum life span is indeterminate (Oeppen and Vaupel 2002; Vaupel 2003; Wilmoth (1997); Wilmoth et al. (2000).

On the other hand, maximum recorded life spans or maximum ages at death may be viewed as outliers not likely to characterize any general population. The upward trend in maximum recorded life span may come to a halt and reverse itself at any time. Past increases in the size of the population at advanced ages cannot assure us that the maximum age at death will continue to rise, because the sizes of birth cohorts a century earlier or survival rates at some higher ages may fall. The slowing of the rise of mortality rates in later life may be due in large part to the heterogeneity in the health composition of the members of a cohort. The concept of heterogeneity here implies that any cohort includes subgroups with different risk levels and, as a result, the subgroups with the higher hazard rates die off in greater numbers earlier in life while those with lower hazard rates survive to later ages. This process transforms the population in the advanced age groups into one with

more uniformly lower mortality levels than the "same" population at the younger ages, which included the less healthy subgroup that was eliminated. This process may also result in populations at the advanced ages that are healthier than the population in the same advanced ages at an earlier date. It can be demonstrated that the rate of increase in death rates with rising age of a population group can fall even though the death rates for the two segments of this group with different hazard rates are rising. However, some analysts argue that the relative declines in age-specific death rates at the advanced ages have been too great to be accounted for wholly by heterogeneity with regard to health (Vaupel 1997; Wachter 2003). They maintain that the deceleration of death rates is due in large part to declines in the hazard rate for individuals.

Temporal Shifts in Age/Cause Pattern of Mortality Rates

In Chap. 6, I described the epidemiological transition as a shift from diseases that are externally caused, acute, and relatively easy to treat (i.e., exogenous causes) to diseases that are internally caused, chronic and progressive, and difficult to treat (i.e., endogenous causes). In the industrialized countries deaths from the former causes typically occur in childhood, youth, and young adulthood. Death rates for these ages have fallen so low in these countries that there is little room for further improvement. Any further major improvement in mortality rates must come from reductions in the death rates at the older ages, where the deaths from endogenous causes are concentrated.

The significance of this shift in the age-cause pattern of mortality is that further reductions in mortality could be increasingly more difficult to achieve because endogenous causes of death are influenced strongly by the biological processes of aging and knowledge of these processes is quite limited. Recall that, while death rates from the major cardiovascular diseases fell sharply in the 1970s and 1980s, the rate of decline since 1990 has been well below that for the earlier decades and the mortality from cancer, diabetes, and several other endogenous diseases had either increased or remained nearly stable from 1970 to 1990 or beyond. It is reasonable to contend, therefore, that the death rates from endogenous diseases are unlikely to decline during the next several decades at the same rate as they declined during the last several decades, and that overall age-specific death rates will not decline in this century at the same rate as they declined during the last century. The rates may not only show far smaller declines but they may even shift direction, as discussed further below.

Relative Changes in Mortality and Longevity

In Chap. 4, I indicated that relatively large declines in age-specific death rates have been associated with relatively small increases in life expectancy at low levels of

mortality. Refer again to Table 4.13 for a comparison of percentage reductions in age-adjusted death rates and percentage increases in life expectancy at birth during the decades from 1900 to 2000 for the United States. The historical relation between these series is not very close, however, so that we cannot say that, when U.S. life tables are analyzed over a range of mortality levels, declining increases in life expectation at birth are strongly associated with increasing reductions in age-specific death rates. That is, these data do not provide evidence of a strong negative correlation between changes in levels of mortality and changes in levels of life expectancy.

Although substantial reductions in death rates and substantial increases in life expectancy at the older ages occurred in the last half of the twentieth century, it appears that at current high levels of life expectancy sizeable percentage reductions in death rates must occur to achieve relatively small increases in life expectancy. For example, an increase in life expectancy from 77 years, the figure for the United States in 2000, to 85 years, or by 8 years, would require the equivalent of a uniform 50-percent reduction in age-specific death rates. Achieving a life expectancy of 100 would require a uniform reduction in present age-specific death rates of about 95% (See also Olshansky et al. 1990). Figure 13.1 illustrates the relation between a (uniform) percent reduction in death rates and the percent increase in life expectancy at birth, based on calculations with the U.S. life table for 2000 ($e_0 = 76.9$ years).[4] It shows a moderately curvilinear relationship between these measures. These calculated increases in life expectations associated with the assumed percent reductions in mortality rates are probably somewhat understated because they were based on a life table for year 2000 that terminated at ages 85 years and over rather than 100 years and over.

[4] Keyfitz' H and Δe formulas, discussed in Chap. 4,

$$H = -\int_0^\infty l(x)\ \ln\{l(x) \div l_0\}dx \div T_0$$

$$\Delta e = -(-k)\ e_0 * H$$

can also be applied to derive a theoretical series of estimates of gains in life expectancy at birth corresponding to uniform percentage reductions in age-specific death rates for the United States in 2000 (H = .1590 and k = assumed percent reduction in death rates):

% reductions in death rates	Gains in life expectancy	Life expectancy
0	–	76.9
.10	1.22	78.1
.50	6.11	83.0
.80	9.78	86.7
.90	11.00	87.9
1.00	12.23	89.1

Again, we observe the curvilinear pattern of the changes in life expectancy. The series terminates with the unrealistic result that, with 100% reductions in death rates, e_0 increases by only 12.2 years.

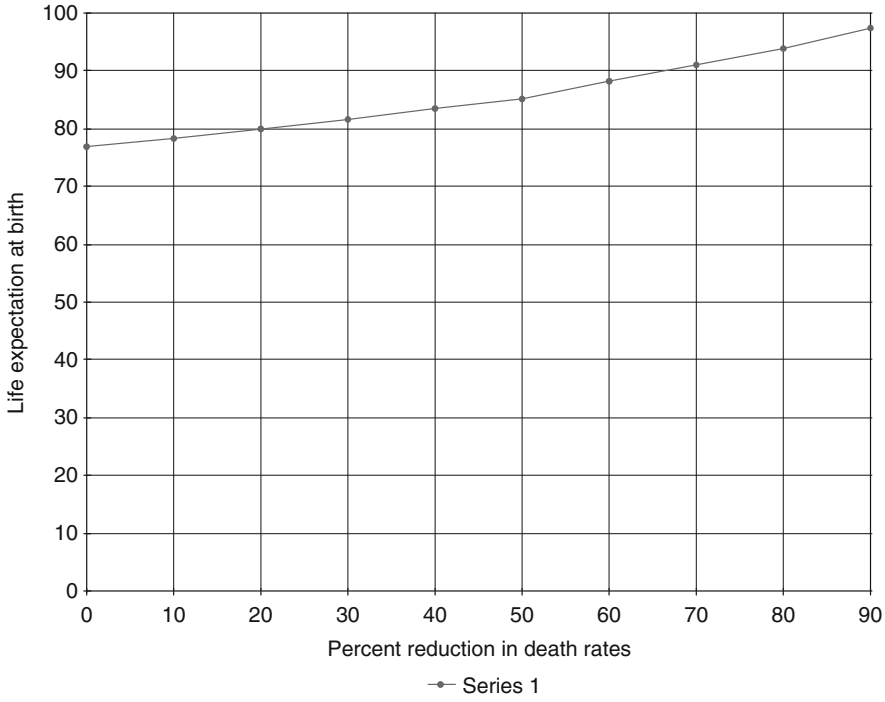


Fig. 13.1 Life expectancy achieved by reductions in life table mortality rates: United States, 2000 (Note: The estimates are biased upward to a small extent because the base life table for 2000 terminates at ages 85 years and over)

As noted, Oeppen and Vaupel (2002) predicted that life expectation in the United States would reach 100 years in 2060. According to the U.S. life table for 2000, only about 1.8% of a birth cohort reaches age 100 (U.S. NCHS 2002a) and the complete expectation of life at age 100 is 102.6. For a life expectation of 100 in 2065, about 56% of the birth cohort would have to reach age 100, age-specific death rates would have to be reduced by about 95%, and the complete expectation of life at age 100 would have to be about 106 years. This would be an amazing achievement! Olshansky et al. (1990) consider that attaining an expectancy figure of 100 is implausible, if not impossible, because it would require the elimination of all deaths from endogenous causes. My empirical calculations (Fig. 13.1) and theoretical calculations based on Keyfitz' H and Δe_0 also suggest that it is highly improbable that this goal can be achieved.

Cause-Elimination and Limits to Life Expectancy

If we track the life expectancy that would have been achieved in the United States between 1960 and 1990 on the assumptions that each of the causes of death

(including the "all other causes" class) was eliminated and that the causes are independent, we would have a nearly horizontal trend line at about 90 or 91 years:

Years	Life expectancy without "gains"	"Gains" in life expectancy	Life expectancy including "gains"
1989–1991	75.4	14.3	89.7
1979–1981	73.9	18.6	92.5
1969–1971	70.8	20.1	90.9
1959–1961	69.9	18.9	88.8

Source: Based on U.S. NCHS decennial life tables for the indicated years

For example, in 1990, the sum of the years gained by eliminating each of about 20 classes of causes of death, 14.3 years, and the current figure for life expectancy, 75.4 years, yields a total of 90 years. The corresponding figures for 1960 sum to 89 years (=69.9 + 18.9). Given the ongoing marked reduction of death rates at the older ages after 1970, and the associated rise in life expectancy, the total number of years gained by the independent elimination of the various causes of death would tend to decline after that year. The addition of the declining number of "years gained" to the rising life expectancy yields a series of hypothetical life expectancies of 89–92 years for the United States in these 30 years, with a mean of 90.5 and no discernible trend.

Inasmuch as this series displays a fluctuating trend rather than a monotonic one and is not fully comparable, it is not possible to extrapolate it with confidence. I would expect the series to be nearly constant if the data were accurate and the figures consistent and comparable. As life expectancy rises and cause-age-specific death rates fall during any period, the increase in life expectation and the reduction in the total gains from eliminating the various diseases should approximately balance.

Competing risks and comorbidity. The independent elimination of causes of death suffers from at least two interrelated paradoxes. The first is the so-called Taeuber paradox, which posits that the elimination of a cause of death in later life would result in increases in other associated causes since the population saved from one disease is exposed to the risk of dying from another. The assumed elimination would result in increases in the size of the population exposed to risk as well as in the rates for other causes that affect various ages in the same general vicinity as the cause eliminated. This is a reflection of the principle of competing risks.

The independence of mortality risks is not a realistic assumption for the relation of cause-specific death rates, as the Taeuber paradox suggests. Because of (1) the competing risks of death, (2) the multiplicity of causes that contribute to or are associated with each death, that is, the comorbidities characteristic of older persons, and (3) the impossibility of human immortality, in the event of the elimination of any cause of death, or even its sharp reduction, the death rates from other causes – particularly those usually associated with the cause eliminated and with average ages of death close to that of the cause eliminated – will tend to rise. Stallard's (2002) study of underlying and multiple-cause mortality showed

conclusively that "death is due not to just one single disease but to a complex set of interacting pathological processes. In these cases the designation of any single disease as the underlying cause of death provides a distorted description of the causal pathways." It is even conceivable that the sharp decline recorded for death rates from cardiovascular diseases between 1970 and 1990 "contributed" to some extent to the lack of substantial progress in reducing cancer during this period, even though cancer is not usually reported as a contributing cause of death from cardiovascular diseases. Accordingly, the net gain from eliminating some major diseases could be quite small, even over the short term.

The second paradox, which we may dub the cause-elimination limit, is closely related to the first. If we eliminate the causes of death as if they were independent of each other, as is assumed in constructing the cause-eliminated life tables of the National Center of Vital Statistics, the total number of years added to life expectancy at birth would amount to about 18 (ranging from 14 to 20), not the infinite number of years implied by the eradication of mortality. This interpretation of the cause-elimination tables suggests limits to the possibilities for extending life expectancy. The total of the gains from cause-elimination can be considered as a current maximum estimate of the possible additions to human life expectancy by eliminating various causes of death.

On the other hand, it could be argued that, if certain causes of death were eliminated, the death rates from some other causes would be lower than recorded because the specified eliminated cause could not have contributed to earlier deaths from these other causes. Furthermore, one might argue that, whatever factors contribute to declines in the risk of death from any single intrinsic cause are likely to have a similar dampening effect on other intrinsic causes of death. I believe, however, that this argument is less cogent than the one presented earlier.

It is unrealistic to assume that the major causes of death will be eliminated in the foreseeable future, but we may reasonably assume that some reduction in them can be achieved. The net effect on total life expectancy from the hypothetical elimination of any particular cause of death in an environment of competing risks can be only roughly estimated because the extent of the linkages of the various causes is not well known. With present knowledge, it is not possible to state how many years, if any, would be added to life expectation under the circumstances of the elimination of any single cause. It may not add any years to life expectancy in the real-life scenario. Hence, it may be hypothesized that the figure of 91 years could be viewed as a rough maximum for total life expectation for either males or females under the grand assumption of the reduction of death rates from a wide range of causes and the principle of competing risks.

Other Demographic Considerations

No chronic degenerative disease has been eliminated, or even nearly eliminated, in spite of all the efforts in this direction. There is no evidence that the age of onset of any major endogenous disease has been raised, although establishing the

age of onset of a disease is rather difficult. We have been able only to manage some of these conditions better so that their disabling effects are postponed and the quality of life for persons having the conditions is improved (e.g., hypertension, diabetes, atherosclerosis). Clearly some additions to life expectancy may result from medical interventions but these interventions can extend life expectancy only within narrow limits. Some classes of diseases may prove to be intractable. For example, the death rates from various forms of violence were reduced by only one-third between 1970 and 2000, so that these causes may be expected to continue contributing substantially to future mortality. As mentioned, the death rate for several leading endogenous diseases (e.g., cancer, nephritis, chronic obstructive pulmonary diseases, Alzheimer's disease) and septicemia, rose in this period and they may be expected to "ride high" for a long time. The obesity epidemic among children and the continuing widespread practice of smoking will contribute to the future difficulties of reducing the death rates of a range of endogenous diseases.

Epidemiological Perspective

The epidemiological view shares the demographic perspective in recognizing the role of the epidemiological transition (i.e., the shift from predominantly exogenous causes of death to the current dominance of the endogenous causes of death) and in having a moderately favorable view with respect to the prospects of reducing the endogenous causes of death. The "overweightness" and obesity "epidemic," especially among children, and persistence of smoking among a substantial share of the population, however, are expected to figure prominently in exacerbating the task of reducing the endogenous causes of death. Another major concern is the changing role of infectious diseases – a reflection of the new epidemiological transition. The resurgence of some old infectious diseases and the emergence of new ones may also make future progress in reducing mortality difficult, both at the younger and at the older ages.

Recent decades have seen the resurgence in the United States of several infectious diseases that had largely been obliterated, specifically measles, tuberculosis, whooping cough, and diphtheria, and the appearance of new infectious diseases, primarily HIV/AIDS, but also SARS, monkeypox, and West Nile virus. In addition, the death rate from influenza has increased considerably in the last decades of the last century. This trend is due to the diminishing effectiveness of many established antibiotics – the result of the mutation of viruses – the transfer of new viruses from animals to humans, hospital-acquired infections, the entry of masses of unskilled and uneducated immigrants, and the increasing numbers of poor persons, especially children, living or working under unhealthful conditions.

As suggested, an ominous threat is the recent sharp increase in "overweightness" and obesity – 37% between 1977 and 1999 – with their numerous negative health implications, such as an increased risk for diabetes, heart disease, hypertension, stroke, osteoarthritis, and various types of cancer (U.S. NCHS/Hoyert et al. (2004).

Olshansky et al. (2005) have estimated that the current negative impact of obesity on life expectancy in the United States will be a minimum of one-third to three-fourths of year, but that it is likely to rapidly approach and could exceed 3.5 years – approximately the negative effect of cancer on life expectancy.

Biological and Biomedical Perspectives

The biological perspective focuses mainly on the considerable age-related somatic deterioration after the peak fertility years in humans and most animal species, the relative contribution of genetic and nongenetic factors in health, disease, and human behavior, the many successes of medical and other human interventions, and the prospects for their continuation and extension.

Senescence

Given the steady accumulation of age-associated diseases with advancing age in the post-reproductive years, one can justifiably maintain that "normal aging" in humans is a roadmap to disaster in later life for most people. Accompanying the passage of chronological time and the accumulation of age-associated diseases is the process of senescence, the molecular and cellular pathogenesis that degrades the functional integrity of the physiological systems of the body. As a result, the longevity of individuals, as well as the life expectancy and average maximum life span of populations, are likely to be limited. Because of the biological processes noted, there is an average 50–80% loss in the functioning of the various physiological systems by one's 80th year relative to their peak capacity in adolescence. These relate to the functioning of the heart (e.g., reduced pumping rate), immune system (e.g., loss of bone marrow, compromised functioning of the system), endocrine system (e.g., declines in testosterone, estrogen, and growth hormone), lungs (e.g., lowered forced expiratory volume), kidneys (e.g., insufficient clearance of waste from blood), and bladder (e.g., reduced capacity) (US National Institute on Aging 2006). There are accompanying declines in muscle mass and strength, bone density, aerobic capacity, and glucose tolerance. The declines in psychological functioning are also enormous by the tenth decade (Baltes 2002).

The Aging Process, Molecular Instability, and Homeostasis

Since the leading causes of death are age-associated, the greatest risk factor contributing to increases in the leading causes of death with advancing age is the aging process itself. With increasing age, molecules, cells, organs, and physiological systems lose their functional capacity. In this way age changes increase the vulnerability to death of humans and other forms of life. Aging in later life may

be regarded largely as a stochastic (i.e., random) process associated with biological changes in the body. Hayflick (1998, 2004) describes the aging process from a biological point of view as the stochastically driven, systemic loss of molecular fidelity that, after reproductive maturation, exceeds repair capacity in animals that reach a fixed size in adulthood. Carey (2002) refers to aging as a shifting pattern of vulnerability to genetic and environmental insults. With advancing age the body finds it more and more difficult to maintain homeostasis (i.e., internal equilibrium of the various physiological systems), in spite of ongoing efforts toward repair, compensation, and accommodation. These efforts become increasingly less effective, and at some point it becomes impossible to achieve the proper functioning of one system without compromising the functioning of another system or systems and life itself is threatened. Biological research suggests, then, that death is the result of a number of probabilistic events, reinforcing one another.

Given these biological forces, Carnes and Olshansky (1993) see age 80 as the limit of the biological warranty period for humans and age 85 as the limit of life expectancy, barring scientific developments that slow the aging process (i.e., modify the biological rate of human aging). In their view survival to an age much beyond these years will be limited until such knowledge is secured. Hayflick (1998, 2000) is equally conservative. He argues that extending human life span can only be achieved by probing into the causes of the aging of cells, i.e., by determining why older cells are more likely to fail than younger cells, and he reminds us that, relatively speaking, not much research effort is being directed at this basic question. He doubts that the human life span can be significantly prolonged by genetic engineering. On the other hand, several other biogerontologists have a far more expansive view of the possibilities of slowing the age process, anticipating a considerable extension of human life span in this century.

Processes Contributing to Molecular Instability

Numerous factors and processes have been identified as contributors to molecular instability (Exhibit 13.2). These factors mainly cause genetic mutations. Among them are oxidative stress resulting from the action of oxygen free radicals and other oxidative processes in the cells – a product of cell metabolism-glycosylation (i.e., attachment of glucose) to proteins, DNA copying errors (including copy number variations), and radiation damage. Other destructive processes that have been identified are loss of heat-shock proteins, which aid in regulating the immune system, and DNA methylation, a process in which methyl groups, a class of air pollutants, attach to genetic material. The myriads of continuous cell divisions add to the risk of copying errors. As a result of these processes, over time genetic mutations accumulate. Some regulatory processes fail (e.g., tumor-suppressor and other disease-suppressor genes become ineffectual), and repair and scavenger operations, a process called autophagy, become weaker and cannot keep up with the destructive processes. In addition, some multifunctional genes that are beneficial early in life become destructive in later life, a process called negative pleiotropy.

A →	B →	C
Maintenance mechanisms	Causes of aging	Signs of aging
Continuous cell division	DNA copying errors Cumulative genetic mutations	Tumor-suppressor genes less effective
Detoxification of harmful chemicals[a]		
Autophagy[b]		Scavenger and repair operations less effective
Normal metabolic processes	Oxidative stress	
Proteostasis[c]	Misfolding of proteins	Proteins cannot perform normal functions
Familial genetic characteristics[d]	Cell senescence[e]	
Negative pleiotropy[f]	Glycosylation[g]	Loss and decreased effectiveness of heat-shock proteins[h]
Gene expression[i]	Mitochondrial dysfunction Methylation[j]	Genes are silenced; genes needed to suppress certain diseases are shut off
Epigenetic controls and stability[k]	Epigenetic defects	Development of tumors
Immune system		Decline in immune response
Temperature control		
Wound repair		
Physiological homeostasis[l]		↓ Cells die ↓ D Biological aging

Note: A factor in column A may lead to change in more than one process in column B and a factor in column B may lead to change in more than one process in column C. Roles of factors and processes overlap. Factors on the same row are not necessarily associated

[a]Including effects of radiation, smoking, and other environmental factors.

[b]Process by which defective and dying cells are cleaned up, often by self-ingestion by their own enzymes.

[c]The proper folding of proteins in cells

[d]For example, combinations of genes in siblings that support longevity

[e]For example, telomere shortening

[f]Multifunctional genes that are beneficial in early life but become destructive in later life

Exhibit 13.2 Principal biological factors in the aging process

[g] Glucose attaching to proteins; also known as cross-linking
[h] Heat-shock proteins (HSPs) are a family of proteins involved in maintaining cell homeostasis and regulating the immune system
[i] For example, genes that "drift away" from their early-life function of controlling development to other functions in later age
[j] Process in which methyl groups (CH_3) attach to DNA
[k] Epigenetics refers to the ways genes are expressed in different cell types and to the processes by which the genetic program is expressed in the development of an organism
[l] Coordinated functioning of all physiological systems, i.e., a healthful equilibrium of these systems

Exhibit 13.2 (continued)

Proteins do much of the work in the body at the molecular level. A protein's functioning depends on proper folding of its sections and alignment with other proteins. If a protein is formed incorrectly or becomes damaged and then misfolds, it cannot perform its normal function(s) or cannot be properly disposed of by cellular machinery. These processes lead to disease.

Another explanation also sees biological aging as primarily a genetic process, but differs from explanations of aging that focus on the accumulation of genetic (DNA) and cellular damage. This view posits that certain genes that control development weaken in time, "drift away" to other functions. Then, when they should respond to disease or injury to the body, they fail to do so and fail to support older cells in regenerating as easily as younger cells. This hypothesis does not assume that the aging process is directed by a few genes that specifically control longevity.

Identification of these factors and processes has given rise to many biological theories of aging, among them the free radical theory, the glycemic theory, the disposable soma theory, the wear-and-tear theory, the mitochondrial theory, the immunological theory, the neuroendocrine theory, and the autophagocytosis theory (the "junk" theory of aging). There are theories also as to how to extend human longevity, such as caloric restriction supported by optimal nutrition. There is an almost one-to-one correspondence between each maintenance mechanism, or the process by which it is compromised, and each theory of aging. For a long time researchers have devoted themselves to showing how one or another of these theories explains human senescence. More recently, it has been recognized that there are multiple causes of aging and that multiple deteriorative processes such as those described are all at work at once. Biogerontologists are taking a more synthetic view of the aging process (Holliday 2007:32). They are seeking new unifying theories at the same time.

One such general theory is Holliday's trichotomy of the availability of resources to sustain life into, first, normal survival functions such as feeding, metabolism, respiration, and other basic survival processes; second, functions of maturing growth and reproduction; and third, maintenance functions. (The theory resembles the disposable soma theory of aging, which states that the amount of energy available to an organism has three components – growth, reproduction, and maintenance.) The allocation of resources between the second and third functions – reproduction and

maintenance – varies widely among species, leading to the combination of early and high fertility with a short life span, or late and low fertility with a long life span (Holliday 2007:41–42). Aging, says Holliday, is "due to the eventual failure of maintenance. Maintenance is essential for the normal development of the adult and for the several decades of adult life. After that . . . we are no longer able to sustain the status quo."

In sum, from the biogerontological point of view, the basis of aging is no longer an unsolved problem. We see that Holliday and Hayflick essentially agree with one another, as do other leading biogerontologists (e.g., Steven Austad, Thomas Kirkwood). As Hayflick (1998) has stated, aging occurs because the complex biological molecules in our bodies become dysfunctional over time as the body's repair systems suffer the same molecular dysfunction and cannot do their job as efficiently. The balance shifts in favor of the accumulation of dysfunctional molecules, and the person manifests the age changes that we associate with aging. The molecular dysfunctional events eventially lead to an increase in vulnerability to age-associated disease.

Genetic Influences on Longevity

We can analyze the problem of aging in terms of the extent to which longevity can be partitioned between genetic and nongenetic influences. Alternatively, we can ask whether humans are programmed for longevity. With the growing knowledge of the human genome, it is thought that it may be possible to identify a gene or combination of genes that are associated with longevity. "Gerontogenes," or "longevity-assurance" genes, are being vigorously searched out by some molecular biologists, such as Guarente and Kenyon (2000). (See also Hall 2003.) So far they claim to have found one or more such genes in some lower-order invertebrates, specifically nematode worms (*C. elegans*) and fruit flies (*Drosophila melanogaster*). Some vertebrates (e.g., tortoises, rockfish, and walruses) live a very long time by human standards, even in the wild, and they may have longevity genes. Molecular biologists have not yet found a longevity gene in humans. Many biologists maintain that there are no longevity genes in humans and question the relevance for humans of the research on aging of nematode worms.

Some studies claim to have evidence that longevity runs in families. Perls and associates found such evidence in data on centenarians from the Boston Centenarian Study. Perls and Terry (2003) examined 444 centenarian pedigrees with at least one member who was 100 years old and over. The centenarians had 2,092 siblings. The researchers found that the relative survival probabilities of brothers of centenarians were 17 times, and of sisters eight times, as great as those of the general population. Figure 13.2 depicts the superior survival of the siblings of these centenarians. The survival curves for male and female siblings of centenarians are compared with the curves for the general male and female populations as represented by the birth cohorts of 1900. The study of Perls and Terry and the studies cited below apply the research approach called demographic selection to centenarians, which analyzes the

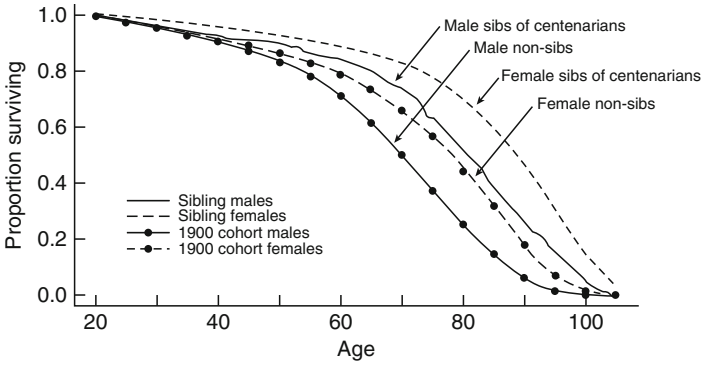


Fig. 13.2 Proportion of siblings of centenarians surviving from age 20 to later ages, compared with the survival of the 1,900 U.S. birth cohort, by sex (Source: Perls et al. (2002); Copyright © 2002 National Academy of Sciences, U.S.A. Reprinted with permission. Primary sources: Boston Centenarian Study; Social Security Administration, Office of the Chief Actuary/Bell and Miller (2002), Table 7)

characteristics and history of persons who are at the upper extreme of longevity to determine the factors that may have contributed to this outcome.

Twin and other family studies have shown a modest genetic effect on longevity (e.g., Ljungqvist et al. 1998; McGue et al. 1993.). A few recent studies suggest that the children of centenarians seem to be unusually healthy (Terry et al. 2004; Barzilai et al. 2001). The children show a markedly reduced prevalence of diseases and conditions associated with aging. Particularly notable are the low risk of cardiovascular diseases and the presence of favorable lipid profiles. According to Perls and Terry (2003), these studies support the hypothesis that phenotypic and probably genotypic characteristics conducive to exceptional longevity are transmitted in long-lived families.

More generally, it appears that a complex set of favorable genetic, environmental, and stochastic determinants of survival need to coexist for survival to age 100 and beyond to occur. Willcox et al. (2008) hypothesize that the high percent of Okinawan centenarians has resulted from just such a coalescence of favorable genetic factors, favorable lifestyle and social psychological factors, including caloric restriction, and a superior public health system.

Biomedical Developments

There are many ongoing and prospective innovative biomedical and sociomedical programs and developments that seek to restore health and reduce mortality. These programs and developments hold out the promise of greatly improving the health status of people of all ages and also increasing human life expectancy. These developments are described in a later section.

Evolutionary Perspective

Some arguments of those who maintain the view that there is a limit for human life expectancy and life span rest on the role of evolutionary biology in aging and senescence. In trying to answer the questions, is there an upper limit to human life, and if so, what is it, evolutionary biology brings to bear the comparative analysis of the longevity of different species, including invertebrates as well as subhuman vertebrates. Comparative analysis of primate life spans conducted by Judge and Carey (2000), correcting for body and brain size, suggests that the evolved life spans of humans probably fell in the range of 72–90 years. These numbers are important because they set a line between the segment of human life span that evolved (e.g., 75 years) and the segment that is very probably an artifact of living in a protected environment (that is, greater than 75 years).

Role of Natural Selection in Prereproductive, Reproductive, and Postreproductive Life

Evolutionary theory informs us that animals have responded to the high levels of extrinsic mortality (i.e., accidents, disease, starvation, and predation) throughout their existence on earth by reproducing early in life – 40 years of age or younger for humans – and then experiencing failure of their bodily structures. This evolutionary mandate dictates an age beyond which the probability of continued survival is extremely low. Only a century ago life expectancy hardly extended past the reproductive years, but today most people survive well beyond the ages of reproduction. Carnes et al. (2003) have called the extended period allowed by evolution for human survival the "biological warranty period." For them, the duration of life needed by humans to achieve maturation and reproduction – the period from about age 13 to about age 40 – helps define the beginning of their biological warranty period.

Evolutionary theory also maintains that somatic degradation is associated with the decline in the force of natural selection in later life. That is to say, after the end of the reproductive period, natural selection does not work to preserve the integrity of body structures and processes with the same level of efficiency as it did before puberty and during the reproductive period. Maturation and reproduction are the key elements in the survival of a species and pressures from extrinsic causes of mortality (i.e., accidents, predation, starvation, infectious diseases) establish a period within which maturation and reproduction must occur. The time scale for reproduction and natural selection vary according to the life span of the species, but the curves of the force of natural selection for sexually reproducing organisms, particularly mammals, look nearly identical to one another, as do those of reproductive success and mortality (Carnes et al. 2006). Figure 13.3 is a generalized portrayal of the decline in reproductive success for humans, with one curve reflecting the declining effectiveness of natural selection during the reproductive period, and with another curve reflecting the rise in mortality and the decline of natural selection in later life.

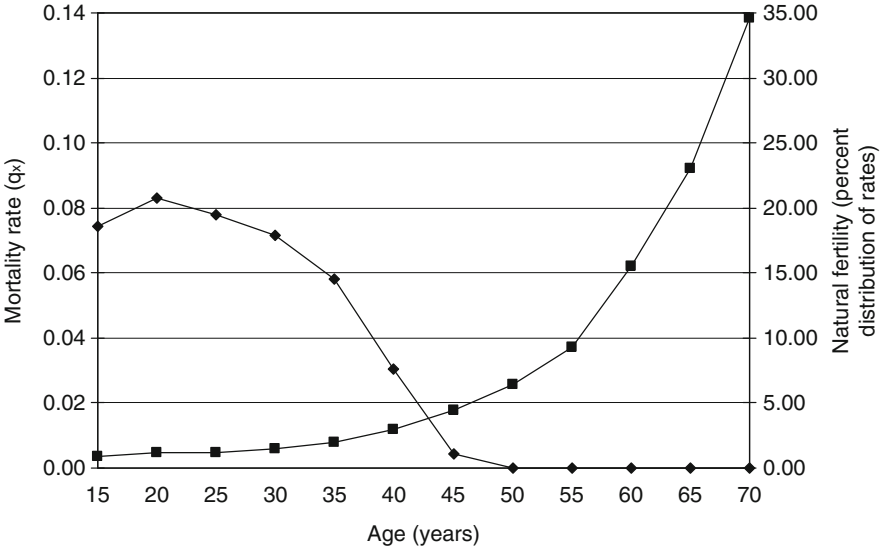


Fig. 13.3 Stylized representation of the age patterns of natural fertility rates and U.S. mortality rates for women (Note: Points are plotted at the initial ages of the 5-year age groups to which they apply;) Source: U.S. NCHS (2002a) and Table 13.2

Reproductive success is represented by the curve of natural fertility, as the latter term was defined in Chap. 9. As one curve falls, the other curve rises.

The lack, or near absence, of selection pressure in the postreproductive period suggests that there is no evolved aging process in this period of the life span. According to evolutionary theory, natural selection operates primarily on traits that affect an organism's ability to reproduce; accordingly, one would not expect evolution to favor genes that extend an organism's life much beyond its reproductive years.

Factors Accounting for Postreproductive Life

Evolutionary biologists have been trying to provide explanations in evolutionary terms as to why humans live as long as they do after the reproductive period. Postreproductive individuals are rarely found in nature except in a few species, and the period of postreproductive survival in nonhumans is relatively short (Austad 1997). In contrast, such individuals are common in captivity, especially among other primates. We seek to know whether the long postreproductive life of humans is an adaptive result of natural selection or is a non-adaptive artifact analogous to the postreproductive life of captive animals. According to Austad (1997), the commonality of postreproductive survival among captive animals does not mean that their increased longevity has evolutionary significance; these animals were simply not designed for such a long survival. Carey (2003) notes that various species have latent mechanisms, set in motion by certain environmental conditions, to

enhance their abilities at somatic maintenance and repair at the cost of reproduction. Among these environmental conditions that may contribute to their longevity is a scarcity of resources that may postpone certain physical developments in the species. Parents' raising younger generations and playing a role in intergenerational support may also prolong postmenopausal life. In contributing to the survival of their children and grandchildren, parents and grandparents serve a function in the evolutionary scheme.

Three factors appear, then, to account principally for the extension of life expectancy of humans beyond the prime reproductive years. First, evolution has endowed the body with a reserve capacity so as to assure the completion of its reproductive mission. It would be inefficient to risk failure in this mission by not building in such a reserve capacity, given the ever-present extrinsic factors vying for the life of the person. This implies that survival well past the end of the reproductive period can occur because we are essentially living on our physiological reserves. Next, evolution may have a "reproductive" role for older people. Most parents become grandparents and are thus available to nurture and train the young, who may go on to achieve successful reproduction (Vaupel 2003; Austad 1997). The extended longevity of *homo sapiens* evolved in a socio-ecological context of favorable environments (e.g., extensive parental care and investment, a high level of sociality), similar to the experience of subhuman primates and eusocial insects (Carey 2002). Finally, trends in life expectancy suggest that human intervention may have contributed to the success of many people in approximating, and of some people in exceeding, their biological warranty period. Survival into the postreproductive years has been achieved by human ingenuity and self-discipline, including public health innovations, medical developments, environmental improvements, increased knowledge of how to care for oneself, and especially lifestyle changes (Carey 2003).

While traditional evolutionary theory maintains that there is a tradeoff between investment in children and a female's survival that distinguishes the various species (Kirkwood and Rose 1991), for given human societies older mothers appear to live longer. Selection for reproductive success may be a principal factor in selection for longevity. Perls et al. (1997) support this interpretation. Mueller (2004) disputes it. She believes that the greater longevity of older mothers can be explained by genetic heterogeneity between family lineages, by differences based on socioeconomic status (that is, the more affluent women are likely to sire their children at older ages and to have higher survival prospects as well), or by both.

Age-Trajectory of Mortality Rates

As noted in Chap. 3 and earlier in this chapter, where the data are reliable, the rise in the curve of mortality rates for humans tends to slow down after about age 85 or 90 and may even level off and reverse direction above age 110. The classical evolutionary theory of senescence generally supports the view that the mortality rates at the higher ages in most species rise steadily, possibly exponentially. However, Carey et al. (1992) and Curtsinger et al. (1992) have shown that the hazard rates at extreme

ages in several invertebrate species (e.g., medflies, nematode worms, and yeast) do not continue to rise exponentially. In numerous species the fundamental shape of the hazard curve is approximated by the Gompertz formula (that is, a pattern of unchanging rates of increase in the hazard rates from young adulthood) only up to a point, after which the hazard rates decelerate (Finch 1990). This interspecies similarity suggests that the same biological patterns apply to human and subhuman species. On the other hand, there is considerable variability among species and within species in the manifestations of aging, senescence, and death.

Summary of Evolutionary Role

In sum, an evolutionary approach to longevity posits that there is a tradeoff between the level of fertility and the level of survival, and links duration of life to the timing of sexual reproduction. Extended human survival in the postreproductive segment of the life span is determined by the excess physiological capacity available after the period of reproductive success. Natural selection and evolution operate only up through the reproductive period, but then other biological, cultural, technological, and social influences begin to dominate. Life expectancies at birth that extend beyond the reproductive years are a recent historical phenomenon. Large-scale survival into the postreproductive period and the associated ascendance of the intrinsic causes of death over the extrinsic causes in the mortality schedule reflect the triumph of human ingenuity over many extrinsic causes and the apparent success of human intervention over the evolutionary process. At present, evolutionary biologists cannot agree on a convincing and consistent explanation for the considerable postreproductive extension of human life, but the several factors noted appear to play important roles (Austad 1997).

Engineering Reliability Theory

Finally, I consider engineering reliability theory. This theory describes the age pattern of the failure rates of manufactured products and has been interpreted as a general theory of aging and longevity. It purports to explain the peculiarities of the age trajectory of mortality rates on the basis of users' experience with manufactured products as these products age. The theory has been adapted to explain the age pattern of living things (Gavrilov and Gavrilova 2004, 2001). The bimodal shape of the curve for human mortality shown in life tables is observed for the curve of failure rates of machines such as automobiles (Siegel 2002). The failure rate of machines follows the Weibull model rather than the Gompertz model:

$$\mu_x = \mu_a \, Bx^c \tag{13.5a}$$

$$\ln \mu(x) = \ln \mu_a + \ln B + c \ln x \tag{13.5b}$$

Here the age-pattern of the force of mortality is assumed to be a power function of x. This formula is applied widely in the analysis of the reliability of manufactured products and in the implementation of proportional hazard models (see Appendix 1):

While machines begin life with all new parts, some come out of the factory with defects. Human beings begin life with all new but many defective parts also. On the other hand, humans have the advantage of many redundant parts, unlike most machines. According to reliability theory this redundancy may help explain the extended life expectancy of human beings after menopause. As with living things, the failure rates of manufactured products show a deceleration of the rates of increase at the advanced ages. Reliability theory has been used to argue that, given that the curve of mortality rise more slowly or even levels off at the advanced ages, there is no fixed limit to life span. This interpretation of reliability theory also rejects mortality compression theory on the basis of the latter's assumption that there is a limit to human life span. On the other hand, some analysts apply the analogy of the "sudden death" of manufactured products (e.g., automobiles) to humans and thereby argue that humans have a fixed life span that cannot be exceeded without slowing the aging process.

Research Programs to Increase Longevity

Two opposing, but not entirely mutually exclusive, positions have been taken as to how to extend human life expectancy and life span. One view, that of clinical medicine and geriatrics, would primarily direct efforts at the reduction of the death rates for specific chronic diseases of later life, such as the heart diseases, malignant neoplasms, diabetes, and Alzheimer's disease. The other view, that of biogerontology, would primarily seek to maximize efforts at understanding the networks of molecules that comprise cells and tissues, the ways these networks are regulated and interact with each other, and the molecular events that lead to health or disease. By devoting resources to understanding the basic processes by which cells become dysfunctional, the dynamics underlying a range of diseases may be discovered and more years may be added to life expectancy than by trying to reduce the incidence of specific diseases.

Biogerontologists maintain almost with one voice that success in the research on the molecular changes in the human body that occur with increasing age could lead to greater extensions of life expectancy and life span than research on specific major causes of death. From their view, study of the individual causes of death as such contributes little to the understanding of the process of aging (Hayflick 2000). Research in understanding why cells age is expected to lead to the control of a number of chronic diseases, thus adding greatly to human longevity. To extend human life to any marked extent, the basic processes of aging have to be understood before any of the chronic diseases of later life can be eliminated. To do this,

> we have to learn how to modify the processes that degrade the surveillance, maintenance, and repair processes of cells, cause disease, and ultimately cause the molecules needed for survival to lose their molecular fidelity. At present the only way to extend life is

> through interventions that manufacture survival time by treating the manifestations of disease processes. Some of the interventions, like an appendectomy, cure the problem by eliminating it. Others like heart bypass surgery and dialysis treat the disease processes but do not eliminate it. There are limits to the number of years that can be added by such interventions (B.A. Carnes, 2004, Private e-mail communication,B.A. Carnes to J.S. Siegel, 2/23/04).

Carnes et al. (2003) maintain that human bodies are not designed for extended operation and the historical ascendence of the intrinsic/endogenous causes of mortality over the extrinsic/exogenous causes suggests that the individuals in the more developed regions may be rapidly approaching their life span potential.

Many approaches are now being pursued by molecular biologists and research geriatricians to extend human longevity. Several of these confront the very processes by which cells are formed, are repaired, and die. One such area of research is the process by which telomeres, the caps at the ends of the chromosomes in the cells, shorten, leading, by their excessive shortening, to cell death (apoptosis). Another such area of research is the process of autophagy, by which used-up proteins, invading microorganisms, and malfunctioning organelles are cleared away in order to keep cells healthy. A third area relates to the processes by which body tissues are regenerated, whether it is nerves (neurogenesis), blood vessels (angiogenesis), or bones (osteogenesis). Knowledge of how to control these processes might enable the researcher to slow the rate of cell death, maintain the efficacy of the "clean-up" operation, or replace tissues that have died.

Improvements in health and increases in longevity could come from further progress in a variety of technical fields: Regenerative medicine, genetic engineering, tissue engineering, nanotechnology, and sonocytology. Regenerative medicine refers to the use of human and other organic materials to stimulate the body's healing power and replace defective parts, and includes such procedures as bone marrow transplantation, joint replacement, organ transplantation, and cartilage cell transplantation. Genetic engineering includes recombinant DNA applications (e.g., gene splicing), in which drugs based on human proteins, such as recombinant human insulin (for diabetes) and interferon alpha (for boosting the immune system in the treatment of hepatitis and cancer) are used to treat common diseases. Engineers in the relatively new field of tissue engineering are working on ways to help the body grow new cartilage, bones, organs, and other tissues damaged by injury or disease. Nanotechnology involves the application of engineering at the molecular level, including the creation of biological robots that can visualize the human body at the subcellular level and identify and prevent disease before symptoms appear. Robots may someday be programmed to maintain homeostasis, keeping body cells in a state of equilibrium. Sonocytology would advance the practice of ultrasonography by providing sonograms for the internal organs at the subcellular level, thus making possible the detection of abnormalities at a very early stage.

Other possible developments include new drugs and medical devices, such as improved medical imaging devices, sensor-driven drug delivery, remote monitoring devices, and implantable neurostimulation systems. We can envision, in addition,

expansion of such areas of research as caloric restriction, preservation of heat-shock proteins, hormone treatments, and control of lifestyle risk factors (e.g., diet and exercise).

For further discussion here I select six promising approaches that may be expected to contribute to human longevity during the next half century: Improvements in the functioning of the immune system, genetic alterations, caloric restriction, organ regeneration and replacement, technological advances in robotics and imaging, and pharmacological interventions. These approaches are generally more consistent with the broad-spectrum view of the molecular biologist than the single-cause approach of the geriatric researcher.

Improving the Functioning of the Immune and Endocrine Systems

Research into the study of the change in the functioning of the immune system as the body ages is driven by the fact that the aging process is associated with a reduction in the ability of the immune system to respond to the demands made on it and that this reduced response contributes to more infections, more inflammatory diseases, more uncontrolled growths, and slower recovery from attacks by external microbial agents. The evidence of the consequences of the reduced efficiency of the immune system is widespread. For example, the elderly are far more likely to contract infectious diseases such as influenza, pneumonia, and gastroenteritis than persons of middle age.

Research is aimed at ways of strengthening the aging immune system, as by modifying the diet, pursuing appropriate exercise programs, reducing stress, avoiding environmental assaults, securing recommended vaccines, and adopting other lifestyle changes. Efforts are being made to achieve more widespread acceptance of existing vaccines for adults, such as the pneumococcal pneumonia vaccine, the HPV vaccine (human papillomavirus vaccine), the adult pertussis/diphtheria/tetanus vaccine, and the new vaccine for shingles; and to develop a new vaccine for HIV/AIDS. Because HPV infections cause most cases of cervical cancer, this disease may be eliminated soon in the industrial countries. These vaccines deal with microbial diseases, however, rather than with the endogenous/intrinsic diseases of later life.

Two other physiological systems, the endocrine system and the neurological system are also extremely important in the aging process, and collaborate closely with the immune system in directing and monitoring the other physiological systems of the body and in maintaining homeostasis of body processes. They are involved in metabolism and waste discharge (i.e., via the gastrointestinal system and urinary system), locomotion and coordination (i.e., via the musculoskeletal system and circulatory system), and sensory responses.

The endocrine system produces hormones that are needed for the effective functioning of all the other systems of the body. Like the immune system, the

endocrine system functions less efficiently with advancing age. Some research has been directed at the effects on health and mortality of different levels of hormones and the effect of supplementing the body's natural production of various hormones. For example, DHEAS (i.e., dihydroepiandrosterone sulphate), the major sex steroid produced by the adrenal glands, is claimed to have anti-aging effects. DHEAS levels decline greatly with age. Low DHEAS has been associated with increased cardiovascular mortality in older men. The relation of DHEAS and mortality in older women has not been firmly established, but a U-shaped relation between the level of the hormone and adverse effects in women has been tentatively shown (Cappola et al. 2006). A U-shaped relation has commonly been seen by researchers in endocrinological conditions. Such a pattern of pathology with deficiency and excess is seen in hyperthyroidism and hypothyroidism, and Cushing's syndrome and adrenal insufficiency. Understanding this relation is important in order to determine whether dietary supplementation of DHEAS is desirable for prevention of premature mortality.

Genetic Alterations

Success in mapping the human genome has opened up a multitude of possibilities for research in associating particular genes with particular diseases and with aging. Some molecular biologists are researching the possibility of the existence of a "longevity gene" or combination of such genes. They hope to find a longevity gene or genes in humans equivalent to those found in a few invertebrates. C. Kenyon, L. Guarenti, M. Rose, and other molecular biologists are anticipating success here and have teamed up with genomic production companies. L. Hayflick, J. Olshansky, and B. Carnes, among others, strongly discount this possibility. The former researchers are trying to find common genetic pathways to aging across species by exploring biologically conserved mechanisms of senescence. They believe that aging is caused, or substantially influenced, by a limited number of critical "gateway" genes, and that if the existence of these genes can be discovered and a way can be found to manipulate their gene products, life in all the species, including humans, can be dramatically extended.[5]

Genetic alterations can now add many days to the life of a mouse. Experiments with mice carried out by Rose suggest that the longevity of mice can be increased by selectively breeding the mice that have offspring late in the reproductive period. There is also evidence that women who have children late in life live longer than

[5] Some scholars oppose calling the genes that influence the duration of life "longevity genes," principally because the use of this term implies that such genes are the direct product of natural selection. They maintain, rather, that these genes are an inadvertent by-product of genetic programs for growth, development, and reproduction, and that manipulating them to extend life will also influence these early life processes.

other women (Doblhammer 2004). If the genetic paths that lead to longer life in humans can be determined, drugs could be devised by which the aging process could be manipulated.

Genetic alterations can also be achieved by gene therapy[6] and by stem-cell therapy.[7] Gene therapy and stem cell research hold promise for treating and even curing a wide range of chronic illnesses. The sequencing of the human genome poses new possibilities for progress through gene therapy, even though at present the functions of most genes, and the combinations of genes that oversee most cellular and tissue functions, are unknown. Gene therapy is being widely explored. For example, gene therapy is being used to stimulate bone growth (i.e., osteogenesis). It is being evaluated for its effectiveness in stimulating the growth of blood vessels (i.e., angiogenesis) and reducing, for example, the effects of peripheral artery disease. Alternatively, gene therapy is being evaluated for its effectiveness in suppressing angiogenesis (i.e., anti-angiogenesis) and thus cutting off the blood supply to unwanted body growths. Anti-angiogenesis therapy in the form of anti-angiogenic drugs is being experimentally used to treat macular degeneration. Scientists have already found genes that are tied to particular diseases, such as Alzheimer's disease, and scores of genes linked to cancers, and they have identified genes that support particular functions, such as angiogenesis.

Like gene therapy, stem cell research holds promise for treating and even curing a wide range of chronic illnesses. Stem cells, particularly embryonic stem cells, can be programmed to grow into a variety of specialized cells and tissues. There is widespread hope, therefore, that embryonic stem cell therapy will make possible cures for Alzheimer's disease, Parkinson's disease, spinal cord injuries, and many other leading chronic diseases. Stem cells are now being experimentally injected into hearts to determine their regenerative capability and their ability to improve the function of hearts that are not receiving enough blood. Stem-cell therapy is now being used in the form of bone marrow transplantation for some types of cancers. It may eventually be used to replace the damaged nerve cells in the brains of Alzheimer's-disease patients.[8] The transplantation of blood-forming stem cells combined with an immune-suppressing drug (rapamycin) is being used to test the

[6]Gene therapy has been made possible by the development of recombinant DNA technology, the process by which DNA can be split into pieces, recombined in new ways, and reinserted in new places in the body.

[7]Stem cells come in several types but all are pluripotent, that is, they can turn into many types of tissue–muscle, bone, skin, and so forth. The most versatile and promising stem cells are embryonic stem cells, but there is a controversy over their use because, in being harvested from embryos, the embryos are destroyed. This issue is discussed further in Chap. 17. I imagine that this controversy will be resolved shortly and that embryonic stem cell research will soon be widely employed in the pursuit of cures for the major illnesses of later life.

[8]For example, M. Tuszynski (Univ. of CA, San Diego) injected the brains of eight early-stage Alzheimer's patients with their own skin cells. These cells had been genetically engineered to produce nerve growth factor, a chemical that boosts repair and regeneration of the brain's nerve cells. The preliminary results showed the technique to be safe and somewhat beneficial (*AARP Bulletin*, July-Aug. 2004:13).

possibility of reversing sickle-cell anemia in adults. A combination of gene therapy and stem cell therapy is currently being tested on diabetic mice in the hope of improving blood flow, with considerable success to date.

Organ Regeneration and Replacement

We have arrived at the era of regenerative medicine and tissue engineering, with the expectation of replacing most organs, tissues, and cells. A continuing problem in medical care is the shortage of body organs for transplant. This is motivating a research effort to determine the conditions under which body organs can be regenerated or humans can receive organs from animals. One research protocol now under way is to transplant the hearts of pigs into baboons, with the intent that at some future time the pig hearts can be transplanted into humans. The process of transplanting animal organs into humans is called xenotransplantation.

Attention is being directed to the study of salamanders, a lizardlike amphibian, that can naturally regenerate its appendages when they are amputated. This is the only vertebrate that has this ability and further understanding of it may give clues as to the ways that regeneration in humans might be accomplished.

Caloric Restriction

Calorie restriction (CR) stands as the single method now known than can increase human longevity, or at least is a reasonably promising device for increasing human longevity. Numerous animal experiments have shown that a nutrient-dense diet, low in calories and low in fat, promotes health, retards aging, and extends life span (Fig. 13.4). It seems to postpone many age-related pathologies in animals. Physiological gains have been firmly demonstrated for rats and mice, but not yet for humans and other primates. A study conducted by Walford et al. (1992) concluded that drastic reductions in blood pressure, cholesterol, triglycerides, and other major risk factors for heart disease, along with reductions in risk factors for diabetes (e.g., "overweightness," high blood sugar levels, low body response to insulin) and possibly other leading causes of death such as cancer, may be achieved in normal individuals by their pursuit of a carefully chosen restricted diet. Specifically, the study showed that a low-calorie nutrient-dense regime produces physiological gains in humans similar to those in other animal species.[9]

[9] According to Walford, daily-caloric intake should be cut to 2000 for average-sized men and 1800 for average-sized women. The goal is to reach a weight that is 10% to 25% below one's "set point." The set point is one's characteristic weight when not overeating or undereating. (Internet, *Bottom Line/Health* interview with Dr. Roy Walford, 1992.)

A small sample of members of the Calorie Restriction Society, an informal association of people presumably observing the Walford-type diet, has been followed for 3 to 15 years by

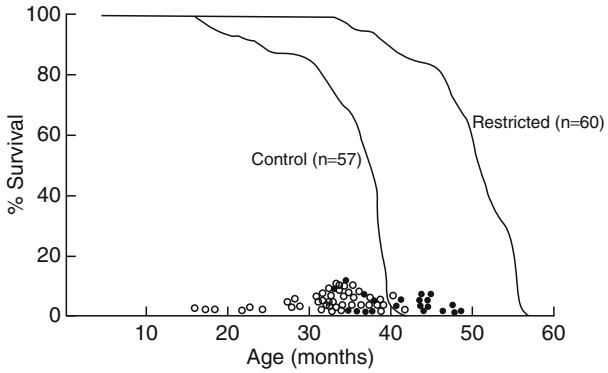


Fig. 13.4 Percent surviving of two groups of mice, one calorically restricted and the second a control group not calorically restricted (Source: R. J. Hodes, Director, U.S. National Institute on Aging, National Institutes of Health, public lecture, Nov. 2005)

A drastic reduction in calories may reduce damage from the harmful byproducts of oxygen metabolism. It may initiate metabolic changes that strengthen the immune system and increase the body's ability to produce new healthy cells. Thereby the process of aging is slowed. Caloric restriction (CR) seems to work the same way as the genetic manipulation of yeast, worms, flies, and mice; this device has been used to extend their life spans. On the other hand, it is questionable whether humans will tolerate on a routine basis the drastic caloric restriction required to achieve substantial benefits. People on these diets are known to feel cold, suffering from a dysregulation of their body temperatures, to feel very hungry, and to be subfecund. A study by Racette et al. (2006) showed that CR was feasible for one year, but the level of CR achieved was less than prescribed. In this study CR and exercise were equally effective in reducing weight and adiposity.

The importance of these studies is that a plan following drastic caloric restriction over many years may significantly extend longevity. If this finding is sustained, the practical application is to encourage more people to eat low calorie diets or, more realistically, to encourage pharmaceutical firms to develop a safe medication that suppresses appetites.

Pharmacological Interventions

Pharmacological interventions have long proved their efficacy in reducing disease and adding to life expectancy. A drug may be the mechanism by which the goal

J. Holloszy of Washington University, St. Louis. This study shows "profound and sustained beneficial" effects in various biomarkers of health (e.g., low cholesterol, low blood pressure, low C-reactive protein) from drastic caloric restriction (1150 to 1950 calories per day).

of caloric restriction is finally achieved. Thousands of drugs are now being tested for their possible efficacy in turning off genes that are involved in cancer, diabetes, heart disease, and other diseases. With respect to cancer, the design of research is to profile cancer genes, determine the basis of their abnormality, and then try to disrupt these abnormal processes by developing appropriate medications. In the case of colon cancer, one drug now available works to starve the cancer cells while another aims to block a protein on the surface of the tumorous cells. With respect to Alzheimer's disease, current drugs deal only with its symptoms, but several clinical trials are under way designed to treat or cure the disease itself, either by preventing the formation of the amyloid beta protein responsible for the short-circuiting of neural transmissions in the brain of Alzheimer's patients, limiting its further growth, or ending the toxicity that it causes.

Developments in human genomic research may make it possible within a few decades to personalize the composition, selection, and administration of various drugs. Genomic profiling of an individual may reveal a given genetic predisposition and suggest a different composition of a medication from the general standard composition. This is the concern of pharmacogenomics. Drugs may be made available for categories of individuals fitting particular genomic profiles. This could result in fewer medication errors and increased efficacy of the medications that are prescribed for an individual. Certain persons have adverse reactions to certain drugs and others do not have such reactions because of their distinctive DNA profile; pharmacogenomics is dedicated to distinguishing these individuals.

We could consider the taking of special foods and dietary supplements as a form of drug intervention, just as we could the taking of vaccines. The virtues of one "food" or another are heralded from time to time as a general magic bullet for health. An ingredient in red wine, resveratrol, has recently been selected for this role, one which was previously assigned to some other widely used dietary supplements, such as echinacea and ginkgo. The putative beneficial health effects of resveratrol for humans include its anti-cancer, anti-viral, anti-inflammatory, and life-extension effects (Sinclair 2006). Such effects have been demonstrated for mice, but they have not been fully demonstrated for humans. The amount of red wine consumed in France may contribute to an increase in longevity, but it is not enough to explain the "French paradox," that is, the low incidence of heart disease in parts of France where the diet includes lots of saturated fats.[10]

Robot-Assisted Surgery and New Uses of Imaging Devices

Technological advances in surgery using guidance techniques are reducing the types of surgery requiring the need for large incisions, and so are reducing the risks of

[10]I have omitted discussion of the fringe movement known as anti-aging medicine. It presumes to prolong life by administering selected dietary supplements and hormones. For a review of this topic, see National Academy on an Aging Society (2004).

surgical complications and the length of stay in the hospital. Minimally invasive surgery gains access to the body through natural orifices, the belly button, or small incisions. Other advantages include minimal blood loss, lower risk of infection, less scarring, and more rapid recovery time.

New research using guided imaging in cardiac surgery is intended to obviate the need to open the chest and stop the heart. One research protocol uses magnetic resonance imaging to guide the cardiac surgeon in operating on a heart requiring an aortic valve. Image-guided radiation therapy in the treatment of lung cancer gives radiologists a clearer, more accurate picture of the lung tumor, allowing a more precise targeting of the tumor and reducing the number of treatments needed for early-stage lung cancer.

Robot-assisted surgery has been introduced in cardiothoracic surgery, gynecological surgery, and urological surgery. It offers greater visualization of the operating site for the surgeon, makes possible greater dexterity and precision in the procedure, and allows a quicker recovery for the patient. A computer manipulates tiny "hands" holding surgical tools inserted in a small incision in the body. A camera also inserted through this incision provides a 3-D view of the site on a large monitor while the surgeon sits at a console controlling the movements of the robotic instruments. Robot-assisted surgery is now being used in the diagnosis and treatment of prostate cancer, ovarian, cervical, and uterine cancer, and kidney diseases.

Electronic Medical Record-Keeping and Bioinformatics

Electronic medical record-keeping is being introduced into medical practice but is still in its infancy. At the least, it would integrate the medical records of an individual so that any of the health-care providers attending an individual would have immediate access to them. At its best, it would integrate both the medical records on persons from all their health-care providers, including the medical records of all major health-care providers, such as the Mayo Clinic, Kaiser-Permanente, and the Veterans Administration, the research results obtained by such organizations as the National Health Interview Survey on self-reported health conditions and self-assessed health, and the findings in innumerable research publications. The new system would add to the safety of treatment by medication since it would reduce, if not eliminate, the risk of drug interactions. It would add to the effectiveness of medical treatment since physicians would have immediate, accurate, and complete information about the patient's history and the latest research findings on the patient's condition. Other alleged advantages of implementing electronic medical record-keeping are the reduction, if not elimination, of the ordering of duplicate medical tests and procedures and hence savings in the cost of medical care. Questions have been raised, however, whether there would be economies and whether privacy could be maintained.

Among the nascent fields of study that are supporting new developments in genetics is bioinformatics. Bioinformatics involves sorting through biological databases by use of the computer and providing data that could be used for genetic

research, genetic counseling, and the new developments in medical research. It does this by furnishing the tools needed to classify, track, and analyze the data developed by geneticists. Among these tools are microarrays, which are grids composed of bits of DNA that are complementary to the messenger RNAs the geneticist finds in a cell. They allow analysts to determine immediately which genes are at work in a particular cell and the strength of their activity. Microarrays are a valuable tool for gene discovery and mapping and are being used to screen thousands of genes rapidly in order to identify mutations that cause diseases. Nanotechnology, the technology used in the study of "supertiny" elements, can be used in conjunction with microarrays to detect diseased genes.

References and Suggested Readings

General Readings

Carey, J. R. (2003). *Longevity: The biology and demography of life span*. Princeton, NJ: PrincetonUniversity Press.

Carey, J. R., & Judge, D. S. (2001). Principles of biodemography with special reference to human longevity. *Population-E*, *13*, 9–40. Paris: INED.

Carey, J. R., & Tuljapurkur, S. (Eds.) (2003). *Life span*: *Evolutionary, ecological, and demographic perspectives*. A Supplement to Vol. 29, 2003, *Population and Development Review*. New York: Population Council. See esp. J.R. Carey, Life span: A conceptual overview, pp. 1–18.

Gavrilov, L., & Gavrilova, N. (1991). *The biology of life span: A quantitative approach*. New York: Taylor & Francis.

Gruman, G. J. (2003). *A history of ideas about the prolongation of life*. In cooperation with the International Longevity Center. New York: Springer.

Metchnikoff, I. I. (2004). *The prolongation of life: Optimistic studies* (P. C. Mitchell, trans.). New York: Springer.

Munnichs, J. M. A., Mussen, P., Olbrich, E., & Coleman, P. G. (Eds.). (1985). *Life-span and change in gerontological perspective*. New York: Academic.

Nesse, R. M., & Williams, G. C. (1994). *Why we get sick: The new science of Darwinian medicine*. New York: Times Books.

Olshansky, S. J., & Carnes, B. A. (2001). *The quest for immortality: Science at the frontiers of aging*. New York: Norton Press.

Robine, J. M. (Ed.). (2001). Biodemographic perspectives on human longevity. *Population-E, 13*(1). Paris: INED.

Siegel, J. S. (2005, January 12–14). The great debate on the outlook for human longevity: Exposition and critical evaluation of two divergent views. In Society of Actuaries (Ed.), *International symposium on living to 100 and beyond*, Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Society of Actuaries (Ed.). (2002, January 17–18). *Living to 100 and beyond: Survival at advanced ages.* Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Society of Actuaries (Ed.). (2005, January 12–14). *International symposium on living to 100 and beyond.,* Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Society of Actuaries (Ed.). (2008, January 7–9). *International symposium on living to 100 and beyond.,* Orlando, FL. CD. Schaumberg, IL: Society of Actuaries.

Wachter, K. W. & Finch, C. E. (Eds.). (1997). *Between zeus and the salmon: The biodemography of aging.* Washington, DC: National Academy Press.

Sources and Quality of the Data

Amundadottir, L. T., Thorvaldsson, S., Gudbjartsson, D. F., Sulem, P., et al. (2004). Cancer as a complex phenotype: Pattern of cancer distribution within and beyond the nuclear family. *PLOS Medicine,* 1, 229–236.

Coale, A. (1991, September). People over age 100: Fewer than we think. *Population Today*, pp. 6–8. Washington, DC: Population Reference Bureau.

Coale, A., & Kisker, E. E. (1990). Defects in data on old-age mortality in the United States: New procedures for calculating mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum, 4*(1, Spring), 2–31.

Gavrilova, N. S., & Gavrilov, L. A. (2005, January 12–14). Search for predictors of exceptional human longevity: Using computerized genealogies and internet resources for human longevity studies. In Society of Actuaries (Ed.), *International symposium on living to 100 and beyond,* Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Kannisto-Thatcher Database on Old-Age Mortality. Available at www.demogr.mpg.de.

Kestenbaum, B. M. (1992). A description of the extreme aged population based on improved Medicare enrollment data. *Demography, 29*(4), 565–580.

Kestenbaum, B. M., & Ferguson, B. R. (2005, January 12–14). Number of centenarians in the United States Jan. 1, 1990, Jan. 1, 2000, and Jan. 1, 2010, based on improved Medicare data. In Society of Actuaries (Ed.), *International symposium on living to 100 and beyond*, Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Leaf, A. (1982). Long-lived populations: Extreme old age. *Journal of American Geriatric Society*, *30*(485–487).

Meslé, F., & Vallin, J. (2002). Improving the accuracy of life tables for the oldest old: The case of France. *Population-E (Paris)*, *57*(4–5), 601–630.

Perls, T. T., Bochen, K., Freeman, M., Alpert, L., & Silver, M. H. (1999). Validity of reported age and centenarian prevalence in New England. *Age and Ageing*, *28*, 193–197.

Poon, L. W. (Ed.). (2008). *The georgia centenarian study: A study of longevity and survival of the oldest old, 1988–2007*. www.geron.uga.edu.

Population Reference Bureau. (2005). *World population data sheet*. Washington, DC: Population Reference Bureau.

U.S. National Center for Health Statistics (NCHS). (2002a). United States life tables, 2000. By E. Arias. *National Vital Statistics Reports, 51*(3). Hyattsville, MD: National Center for Health Statistics.

U.S. National Center for Health Statistics (NCHS). (2002b). Deaths: Final data for 2000. By A. M. Miniño, E. Arias, K. D. Kochanek, S. L. Murphy, & B. L. Smith (Eds.). *National Vital Statistics Reports*, *50*(15). Hyattsville, MD: National Center for Health Statistics.

U.S. National Center for Health Statistics. (2007). Deaths: Final data for 2004. By A. M. Miniño, M. P. Heron, S. L. Murphy, & K. D. Kochanek. *National Vital Statistics Reports*, *55*(19). Hyattsville, MD: National Center for Health Statistics.

U.S. Social Security Administration, Office of the Chief Actuary. (2002). Life tables for the United States social security area, 1900–2100. By F. C. Bell & M. L. Miller. *Actuarial study* No.116. Baltimore, MD: Office of the Chief Actuary, Social Security Administration.

U.S. Social Security Administration, Office of the Chief Actuary. (2005). Life tables for the United States social security area, 1900–2100. By F. C. Bell & M. L. Miller. *Actuarial Study* No.120. Baltimore, MD: Office of the Chief Actuary, Social Security Administration.

Vincent, P. (1951). La mortalité des vieillards, *Population*, *6*(2), 189–190. Paris: INED.

Willcox, D. C., Willcox, B. J., He, Q., Wang, N. C., & Suzaki, M. (2008). They really are that old: A validation study of centenarian prevalence in Okinawa, *Journal of Gerontology: Biological Sciences*, *63A*(4), 338–349.

Yue, C.-S. J. (2005, January 12–14). Data quality of oldest-old population in Taiwan: 2003 census for ages 89 and above. In Society of Actuaries (Ed.), *International symposium on living to 100 and beyond*, Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Demographic Measures and Trends

Barbi, E., Caselli, G., & Vallin, J. (2003). Trajectories of extreme survival in heterogeneous populations. *Population-E (Paris)*, *58*(1), 43–66.

Finch, C. E., & Pike, M. C. (1993). Maximum life span predictions from the Gompertz mortality model. *Journal of Gerontology: Biological Sciences*, *51A*(3), B183–194.

Horiuchi, S., & Wilmoth, J. R. (1997). Age patterns of the life table aging rate for major causes of death in Japan, 1951–1990. *Journal of Gerontology: Biological Sciences*, *52A*, B67–B77.

Horiuchi, S., & Wilmoth, J. R. (1998). Deceleration in the age pattern of mortality at older ages. *Demography*, *35*, 391–412.

Lee, R. D., & Carter, L. (1992). Modeling and forecasting U.S. mortality. *Journal of American Statistical Association*, *87*, (419): 659–671.

Martelin, T., Koskinen, S., & Valkonin, T. (1998). Sociodemographic mortality differences among the oldest old in Finland. *Journal of Gerontology: Social Sciences*, *53B*(2), S83–S90.

Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, *296*, 1029–1031.

Olshansky, S. J., & Carnes, B. A. (1997). Ever since Gompertz. *Demography*, *34*(1), 1–15.

Olshansky, S. J., Carnes, B. A., & Grahn, D. (1998). Confronting the boundaries of human longevity. *American Scientist*, *86*, 56–61.

Robine, J. M., & Saito, Y. (2003). Survival beyond 100: The case of Japan. In J. R. Carey & S. Tuljapurkur (Eds.), Life span: *Evolutionary, ecological, and demographic perspectives* (A Supplement to Vol. 29, pp. 208–228). 2003, *Population and Development Review*. New York: Population Council.

Robine, J. M., & Vaupel, J. W. (2002, January 17–18). Emergence of supercentenarians in the low mortality countries. In Society of Actuaries (Ed.), *Living to 100 and beyond*, Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Rogers, R. (1995). Socio-demographic characteristics of long-lived and healthy individuals. *Population Studies*, *21*(1), 33–58.

Stallard, E. (2002, January 17–18). Underlying and multiple-cause mortality at advanced ages: United States, 1980–1998. In Society of Actuaries (Ed.), *Living to 100 and beyond*, Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Vaupel, J. W. (1997). The average French baby may live 95 or 100 years. In J. M. Robine, J. W. Vaupel, B. Jeune, & M. Allard, et al. (Eds.), *Longevity: To the limits and beyond* (pp. 11–27). Berlin: Springer.

Vaupel, J. W. (2003). Post-Darwinian longevity. In J. R. Carey & S. Tuljapurkur (Eds.), *Life span: Evolutionary, ecological, and demographic perspectives* (A Supplement to Vol. 29, pp. 258–269). *Population and development review*. New York: Population Council.

Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., et al. (1998). Biodemographic trajectories of longevity. *Science*, *280*(5365), 855–860.

Wachter, K. (2003). Hazard curves and life span prospects. In J. R. Carey & S. Tuljapurkar (Eds.), *Life span: Evolutionary, ecological, and demographic perspectives* (A Supplement to Vol. 29, pp. 270–291). *Population and development review*. New York: Population Council.

Wilmoth, J. R. (1997). In search of limits. In K. W. Wachter & C. E. Finch (Eds.), *Between zeus and the salmon: The biodemography of longevity* (pp. 38–64). Washington, DC: National Academy Press/National Research Council.

Wilmoth, J. R. (1998). The future of human longevity: A demographer's perspective. *Science*, *280*, 395–397.

Wilmoth, J. R., & Robine, J. M. (2003). The world trend in maximum life span. In J. R. Carey & S. Tuljapurkar (Eds.), *Life span: Evolutionary, ecological, and demographic perspectives* (A Supplement to Vol. 29, pp. 239–257). *Population and development review*. New York: Population Council.

Biological, Evolutionary, and Reliability Perspectives on Aging

AARP. (2004, July–August). *AARP Bulletin*, p. 13.

Austad, S. N. (1997). Postreproductive survival. In K. W. Wachter & C. E. Finch (Eds.), *Between zeus and the salmon* (pp. 161–174). Washington, DC: National Academy Press.

Cappola, A. R., Xie, Q. L., Walston, J. D., Leng, S. X., et al. (2006). DHEAS levels and mortality in disabled older women: The women's health and aging study I. *Journal of Gerontology: Medical Sciences, 61A*(9), 957–962.

Carey, J. R. (1997). What demographers can learn from fruit fly actuarial models and biology. *Demography*, *34*(1), 17–30.

Carey, J. R. (2002). The importance of including biodemography in the demography curriculum. *Genus, 58*(3–4):189–199.

Carey, J. R., Liedo, P., Orozco, D., & Vaupel, J. W. (1992). Slowing of mortality rates at older ages in large medfly cohorts. *Science, 258*, 457–461.

Carnes, B. A., & Olshansky, S. J. (1993). Evolutionary perspectives on human senescence. *Population Development Review, 19*, 793–806.

Carnes, B. A., Olshansky, S. J., Gavrilov, L. A., & Gavrilova, N. S. (1999). Human longevity: Nature v. nurture – fact or fiction. *Perspectives in Biology and Medicine, 42*(3), 422–441.

Carnes, B. A., Olshansky, S. J., & Grahn, D. (2003). Biological evidence for limits to the duration of life. *Biogerontology, 4*(1), 41–35.

Carnes, B. A., Holden, L. R., Olshansky, S. J., Witten, T. M., & Siegel, J. S. (2006). Mortality partitions and their relevance to research on senescence. *Biogerontology*, *7*, 183–198.

Curtsinger, J. W., Hidenor, H. F., Townsend, D. R., & Vaupel, J. W. (1992). Demography of genotypes: Failure of the limited life-span paradigm in *Drosophila melanogaster*. *Science, 258*, 461–463.

Doblhammer, G. (2004, April 1–3). *Why do late parents live longer?*. Paper presented at the annual meeting of the Population Association of America, Boston, MA. Rostock, Germany: Max Planck Institute for Research.

Finch, C. E. (1990). *Longevity, senescence, and the genome*. Chicago: University of Chicago Press.

Finkel, D., Pedersen, N. L., Reynolds, C. A., Berg, S., et al. (2003). Genetic and environmental influences on decline in biobehavioral markers of aging. *Behavior Genetics, 33*(2), 107–123.

Finch., C. E., & Kirkwood, T. B. L. (2000). *Chance, development, and aging*. New York: Oxford University Press.

Frolkis, V. V., & Muradian, K. K. (1992). *Experimental life prolongation*. Boca Raton, FL: CRC Press.

Gavrilov, L., & Gavrilova, N. (2001). The reliability theory of aging and longevity. *Journal Theoretical Biology, 213*(4), 527–545.

Gavrilov, L., & Gavrilova, N. (2004, September). Why we fall apart. *IEEE Spectrum, 41*, 31–35.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies. *Philosophical Transactions of the Royal Society of London, 115*, 513–585.

Guarente, L., & Kenyon, C. (2000). Genetic pathways that regulate aging in model organisms. *Nature, 408*, 255–262.

Hall, S. S. (2003). *Merchants of immortality: Chasing the dream of human life extension*. New York: Houghton Mifflin.

Hayflick, L. (1998). How and why we age. *Experimental Gerontology, 33*, 639–653.

Hayflick, L. (2000). The future of aging. *Nature, 408*, 267–269.

Hayflick, L. (2004). From here to immortality. *Public Policy and Aging Report, 14*(2), 1, 3–7. National Academy on a Aging Society.

Holliday, R. (2007). *Aging: The paradox of life. Why we age*. Dordrecht, the Netherlands: Springer.

Judge, D. S., & Carey, J. R. (2000). Postreproductive life predicted by primate patterns. *Journal of Gerontology: Biological Sciences, 55A*, B201–B209.

Kirkwood, T., & Rose, M. (1991). Evolution of senescence: late survival sacrificed for reproduction. *Philosophical Transactions of the Royal Society of Medicine, 332*, 15–24.

Masoro, E. J. (2000). Don't kill the messenger – learn to evaluate the message. *Gerontologist, 40*(2), 240–242. Review of books on the biological basis of aging and the causes of aging.

National Academy on an Aging Society. (2004). *The pros and cons of anti-aging medicine*. Public Policy and Aging Report. Papers by L. Hayflick, R. H. Binstock, C. Haber, E.T. Juengst, & R. Ponsaran, T.E. Johnson, and S. G. Post. Washington, DC: National Academy on an Aging Society.

Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., et al. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine, 352*(11), 1138–1145.

Olshansky, S. J., Carnes, B. A., & Cassel, C. (1990). In search of Methusaleh: Estimating the upper limits to human longevity. *Science, 250*, 634–640.

Olshansky, S. J., Carnes, B. A., & Désquelles, A. (2001). Still in search of Methuselah: Prospects for human longevity in an aging world. *Science, 291*, 1491–1492.

Olshansky, S. J., Carnes, B. A., & Grahn, D. (1998). Confronting the boundaries of human longevity. *American Scientist, 86*(1), 52–61.

Racette, S. B., Weiss, E. P., Villareal, D. T., Arif, H., et al. (2006). One year of caloric restriction in humans: Feasibility and effects on body composition and abdominal adipose tissue. *Journal of Gerontology: Medical Sciences, 61A(9)*, 943–950.

Siegel, J. S. (2002). *Applied demography: Applications to business, government, law, and public policy*. San Diego: Academic.

Sinclair, D. (2006, November 16–20). *Longevity drugs: Myth and reality*. Paper presented at annual meeting of the Gerontological Society of America, Dallas, TX.

Strehler, B. I. (1979). Aging research: Current and future. *Journal of Investigative Dermatology, 73*, 2–7.

U.S. National Institute on Aging. (2006). *Aging under the microscope: A biological quest*. Bethesda, MD: U.S. National Institute on Aging.

Vetta, A., & Courgeau, D. (2003). Demographic behaviour and behaviour genetics. *Population-E, 58*(4–5):401–428. Paris: INED.

Walford, R. L., Harris, S. B., & Gunion, M. W. (1992). The calorically restricted low-fat nutrient-dense diet in Biosphere 2 significantly lowers blood glucose, total leucocyte count, cholesterol, and blood pressure in humans. *Proceedings of the National Academy of Sciences, 89*, 11533–11537.

Centenarians and Superlongevity

Anderson-Ranberg, K., Lone, V., & Jeune, B. (2001). Dementia is not inevitable: A population-based study of Danish centenarians. *Journal of Gerontology: Psychological Sciences, 56B*(3), P152–P159.

Baltes, P. B. (2002, March 31–April 4). Paper presented at the Advanced Technical Conference on Aging, Valencia, Spain. Precursor to the United Nations World Assembly on Aging, Madrid, Spain, April 7–11.

Barzilai, N., Gabriely, I., Gabriely, M., et al. (2001). Offspring of centenarians have a favorable lipid profile. *Journal American Geriatrics Society, 49*, 76–79.

Calvert, J. F., Hollander-Rodriguez, J., Kaye, J., & Leahy, M. (2006). Dementia-free survival among centenarians: An evidence-based review. *Journal of Gerontology: Medical Sciences, 61A*(9), 951–956.

Hagberg, B., Alfredson, B. B., Poon, L. W., & Homma, A. (2001). Cognitive functioning in centenarians: A coordinated analysis of results from three countries. *Journal of Gerontology: Psychological Sciences, 56B*(3), P141–P151.

Ljungqvist, B., Berg, S., Lanke, J., McClearn, G., & Pedersen, N. L. (1998). The effect of genetic factors for longevity: A comparison of identical and fraternal twins in the Swedish Twin Registry. *Journal of Gerontology*: *Medical Sciences, 56A*, M441–M446.

McGue, M., Vaupel, J. W., Holm, N., & Harvald, B. (1993). Longevity is moderately heritable in a sample of Danish twins born 1870–1880. *Journal of Gerontology, 48*, B237–B244.

Mueller, U. (2004). Does late reproduction extend the life span? Findings from European royalty. *Population and Development Review, 30*(3), 449–465.

Perls, T., & Terry, D. (2003). Understanding the determinants of exceptional longevity. *Annals of Internal Medicine, 139*, 445–449.

Perls, T. T., Albert, L., & Fretts, R. C. (1997). Middle-aged mothers live longer. *Nature, 389*, 133.

Perls, T. T., Wilmoth, J., Levinson, R., Drinkwater, M., et al. (2002). Life-long sustained mortality advantage of siblings of centenarians. *Proceedings. of National Academy of Sciences, 99*(12), 8442–8447.

Poulain, M., Pes, G. M., Grasland, C., Carru, C., et al. (2004). Identification of a geographic area characterized by extreme longevity in the Sardinia Island: the AKEA study. *Experimental Gerontology, 39*, 1423–1429.

Robine, J. M., & Vaupel, J. W. (2002, January 17–18). Emergence of supercentenarians in the low mortality countries. In Society of Actuaries (Ed.), *Living to 100 and beyond*, Lake Buena Vista, FL. CD. Schaumberg, IL: Society of Actuaries.

Silver, M. H., Jilinskaia, E., & Perls, T. T. (2001). Cognitive functional status of age-confirmed centenarians in a population-based study. *Journal of Gerontology*: *Psychological Sciences, 56B*(3), P134–P140.

Smith, D. W. E. (1997). Centenarians: Human longevity outliers. *Gerontologist*, *37*(2), 200–206.

Terry, D. F., Wilcox, M. A., McCormick, M. A., & Perls, T. T. (2004). Cardiovascular disease delay in centenarian offspring. *Journal of Gerontology: Medical Sciences, 59A*, M385–M389.

U.S. Census Bureau. (1999). Centenarians in the United States. In C. A. Krach & V. A. Velkoff (Eds.), *Current population reports*. Series P-23-199RV. Washington, DC: U.S. Government Printing Office.

Wilmoth., J. R., Deegan, L. J., Lundström, H., & Horiuchi, S. (2000). Increase of maximum life span in Sweden, 1861–1999. *Science*, *289*, 2366–2368.

Wilmoth, J., Skytthe, A., Friou, D., & Jeune, B. (1996). The oldest man ever? A case study of exceptional longevity. *Gerontologist*, *36*(6), 783–788.

Yashin, A. I., & Iachine, I. A. (1997). How frailty models can be used for evaluating longevity limits: Taking advantage of an interdisciplinary approach. *Demography*, *34*(1), 31–48.

Chapter 14
Models of Aging, Health, and Mortality, and Mortality/Health Projections

Models of Aging, Health, and Mortality

Reprise on Causation of Health Outcomes

As we saw, the life span of individuals is determined by a complex interplay of evolutionary, genetic, epigenetic, social/individual/ecological, and stochastic factors operating roughly within the limits of the genetic predispositions of these individuals. The stochastic factor represents both a residual category that jointly encompasses those influences that cannot be assigned to the other categories and those influences that occur by chance. Chance events are a major influence in the determination of the life span and, like evolutionary and genetic influences, may be considered at least in part beyond society's or the individual's control. Pending the personalization in health planning of the genomic characteristics of each individual and developments in gene therapy, therefore, the focus should be on social/personal/ecological factors as areas where major changes in health status can now be largely effected.

The question as to which factor is more important, genetics/epigenetics, "environment," or chance in the complex interplay I have referred to is relevant only with respect to a fourth element, namely health outcomes. (Just as the question, which is more important in a rectangle, the length or the width, has no meaningful answer except in relation to the some use of the rectangle.) These factors interact in multiple ways and at multiple levels to modify sharply the effect of the other in influencing different health outcomes. To quote George Martin (Unpublished Conversation with the Gerontological Society of American, Spring 2003), distinguished American biogerontologist, "...gene action during early development and puberty can set the stage for differential rates of the emergence of senescent phenotypes...environmental influences play a powerful role in gene expression. They can vary from the mother's nutritional status during her pregnancy to parental neglect and psychological abuse during early childhood."

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_14, © Springer Science+Business Media B.V. 2012

Concepts of Individual Aging Trajectories

It is canon among biogerontologists to distinguish aging from disease. Although most diseases are related to the passage of time – and hence they are known as age-related or age-associated diseases, most biogerontologists do not consider aging as a disease. They are quick to point out that, while the molecular and biological changes that occur at the level of cells, tissues, and organs during the course of life can and often do make people more susceptible to the expression of disease, this is not always the case. In fact, a small number of people who live to extreme old age appear to be largely free from heart disease, Alzheimer's disease, and other major chronic degenerative diseases, even though their bodies and minds have aged considerably since they were born.

If, as this view holds, aging is not a disease, then it could be inferred that normal aging is nonpathological. This inference has to be reconciled with the fact that nonpathological aging is a rare occurrence. Nearly all aged persons have multiple chronic morbidities that have accumulated with advancing age. Hence, it is unclear how normal aging can be synonymous with nonpathological aging. Just as normal aging is associated with increasing morbidity, biological aging emphasizes the increasing vulnerability to disease and death that occurs with the passage of chronological time and with increasing age. When attention is directed to the progressive dysregulation of body processes and the decline in the integrity of the cells, tissues, and organs of the body with increasing age, the term senescence is applied. The damage characterizing senescence begins early in life and is the inevitable product of the biochemistry of life itself. Senescence results in the gradual disruption of bodily homeostasis, i.e., the equilibrium between the body's systems, and finally death.

These biological concepts of aging are troubling to some analysts because they define aging entirely in negative terms. A non-biological definition of aging proposed as an antidote to this approach is that aging represents the changes, both positive and negative, that individuals experience with advancing age. This concept allows for the sense of freedom from responsibility, and gains in wisdom, judgment, and other positive traits that individuals experience as they grow older. Many gerontologists have chosen to describe and popularize idealistic and inspirational concepts of aging that emphasize targets or goals rather than the facts of aging. Several terms have been used to describe these forms of aging, such as healthy, productive, active, positive, functional, and successful. When aging is viewed in these terms, there is an activist orientation to the description setting forth a goal designed to blur and overshadow the negative consequences of aging processes as we know them. The most euphemistic of these expressions of aging – the one with the most grandiose inspirational quality – is "successful" aging. Most people experience similar but not identical changes in their bodies and minds as they grow older, and these changes do not conform to these idealistic descriptions. The processes of aging that most people experience often have a generally negative quality, and for good reason, particularly with respect to health. That is why the identification of a healthy aging phenotype is often held up as the gold standard for successful aging.

Normal Aging and Pathological Aging

The concept, normal aging, can be used in three ways. It can be used to refer to the increasing vulnerability to disease and death that accompanies advancing age, "average" health for one's age, and a standard of health for one's age. The first definition corresponds to what has been called biological aging. The expression is, however, more commonly interpreted to mean average health for one's age. The third definition is a normative, goal-oriented use of the concept. These three definitions are distinct, even antithetical, and so the expression, normal aging, is ambiguous. Since nearly all elderly persons have multiple chronic diseases or disorders, normal aging in the sense of "typical" implies, not only the aging changes that are not pathological, but also those that are due to disease. Normal aging implies to most physicians and laypersons about the same state of health as is typical of other persons of the same age. One hears this use when the physician tells a patient that his or her condition is "about normal for your age." This report should not usually be taken as good news by an elderly patient.

As noted, to emphasize that aging is not a disease, some biogerontologists define the terms normal aging and biological aging to mean increasing *vulnerability* to disease. However, both normal aging and biological aging, especially the latter, are also sometimes used to refer to the accumulated damage to the building blocks of life (e.g., proteins, fats, carbohydrates, DNA, etc.) that is called senescence. In keeping with the latter concepts, biological aging and senescence have also been described as the "result of failure of homeostatic systems, resulting in an increasing probability of death" (Blumenthal 2003). These terms emphasize the fact that health is a dominant concern of older persons and its loss or the reduced integrity of the body in later life mainly defines the condition of older people.

Some claim that aging and disease are distinct and dichotomous states (Hayflick 2001; Strehler 1977; Butler 1977; Williams 1992). They are, in effect, drawing a sharp line between biological or physiological aging and pathological aging. Others claim that the bounds are fuzzy and that the states merge into one another in a continuum (Kleemeier 1965; Ludwig 1980; Rattan 1991). Still others have ambiguous views of the matter. According to views on this issue compiled by Blumenthal (2003), the separation of biological aging from pathological aging is based on such principles as that diseases during the lifetime of an individual have the capacity "of adding injury to the insult of biological aging" and that "…the diseases with a high prevalence in the aged population are the consequence of the aging-linked weakening of critical biological defense mechanisms, which then permit the effects of extrinsic agents to take hold." Aging is described as universal and unavoidable physiological decline whereas diseases occur in some persons and not in others, are associated with specific etiological risk factors, and hence are preventable.

Views on the opposite side are expressed in the statements that physiological and pathological aging are "…so interrelated as to make attempts [to separate them] relatively abortive. It would be far more relevant to accept the existence of a continuum of aging phenomena" (Hall 1984). "…the distinction between so-called

natural aging and the pathologies that are common in old people is artificial" (Holliday 1988). Underlying both aging and disease are progressive dysfunctionality of intracellular parts and biological processes such as inflammation, oxidative stress, and defective repair of gene mutations. An ambiguous position is expressed in the following statement: "In the future, studies of aging must better attempt to capture the interplay between diseases and aging phenomena" (Fozzard et al. 1990).

The difficulty of drawing a sharp line between aging and disease has been graphically illustrated by Ling et al. (2007) with reference to a major chronic disease of later life, osteoarthritis. The dichotomy between aging and disease is appropriately tested with osteoarthritis because advanced age is the strongest risk factor for the development of this disabling disease. Ling et al. report that highly sophisticated imaging devices are beginning to define the anatomic boundaries of the gradual shift from a normally aging joint to early osteoarthritis to full-blown pathology.

The distinction between aging and disease is an important theoretical and heuristic one but in practical terms the distinction fades away. Often with increasing age, a health condition that once fell within acceptable bounds as a risk factor (i.e., physiological aging) or even as a disease entity may have deteriorated and now merits a classification as disease (i.e., pathological aging). Moreover, with the passage of time, as medical knowledge increases, the nature of a condition may be reevaluated and the condition may be reclassified from risk factor to disease or vice versa. The history of obesity or high blood pressure may be illustrative of this change in diagnostic roles. My own view is closer then to the "continuum" position.

Normative aging. Another meaning of the expression normal aging is as a "standard," that is, a standard of health at a given age to be met or emulated (*cf.*, a norm), and to be compared with the actual findings for the person or the age group. A set of health conditions needs to be selected to serve as the standard for judging the health condition of the person in question. Normal aging in this sense implies the absence of the pathologies typical of later life and possession of a state of health typical of an earlier more robust age. The "acceptable" ranges listed on medical reports for the results of laboratory tests do not generally qualify for this purpose; they are designed to reflect the ranges usually found at the particular age. This third concept of aging may better be labeled normative aging.

To measure changes in health with advancing age, indicators called biomarkers, measuring biological changes that characterize the aging process, are quantitatively assessed. For a selected list of biomarkers, see Exhibit 5.2. There are numerous biomarkers, but none can be confidently applied to a particular individual as a basis of judging his or her biological age or expected duration of life, or to a particular group as a basis of judging its life expectancy. A package of these, however, could be measured for an individual and the results compared with the corresponding values for the population in some younger age group. A considerable body of scientific research on designing theories of health and predicting health outcomes is based on scores for groups of biomarkers or, as they are termed collectively when cumulated over the life course, allostatic load. Although a single biomarker

is an undependable and volatile measure, scores for combinations of them could be averaged (equally or weighted) for an individual or a group of individuals at a given age group, for comparison with and evaluation against the corresponding combined score for a younger age group. In this way a relatively risk-prone adult group could be compared with a relatively risk-free adult group, in order to measure their relative state of health and longevity prospects.

Multitudinous discomforts and false alarms. Even in the absence of serious illness, aging is invariably accompanied by a multitude of physical and mental discomforts that place restraints on activities during the later years or make it uncomfortable to perform them. These are not commonly considered pathological, however, unless they fall outside certain wide boundaries and result in limitations of functioning. Moreover, aging is often accompanied by "false alarms," apparent symptoms of serious life-threatening illnesses that, happily, turn out to be negative or inappropriately diagnosed. The discomforts and frequent "symptoms of older age" include, but are not limited to, reduced visual acuity, hearing loss, dental problems, skin lesions, intermittent backaches, debilitated knees, inability to stand or walk comfortably more than briefly, other similar musculoskeletal problems, gastrointestinal irregularities, easy fatigue, inability to concentrate, headaches, weakening grip, balance problems, short-term forgetfulness, nocturia, dysuria, irritable bladder syndrome, hemorrhoids, insomnia, other sleep disorders, increased tendency to catch infectious diseases or sustain small injuries, and slower recovery from injuries and infections. The points at which these discomforts and symptoms become identifiable pathologies, e.g., benign prostatic hypertrophy, irritable bowel syndrome, ocular migraine, depression, osteoarthritis, hypertension, bradycardia, etc., are arbitrary and ill-defined. The pathologies may even precede the discomforts and symptoms. In themselves the discomforts limit effective activity below desired levels, reduce productive time and productivity, and force preoccupation with health matters on the part of many older persons.

Successful Aging

Among the goal-oriented, idealistic, and inspirational concepts of aging, the concept that has grabbed the imagination of more gerontologists than the others is that of "successful aging." Hence, I select this one for principal discussion. Rowe and Kahn (1987, 1998), who popularized the concept of successful aging, define it as the ability to maintain three keys behaviors or characteristics throughout most of the life span: Low risk of disease and disease-related disability, high mental and physical functioning, and active engagement with life. They consider each factor important in itself and to some extent independent of the others, but the combination of all three factors represents the concept of successful aging most fully. Rowe and Kahn maintain that successful aging is largely determined by individual lifestyle choices in diet, exercise, the pursuit of mental challenges, self-efficacy, and involvement with other people, not by genetic inheritance. The authors claim that, even though

many health conditions are affected by genetic influences, the effects of the genes can be modified by appropriate lifestyle changes and medical interventions.

The concept of successful aging has been widely adopted. Numerous books and research projects have built on and extended or modified the concept (e.g., Morrow-Howell et al. 2001; Callahan 2003). For example, Inui (2003) poses a definition following the humanist perspective that does not require the absence of disease or disability, but emphasizes the "preservation of key capacities to perform in domains that are important to the individual in his or her niche."

On the other hand, the validity and utility of the concept have also been challenged. Some see it merely as a public-relations ploy or inspirational slogan designed to mobilize people into action, to make them feel responsible for and to take charge of their own health, and to challenge them to pursue a healthy lifestyle. This view of the concept is not concerned with its validity, only its effectiveness as a "political" tool. Some authors, particularly Masoro (2001), have pointed out some of the basic limitations of the concept. His main argument is that few people reach advanced age without significant age-related diseases and that most people will reach these ages with numerous comorbidities and appreciable physiological deterioration. Even the "successful agers" are often plagued by a host of health conditions that greatly limit the range of their activities in their later years, or they endure various serious health conditions that are asymptomatic (i.e., not known to the person and not yet seriously affecting their functioning) until the conditions plunge them into a morbid, disabled, or dependent state. Masoro also notes that the concept of successful aging understates the role of genetics and chance in their effect on late life health and functionality.

"Successful aging" makes an unfortunate subjective evaluation of an individual's efforts at success or failure in achieving a long life, as if his or her efforts are wholly responsible for reaching an advanced age. One is tempted to ask, is living a brief life with a healthy life history, or living to an advanced age but in a disabled or dependent state, to be labeled unsuccessful aging? A more realistic interpretation of aging acknowledges that attaining a long and healthy life is as much a matter of genetics and luck as individual mastery over a healthy lifestyle. One's genetic endowment combined with environmental and stochastic forces largely beyond one's control bear much responsibility for one's longevity. Among the environmental factors exerting a strong influence on individuals' adult health is the socioeconomic status – education, income, and occupation – of their parents during the individuals' formative years and the latters' health record in childhood and adolescence. In sum, individuals have only limited control over their health in later life and the phenotypic expression of their genetic and familial endowment during the course of their lives.

An operational definition of the concept of successful aging is needed if we are to be able to classify individuals as successful/unsuccessful agers and to measure the growth of the successfully aging population. Tate et al. (2003) have tried to arrive at a definition of successful aging on the basis of the replies of elderly male respondents to a question on the concept in the Manitoba (Canada) Follow-up Study. The most common response defined the concept in terms of health and freedom from

disease, and the next most common response was "happiness, enjoying life, having a satisfying lifestyle." Tate et al. caution us that an inherent problem with defining successful aging in this way is that success is a subjective judgment. Success implies different standards for different people and hence it can carry vastly dissimilar meanings. Successful aging is dependent on the value system of the individual, and that value system is molded by the value system of the society in which the individual lives. We really need a more objective definition of the concept or ought to discard it.

Other Concepts of Aging

As noted, other expressions, such as positive aging, productive aging, active aging, healthy aging, functional aging, and aging well, have been used as synonymous with, or alternatives to, successful aging. Although all these expressions have been widely employed in gerontological writing, I suggest that a single concept of this kind should be selected for the scientific lexicon of gerontology, be given formal recognition, and be developed operationally for statistical and analytic purposes. For this purpose, I prefer "healthy aging," although this expression itself is an oxymoron of a sort. The definition of health could follow the WHO definition and encompass physical, mental, and social health. The term healthy aging avoids the judgmental connotation of some of the other expressions and can be given an operational definition based on classifying individuals' health condition according to specified criteria.

Concepts of Mortality/Morbidity Trajectories

Compression of Morbidity and Successfully Aging Populations

There is greater merit in the concept of a successfully aging population than in the concept of a successfully aging individual. We may demographically define a successfully aging population as a population with a relatively limited period of late-life illness. The period of chronic morbidity in later life may be reduced if life expectancy at birth rises more slowly than the age of onset of chronic morbidity, or if life expectancy at birth is stable while the ages of onset of the chronic late-life diseases are delayed to later and later ages, approaching the unchanging life expectancy. Under either scenario late-life morbidity is compressed into a shorter and shorter period.

Such scenarios are variations of the compression-of-morbidity model first proposed by Fries (1980, 1987, 1989). It is reasonable to hypothesize that a rise in life expectancy would be accompanied by a rise in the age of onset of late-life morbidity, at least until life expectancy reaches a presumed maximum. Fries believed that there was such a maximum and considered it to be 85 years. Some analysts were

quick to criticize the Fries theory of mortality compression as being unrealistic and perhaps Panglossian (Schneider and Brody, 1983; Schneider and Guralnik, 1987). Given the life expectancies already achieved by various national populations and the intense biomedical research being pursued in laboratories throughout the world, Fries' maximum age now appears low for the longevity prospects of many countries. (See discussion in Chap. 13.) Moreover, mortality rates, still quite high at the ages over 65 in the first decade of the twenty-first century, continue to fall at these ages, so that life expectancy at birth and life expectancy at the higher ages continue to rise. A rising life expectancy at birth may itself be viewed as "converging" to the average maximum life span (i.e., average age of life-table survivors aged 100 or more) since the former has been rising more rapidly than the latter.

Evidence that has become available since Fries first published his theory appears to confirm an important part of it, at least for the United States. Disability ratios as reported by the U.S. National Long-Term Care Survey declined during the 1980s and 1990s, and the age of onset of disability appears to have risen more rapidly than life expectancy at birth in this period (Fries 2003). The finding of morbidity compression is not universal, however, according to a OECD study that reviewed trends in ADL disability at age 65 and over in 12 OECD countries during the 1990s (Robine et al. 2008a). This study provided clear evidence of a decline in disability among elderly persons in only five of the twelve countries studied (including the United States). Furthermore, the decline in ADL disability ratios of older Americans in the United States has not continued into the first decade of this century (Fuller-Thomson et al. 2009).

The situation is more complex than reflected in changes in levels of ADL disability. A decline in the level of ADL disability does not necessarily mean a compression of morbidity. In fact, an expansion of morbidity may accompany a decline in disability if the individuals saved from disability suffer from various serious chronic diseases whose ages of onset have remained stable. It is not known generally whether or not the ages of onset of such diseases changed during the 1980s and 1990s although this is a researchable question. There is also a question of the meaning and comparability of data on a decline in disability. The decline may not signify improved health; it may signify, instead, increased functional independence resulting from technological and environmental changes.

Compression of Mortality and Rectangularization of the Survival Curve

The morbidity-compression model as originally proposed by Fries was accompanied by the concept of mortality compression. Linking the paths of survival and mortality, we can formally measure mortality compression in terms of the variation in the distribution of ages at death around their mean or median age or as the tendency of the survival curve to become more and more rectangular in shape at the higher ages. In the Fries' theoretical construct, mortality compression involves

a closing of the gap between current life expectancy at birth and the assumed maximum life expectancy of about 85. This shift implies that the survival curve is becoming more and more rectangular in shape. In simple terms, more and more people are expected to live to an advanced old age approximating age 85, and then die within a narrow range of ages on either side of that age (Fries 1980; Fries and Crapo 1981).

As indicated in Chap. 4, rectangularization-of-the-survival-curve/mortality-compression can be measured and identified in several ways. Among them are a reduction in the interquartile survival range for deaths, a reduction in the standard deviation of deaths around the mean age of deaths, an increase in the (negative) slope of the survival curve after the mean age of death, and a rise in the ratio of life expectancy at birth to total life expectancy at age 100. The first measure, the interquartile range, decreased from about 65 years in 1751 to 15 years in 1998 in Sweden (Wilmoth and Horiuchi 1999). Most of the decline occurred between the 1850s and 1950s. Since then, variability in age at death has been nearly constant in Sweden, defying predictions of a continuing rectangularization in that country. Compression of mortality also appeared in historical data for several European countries, but not in Japan (Robine et al. 2008b). The United States has experienced a steady compression of mortality through the last century although it is characterized by a high degree of variability in age at death compared with both Sweden and Japan (Wilmoth and Horiuchi 1999). I (Siegel 2005) found a century-long trend toward compression of mortality in the United States on the basis of the relative interquartile range (Table 4.8). Although this trend continued to the very end of the century, the pace of the compression has slowed in the more recent decades.

Other researchers came out with different results. Rothenberg et al. (1991) examined the coefficient of variation around the mean age of deaths in the United States from 1962 to 1984. They found that, as the mean age at death increased and the force of mortality decreased over time, there was a relative increase in the variability of age at death among those of advanced age (85 and over). Myers and Manton (1984) came out with the same conclusion using similar data and measures.

An important factor influencing the compression of mortality is the measured or assumed increase in longevity at the highest ages. Changes in mortality rates at ages 95 and over or 100 and over affect the distribution of deaths and hence the degree of mortality compression, even though the share of deaths at these ages among all deaths or even adult deaths is small and the data are tainted by serious reporting errors. The various measures of the variability of the age of deaths should take account of deaths at these advanced ages because, if average maximum life span and life expectation at age 95 or 100 were to rise sharply, the survival curve would be pulled to the right and rectangularization would tend to be slowed or even reversed. Accordingly, a measure of mortality compression that takes more direct account of mortality changes at the extreme ages of life seems useful. The ratio of

the life expectancy at birth to the total life expectancy at age 100 (per 100) is such a measure:

$$\frac{e_0 + 0}{e_0 + 100} * 100 \tag{14.1}$$

Estimates of mortality compression using this measure for the United States over the last century and the next (for females) are as follows:

Ratio per 100	
1900	48.96 ÷ 101.61 = 48.2
1950	71.13 ÷ 101.92 = 69.8
1980	77.52 ÷ 102.42 = 75.7
2000	79.39 ÷ 102.26 = 77.6
2050	83.22 ÷ 102.77 = 81.0
2100	86.40 ÷ 103.57 = 83.4

Source of basic life table data: U.S. Office of the Actuary, *Actuarial Study* No. 120, 2005.

Both life expectancy at birth and total life expectancy at age 100 rise, but life expectancy at birth rises much more rapidly. As a result, the series rises. This trajectory indicates that the survival curve has been becoming and continues to become more rectangular. It would take an extremely large increase in total life expectancy at age 100 to produce a turnaround in the measured compression of mortality. For example, even if total life expectation at age 100 were to rise to 105 in 2050 and to 108 in 2100, the percentage would not turn around: (84 ÷ 105)*100 = 80 for the year 2050 and (88 ÷ 108)*100 = 81 for the year 2100.

Expansion of Morbidity and Frailty

Life expectancy and the period of late-life morbidity may move up at the same rate, maintaining a relatively fixed relation to one another. Under this scenario the period of morbidity is neither compressed nor expanded but remains relatively unchanged. This is a realistic possibility, given appropriate lifestyle modifications and/or medical developments. In a more problematic course of events, survival is extended because of the influences mentioned but the ages of onset of serious chronic illness, disability, or frailty do not change. In such a scenario the duration of late-life morbidity increases, growing at a faster pace than the rise in life expectancy. The period of chronic morbidity is not compressed but expands. The population lives longer but is "sicker" for a longer period (Gruenberg 1977; Verbrugge 1984; Olshansky et al. 1991). It is apparent that simply improving the management of the manifestations of aging increases the risk of the occurrence of a scenario in which both the period of old age and the period of disability and frailty lengthen.

As we saw, frailty denotes a condition of general physical weakness, including difficulty in moving about and performing tasks alone. Frail persons are disabled because of a multitude of age-related health conditions rather than just one or two.

An expansion of the period of morbidity is likely to mean a large increase in the number of frail elderly in the population, requiring the time and commitment of numerous family members and an expansion of long-term care facilities.

So far we have assumed that life expectancy continues on its historical upward trajectory, but this course is not inevitable and there are signs that it may not continue (Olshansky et al. 2005). Hence, the range of possible scenarios needs to be extended to include those in which life expectancy stalls or even reverses direction. Under such circumstances, whether morbidity is compressed or expands depends on the course of the major chronic diseases. If life is shorter, the period of sickness may be shorter as well. However, the more likely scenario is that the period of sickness will be expanded. Inasmuch as the leading probable cause of a prospective reversal of life expectancy is an increase in the incidence of diabetes morbidity, the diabetes death rate, and the associated conditions (including kidney disease, cardiovascular diseases, neuropathy), the average age of onset of these diseases may be expected to fall and an expansion of morbidity from these diseases may be expected to occur. Even if the rise in life expectancy does not halt, given the outlook for diabetes and several other diseases for which there are no cures at this time, the period of morbidity may widen as we go into the future.

Mortality and Health Projection Models

Different researchers project quite different levels and patterns of mortality for the next several decades. Some see life expectancy as reaching an upper limit at age 85 (Olshansky et al. 1990; Fries 1980, 1987) and others project it to 100 by the third or fourth quarter of this century (Oeppen and Vaupel 2002; Wilmoth 2000; Manton and Woodbury, 1992).Some evolutionary and molecular biologists, such as C.Kenyon, L. Guarante, and M. Rose, have more expansive expectations (Guarente and Kenyon 2000). While their research has primarily dealt with small invertebrate species such as fruit flies and nematode worms, they anticipate maximum average life spans of 200 years or more for humans in the next century (Hall 2003). An even more extreme view is held by A. de Grey et al. (2002), the British geneticist and associates, whose SENS (Strategies for Engineered Negligible Senescence) program calls for implementing a set of biomedical and biogerontological activities that could arguably extend human longevity by biblical proportions. The SENS program seeks to introduce a series of rejuvenation therapies sufficient to restore older laboratory animals to middle age and, in due course, humans, to the physical and mental robustness of young adults. The rejuvenation therapies include combating oxidative stress and mitochondrial mutation, rejuvenating the immune system, tissue engineering, gene therapy, organ and tissue regeneration (e.g., neuroregeneration), and DNA repair.

The “conservative” or “centrist” projections of the older U.S. population and the “optimistic” projections are far apart, but even the “centrist” positions are not close. For example, for the population 65 years and over in 2050, the Census

Bureau (2009) projected 88.5 million in its middle series, the Social Security Administration (2008) projected 80.8 million (including overseas covered workers) in its intermediate series, and Ahlburg and Vaupel (1990) projected 92 million, For the population 85 years and over in 2050, the Census Bureau (2008) gave 19.0 million, the Social Security Administration offered 15.1 million as its intermediate figure, and Ahlburg and Vaupel (1992) predicted 39 million. On the other hand, Manton and Woodbury (1992) projected over 135 million for the population 65 years and over in 2050 and 54 million for the population 85 years and over in 2040.

Mortality Projections

Mortality projections are commonly prepared in connection with the generation of population projections as a component of population change, along with the other components, fertility and migration. This is the context for the mortality projections prepared by the U.S. Census Bureau and the Actuary's Office, U.S. Social Security Administration. These population projections are used by numerous agencies in planning public programs (e.g., hospital or road construction, social security planning). Mortality projections may also be generated in their own right, for example, for public health planning or planning the expansion of the funeral industry, or, much less commonly, as an intermediate step in preparing projections of the health status of the population. Many different methods, assumptions, and data have been used to project mortality. The following sections describe the principal methods and their variants, particularly where mortality projections are prepared as a component of change in preparing population projections.

Methods of Mortality Projections[1]

The methods for projecting mortality may be roughly classified as demographic, actuarial, statistical, and epidemiological. Just as the scopes of these disciplines overlap, a particular projection design may contain elements of more than one of these methods. Hence, the methods are not always distinguishable in terms of the above captions, are not mutually exclusive, and may be combined in various ways, as will become evident from the descriptions given. Most methods are designed to project mortality as an independent variable, but some are designed to project mortality as an extension of morbidity projections. In this discussion I give principal attention to demographic methods and their variants since they are employed most commonly. The method selected by an analyst in an actual application tends to vary according to the academic specialty of the analyst.

[1]It is not practical to describe the methods of preparing mortality and population projections in detail here. The reader is referred to other publications in which these methods are discussed more fully. See, for example, Siegel and Swanson (2004); Ahlburg and Land (1992); Smith et al. (2001); and Siegel (2002).

Demographic methods. In the demographic model, the mortality projections are typically made within the framework of a cohort-component method of making population projections. The component method calls for combining births and immigration with the base population and removing deaths and emigration. The cohort-component method carries out the calculations for the population and the components with data disaggregated by age and sex. Projections of mortality, fertility, and migration are applied serially to the prior age/sex distribution of the population, to derive the population projections for age and sex groups at each subsequent date. More specifically, the base or current population, disaggregated by age and sex, is carried forward by projected age-specific maternal birth rates, age-sex-specific survival rates, and amounts or rates of net migration for age-sex groups, to generate population projections for age-sex groups at the next date.

A deterministic model has been commonly employed in making the projections. In general terms, a deterministic projection model is a model for projecting possible outcomes with specified assumptions for certain variables, in which any specified set of assumptions fully determines a single outcome for the variables. To allow for the uncertainty of a single projection, called a middle or intermediate forecast and viewed as most likely, two other series, designed to cover a plausible range around the forecast, have traditionally been prepared in addition. These two series have also generally followed the deterministic model.

Leslie matrix. A compact mathematical summary of the cohort-component method of preparing population projections is given by the Leslie matrix and a single projection equation. The Leslie matrix is:

$$
{}_5\mathrm{S_x} = \frac{{}_5\mathrm{L_x}+5}{{}_5\mathrm{L_x}} \qquad L = \begin{vmatrix}
0 & 0 & \frac{\mathrm{L_0}}{2\mathrm{l_0}}\left(\frac{\mathrm{L_{15}}}{\mathrm{L_{10}}}\mathrm{F_{15}}\right) & & \frac{\mathrm{L_0}}{2\mathrm{l_0}}\left(\mathrm{F_{15}}+\frac{\mathrm{L_{20}}}{\mathrm{L_{15}}}\mathrm{F_{20}}\right) & \cdots\cdots & 0 & 0 \\
{}_5\mathrm{S_0} & 0 & 0 & & 0 & \cdots\cdots & 0 & 0 \\
0 & {}_5\mathrm{S_5} & 0 & & 0 & \cdots\cdots & 0 & 0 \\
0 & 0 & {}_5\mathrm{S_{10}} & & 0 & \cdots\cdots & 0 & 0 \\
0 & 0 & 0 & & {}_5\mathrm{S_{15}} & \cdots\cdots & 0 & 0 \\
\vdots & & & & & & & \vdots \\
\vdots & & & & & & & \vdots \\
0 & 0 & 0 & 0 & 0 & {}_5\mathrm{S_{75}} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & {}_\infty\mathrm{S_{80}} & 0
\end{vmatrix} \tag{14.2}
$$

where L stands for the Leslie matrix, ${}_5S_x$ represents a 5-year survival rate for ages x to $x + 5$, and $\mathrm{F_x}$ represents an age-specific birth rate (boy or girl babies) for women aged x to $x + 4$. To project the population for a series of 5-year periods into the future, apply the following equation in accordance with the Markovian model:

$$
\mathbf{N}_{t+5} = \mathbf{L}_t^{*}\mathbf{N}_t + \mathrm{M}_t \tag{14.3}
$$

where $\boldsymbol{N}_{t+5}$ is the column vector containing population counts for age groups at time $t + 5, \boldsymbol{N}_t$ is the corresponding column vector for population age groups

at time t, and M_t is the period allowance for net migration. The Leslie matrix and Eq. 14.3 express in matrix algebra the iterative process of multiplication of a base population by a set of projected survival rates and a set of age-specific maternal birth rates, and the addition of net migration with each iteration.[2]

Setting targets for mortality. Focusing on the mortality projections, the survival rates needed to carry the base population forward must be extrapolated to future dates to allow for prospective changes in mortality levels. This type of projection has been made in terms of a set of age-sex-specific death rates that have been estimated on the basis of predetermined target values for life expectation at birth for some distant future year, e.g., 2075. Alternatively, the set of target age-sex-specific death rates may be determined initially for the target date and then the life expectation values and the survival rates for the target date that they imply are calculated from them.

Several devices for setting the target values for the age-sex-specific death rates or the expectation of life at birth for the sexes may be considered. Projected age-specific death rates for a population may be linked to the age-specific death rates or life expectation of another population that has a more advanced longevity record. Structural equation models, the type of models usually employed in econometric studies, may be employed. These equations link changes in mortality to changes in various demographic, socioeconomic, and other variables in multivariate regression equations. The age-specific death rates or life expectation may be extrapolated by statistical methods on the basis of their own historical time series. The model fitted may be an exponential curve, a logistic curve, ARIMA-type time series (see below), or other model.

For the United States, we may consider as the target value for life expectation the current e_o value for the best state, the e_o value obtained by combining the lowest death rates at each age for the states, the highest current e_0 value in any country, or the e_o obtained by combining the lowest death rates at each age for the countries having reliable data. Another possible target is the value obtained for e_o from the endogenous (intrinsic) death rates at each age for the United States, or the value obtained by eliminating all deaths due to unhealthful lifestyles and behaviors. The figure obtained by selecting the lowest age-specific rates now observed internationally and combining them to derive a new life table is more favorable than are the values based on endogenous mortality but less favorable than the value based on the years added by pursuit of a healthful lifestyle and behavior. Judgments will differ as to the causes of death considered as endogenous

[2]The implementation of this matrix equation and the Leslie matrix requires an abridged life table, a set of age-specific maternal birth rates, and a population age-sex distribution. Five-year survival rates are shown in the subdiagonal and the birth rates are shown in row one; the age distribution for one of the sexes is placed on the right as a column vector. The birth rates in row one are modified so as to allow for the sex of the births, the shift in the age distribution of the female population over a 5-year period, and the survival of the births to ages under 5 at the end of the 5-year period. For further explanation, see Keyfitz (1977:31).

Table 14.1 Alternative projections of life expectation at birth, for U.S. males and females, prepared by the U.S. Census Bureau in 1999 (Assumptions used in interim national population projections released in 2004)

	Male				Female			
Year	Lowest	Middle	Highest	Range[a]	Lowest	Middle	Highest	Range[a]
2006[b]		75.1				80.2		
1999	74.0	74.1	74.1	0.1	79.7	79.8	79.8	0.1
2025	76.5	77.6	79.1	2.6	82.6	83.6	84.6	2.0
2050	79.5	81.2	83.8	4.3	84.9	86.7	88.4	3.5
2100	85.0	88.0	92.3	7.3	89.3	92.3	95.2	5.9

Source: U.S. Bureau of the Census (1999). *Population Division Working Paper* No. 38. www.census.gov/population/www/projections
[a]Difference between highest and lowest series
[b]Current estimates

Table 14.2 Alternative projections of life expectation at birth and at age 65 for males and females, Prepared by the Actuary's Office, Social Security Administration: 2007–2060

	Male				Female			
Year	Low cost	Intermediate	High cost	Range	Low cost	Intermediate	High cost	Range
At birth								
2007[a]		75.2				79.9		
2010	75.4	75.7	75.9	0.5	79.8	80.0	80.2	0.4
2020	76.0	76.9	77.9	1.9	80.1	80.9	81.7	1.6
2030	76.5	78.0	79.5	3.0	80.5	81.8	83.1	2.6
2040	77.0	79.0	81.1	4.1	80.9	82.6	84.5	3.6
2050	77.5	80.0	82.6	5.1	81.3	83.4	85.7	4.4
2060	77.9	80.8	83.9	6.0	81.6	84.2	86.9	5.3
At age 65								
2007[a]		16.7				19.2		
2010	16.8	16.9	17.1	0.3	19.2	19.3	19.4	0.2
2020	17.0	17.6	18.2	1.2	19.3	19.8	20.3	1.0
2030	17.3	18.2	19.2	1.9	19.5	20.3	21.3	1.8
2040	17.5	18.8	20.2	2.7	19.7	20.9	22.3	2.6
2050	17.7	19.3	21.2	3.5	19.9	21.4	23.3	3.4
2060	17.9	19.8	22.1	4.2	20.1	21.9	24.1	4.0

Source: *The 2008 annual report of the board of trustees of the federal old-age and survivors insurance and disability insurance trust funds*, 2008
[a]Preliminary or estimated

or causes of death due to unhealthful lifestyles and behaviors, but even an expansive definition of these categories yields conservative target values in comparison with an exponential extrapolation of the 100-year U.S. trend in life expectancy at birth – 100 years or so in 2060. Projected values for the expectation of life at birth in future years according to the U.S. Census Bureau and the Social Security Administration are shown in Tables 14.1 and 14.2. Their values for 2050 or 2060 are lower than a simple arithmetic (95 years) or exponential (100 years) extrapolation of the trend of e_0 in the United States in the last 100 years.

The value for best U.S. state is 78.2 (Hawaii, 1989– 1991). Assuming that Hawaii continued to progress as has the United States as a whole since 1989–1991, its figure would be 80.6 in 2006. Life expectation at birth derived by contructing a life table with the lowest age-specific death rates internationally (83) is hardly different from the present-day value for Japan (83), whose life expectancy for females is now a few years ahead of that of the other leading countries. A study of Seven Day Adventists conducted by Fraser and Shavlik (2001) concluded that optimal health-related behaviors in the United States would add 10 years to life expectancy. This estimate assumes improvements in diet, a program to bring everyone's weight below the overweight line, universal practice of adequate exercise, and the elimination of tobacco use. This regimen would yield a much higher life expectancy (88 years) than at present, but not the highest of the proposed projections for 2060 (100 years).

In deriving target values for mortality, the analyst may decide to adopt a set of age-specific death rates from another population. This "model" population should resemble the adoptive population in socioeconomic, cultural, and other characteristics, such as technological advancement and the general profile of the main causes of death, but be more advanced with respect to health status.

Age-delay method. A variant of the concept of "borrowing" age-specific death rates from other populations is the age-delay model. Here, the donor population is the same as the recipient population. In this method it is assumed that medical developments and improvements in lifestyles and behaviors postpone the death rate at a given age in a given year to a later age in a later year. The mortality level of later (i.e., younger) cohorts in a given year is selected to represent the future mortality level of earlier (i.e., older) cohorts. Accordingly, with the passage of time each cohort experiences lower death rates when it reaches a given age than the previous cohort (Olshansky 1987; Manton et al. 1990; Bongaarts 2004). Projections of mortality rates can be made by assuming a delay of a specified number of ages and years. With an assumption of a delay of one age every three years – a conservative improvement, age-specific death rates at ages 60–63 in year y would be projected as follows, assuming a linear decrease in the intermediate years:

	year						
Age	y	y + 1	y + 2	y + 3	y + 4	y + 5	y + 6
60	.0100						
61	.0121	.0113	.0107	.0100			
62	.0137	.0132	.0126	.0121	.0113	.0107	.0100
63	.0153	.0148	.0142	.0137	.0132	.0126	.0121

Here the death rates at ages 60 to 63 in year y are repeated in future year $y + 3$ at the next higher age, and in year $y + 6$ at the second higher age; intermediate values were obtained by linear interpolation of the values at the same age.

An extreme variant of this method of making mortality projections is to eliminate one or more causes of death completely. Competing risks among deaths and other

factors make this assumption problematic, but not impossible, to implement and interpret. The results would be of purely analytical value. The availability and description of cause-elimination life tables that can be used to implement such an assumption were discussed in Chap. 4. Alternatively, a major risk factor such as obesity could be gradually eliminated to evaluate the implications of such changes. Projections of mortality based on the assumption that obesity is eliminated would in effect reduce the impact of deaths from diabetes, cardiovascular diseases, kidney disease, and certain other diseases on life expectancy, but measuring this impact is difficult and subject to considerable error.

As indicated earlier, the projections of the components are often based on long-term assumptions relating to a summary measure for the particular component, e.g., the total fertility rate for births, life expectation at birth for deaths, and the absolute annual amount or rate of net immigration. In its earlier projections of mortality, the U.S. Census Bureau gave target values in terms of life expectancy at birth for each sex and race in the twenty-second century. Age-sex-race-specific death rates consistent with these target values are determined for the terminal year and death rates for intermediate years are derived by interpolation between the base year and the terminal year. In its most recent projections of mortality, the U.S. Census Bureau (2009) employed time series analysis on the basis of historical trends for 1984–2003 for the principal racial-ethnic group (non-Hispanic non-black) to develop age-sex-specific mortality schedules to 2075. Death rates for the other racial/ethnic groups were estimated by logistic interpolation between 2003 and 2075 on the assumption of convergence of the mortality of these groups to that of the non-Hispanic non-black group.

Actuarial model. Some analysts may differentiate the actuarial model from the demographic model, but these two approaches can hardly be distinguished. Like the demographic model, the actuarial model is a cohort-component method that sets target values for the components or projects the components by time series analysis. On the other hand, while demographic projections tend to be general-purpose projections, the actuarial model is intended to serve specific actuarial uses (e.g., calculation of actuarial tables), tends to project selected populations (*i.e.*, "Social Security" population, insured population, non-smoking population), and often employs rather complex mathematical formulas in the projection of mortality.

The mortality projections made in 2008 by the Actuary's Office of the U.S. Social Security Administration (SSA) for ages under 65 were based on NCHS data using delphic-type assumptions as to the level of age/sex/cause-of-death-specific death rates for a terminal year about 80 years ahead. Seven categories of causes of death were modeled in the assumptions. In the intermediate series of SSA projections, current average annual percent reductions in age/sex/cause-of-death-specific death rates were assumed to change to the ultimate average annual reductions assumed for 2032 and later years on the basis of the average annual reduction in the age-adjusted death rate between 1979 and 2004 (0.86%). Between 2032 and 2082 the age-adjusted death rate for all ages was assumed to decline at an average annual rate

of 0.73%. Future death rates at ages 65 and over were based on a historical series of Medicare data for 1979–2004, which declined at an annual average rate of 0.66%. The rates were assumed to decline at an annual average rate of.65% between 2032 and 2082.

Mathematical and statistical extrapolations of mortality. Several analysts have proposed mathematical or statistical extrapolations of past series of age-specific death rates without setting targets initially. Thatcher et al. (1998), Thatcher (1999), and, more recently, Bongaarts (2004) proposed a logistic formula to extrapolate death rates. Lee and Carter (1992) and McNown and Rogers (1992) made projections of mortality by ARIMA time-series techniques.[3] These techniques involve various devices to transform the original series into a stationary form, that is, a series whose mean, variance, and autocorrelation structure do not change over time. In general, this means a flat-looking series without a trend and periodic fluctuations. One ARIMA device or a combination of ARIMA devices may be applied so as to achieve an approximately stationary form. A time series can usually be converted into a stationary series by calculating differences between the data points and/or applying one or two other ARIMA processes to the series.[4] Then the residuals are modeled.

Stochastic demographic forecasting (Lee-Carter method). The Lee-Carter method (Lee and Carter 1992), also called stochastic demographic forecasting, is a cohort-component method in which the series of fertility and mortality rates are modeled as stochastic processes and projected by ARIMA time-series techniques of the Box-Jenkins type. Stochastic forecasting incorporates probabilistic calculations into the method; that is, it generates confidence intervals for the forecasts of fertility and mortality stochastically. In general, a stochastic projection model, in contrast to a deterministic projection model, is a model that projects probability distributions of possible outcomes by allowing for random variations in one or more variables (e.g., age-specific death rates) over time. The random variations are generally based on fluctuations observed in historical data for a defined past period.

To apply stochastic forecasting for United States mortality and fertility, Lee and Carter estimated stochastic time-series models for the principal demographic variables (e.g., age-specific birth rates and age-specific death rates). For these variables, data were analyzed for a lengthy historical base period, namely 1933–1987. The application of many simulations produces multiple outcomes that serve

[3] ARIMA is the acronym for autoregressive integrated moving average. The terms refer to the various processes by which the time series is transformed for the purpose of preparing it for projection. In autoregression, the data points are "regressed" against a previous data point or points in the same time series; "integration" refers to the order of differencing of the data points (e.g., first or second differences); and "moving average" refers to the substitution of moving-average values for the original values (defined by the number of years included in the moving average).

[4] The original series may be transformed by taking first or second differences, calculating natural logarithms or square roots, or fitting a straight line. Moving averages may be calculated from the transformed series or the series may be "regressed on itself" with a specified time lag (1 or 2 years).

as a probability distribution for the forecasts.[5] It is important to recognize that such confidence intervals do not represent statistical probabilities in the conventional sense, but rather, they provide a measure of how confident one is that future changes in mortality or fertility will resemble those observed in the past.

The Lee-Carter projection method employs the following mortality model:

$$\ln m(x,t) = a(x) + b(x)^{*}k(t) + E(x,t) \tag{14.4}$$

where $m(x,t)$ = the central death rate at age x and time t, $a(x)$ = mean ln m_x (age-specific constants describing the general pattern of mortality by age), $b(x)$ = age-specific constants representing the relative change in mortality rates at a given age, $k(t)$ = an index of the level of mortality at time t, and $E(x,t)$ = an error term or residual. This model was found to fit past age-specific death rates from 1933 to 1987 in the United States well, explaining 93% of the within-age-group variance in this period (Lee and Carter 1992).

A projection requires the extrapolation only of k(t), the index of the intensity of mortality. Two of the parameters, $a(x)$ and $b(x)$, are held constant for the projection period after being determined from the historical series by the estimating equation. To project $k(t)$, Lee and Carter employed ARIMA time-series techniques and assumed a random walk with drift.[6] More specifically, a linear trend in $k(t)$, varying with ln $m(x.t)$, is assumed for the future, with the result that the projected mortality rates at every age follow an exponential decline. While the projected proportional rate of mortality decline varies by age, it is assumed to be equal at each age to the rate observed in the historical period.

Once k is modeled as a stochastic time series process, its use in estimating m_x's means that the m_x's are also modeled as a stochastic process. Employing the equation where m_x is a function of k and inserting into it the estimated values of the age-specific coefficients a and b, we can solve for the forecast values of m_x and

[5] A simulation is a form of model testing, i.e., a calculation, normally carried out by computer, in which a set of assumptions or conditions, either theoretical or realistic, are applied to the model in order to generate one or more representative scenarios where it is not possible or desirable to test all conditions pertaining to the model.

[6] The equation for a random walk with drift is,

$$k_t = c + k_{t-1} + E_t$$

where c is the drift term and E_t is a normally distributed random variable with mean zero and variance σ^2. A random walk with drift refers to the way the ARIMA model is applied. Initially, the first differences of the series are examined to see if a predictable pattern can be discerned. The transformed series tends to appear stationary and quite random. The random-walk model assumes that the projected figures take a random step from their last recorded level equal to the average of a random walk-with-drift-difference in the past. If this average difference is zero, it is a random walk without drift. In a random walk with drift, the projected series includes a non-zero constant term, assuming an upward or downward trend. In ARIMA terms, this is a "(0,1,0) model with constant," where the 1 refers to first differences.

compute all desired life-table functions. Applying the Lee-Carter method to U.S. data resulted in an increase in life expectancy from 1990 to 2065 of 10.5 years, or in a life expectancy of 86.1 years in 2065, with a confidence band of +3.9 years and −5.6 years.

The Lee-Carter method has been widely adopted. See, for example, Tuljapurkar et al. (2000), United Nations (2003), U.S. Census Bureau (2008), U.S. Social Security Technical Advisory Panel (2003), Tabeau et al. (2002), and Torri and Vignoli (2007).

The Lee-Carter model has certain strengths. It has only three parameters and represents the trends and patterns of mortality of the past well. The projection equation is based on data for an extended past period, over 50 years. The method provides confidence intervals for the point forecasts. Moreover, as evaluation tests show, the method is relatively accurate over short projection periods. The method has an important weakness, however. The assumption that rates of mortality improvement at each age remain unchanged over time is not supported by the experience of several countries (Bongaarts 2005). In fact, rates of mortality improvement have tended to fall off at the younger ages and to rise at the older ages. (See Booth et al. 2002.) As a result, the Lee-Carter method may project unreasonable age patterns of mortality over the long run. It is desirable, therefore, to modify the Lee-Carter assumption of constancy in rates of mortality improvement in the future.

Shifting logistic (Bongaarts) method. Bongaarts (2005) has proposed a "shifting logistic model" to overcome this presumed weakness of the Lee-Carter method. This logistic model has three parameters. The "slope" parameter is assumed to be fixed over time while the "level" parameter and the "exogenous-mortality" parameter vary with time. The equation is:

$$\mu(x,t) = \frac{k(t)e^{bx}}{1 + k(t)e^{bx}} + R(t) \tag{14.5}$$

where $k(t)$ is the level time-varying parameter, $R(t)$ is the exogenous-mortality time-varying parameter, and b is the slope time-constant parameter. The level and exogenous-mortality parameters have to be extrapolated. In this model the constancy of the slope parameter means that the intrinsic (nonexogenous) component of total mortality shifts to higher or lower ages as mortality conditions improve or deteriorate. Estimates of the parameters were obtained by fitting the logistic model to mortality schedules for an extended past period, 1950–2000, for 14 countries. (For this purpose a nonlinear least squares curve was fitted to the data by the program package STATA.)

Epidemiological methods. Methods of projecting mortality that take account of the causes of disease or death as vehicles for making projections of mortality may be labeled epidemiological. Epidemiological methods are now particularly concerned about the effect on mortality trends of the ongoing epidemiological transition, reflecting the new dominance of endogenous causes of morbidity and mortality over the exogenous causes. They may also be concerned about the emerging

epidemiological transition, reflected in the resurgence of several old infectious diseases, the emergence of some new ones, particularly HIV/AIDS, and the obesity epidemic among children. Hence, the SSA method, the cause-delay technique, the McNown/Rogers method, and the health process method can be considered epidemiological. The latter two methods are described below.

McNown and Rogers (1992) projected mortality on the basis of an analysis of age-sex-specific death rates disaggregated by cause. The age profiles of death rates for five leading causes of death are represented by parametrized mortality schedules at past dates. More specifically, a multiexponential model with nine parameters was fitted to the age profiles of cause-specific death rates for three parts of the age distribution over the 1960–1985 period. Three of the nine parameters, namely, the "level" parameters for the three parts of the age distribution, were assumed to vary with time and were projected by ARIMA time series methods. The projected values of the parameters were then used to determine the multiexponential model for future dates and the future age-sex-cause-specific death rates.

The method referred to as the health process model, and proposed by Manton and Woodbury (1992), Manton and Stallard (1993, 1994), and Manton et al. (1990), is another type of epidemiological model. It derives mortality projections by analyzing the health processes that result in death from the various causes. The method combines several of the analytic and statistical methods described earlier: Time-series analysis, stochasticity (i.e., providing confidence limits for the projections obtained), and a structural equation model (which provides measures of the future impact of risk factors on health status). The method claims to have the ability to forecast turning points in health, mortality, and population and, because it uses much more information on aging and health than alternative methods, it claims to be superior to methods that are based merely on cause-specific death rates (such as the SSA or McNown/Rogers methods). (See section "Health Projections" below for further discussion.)

The desirability of preparing health-status projections as a prior step in preparing mortality projections is a logical basis as to how to proceed. It uses and provides additional information that enlightens the analyst regarding the prospects for mortality and population changes. In this sense it has additional analytic utility. Evidence that health analysis adds to the accuracy of mortality projections is lacking, however.

Econometric/structural equation models. In the econometric model, regression analysis is the basis for the projections of mortality. A system of simultaneous multivariate linear equations links a mortality or longevity measure with several independent variables. The longevity measure (or a measure of the change in longevity over a specified past period) serves as the dependent variable to which are related several independent explanatory variables, such as socioeconomic status, medical technology, and health organization. Once the parameter values are determined from historical data, it is necessary to develop projections of the independent explanatory variables. The projections of the independent variables can now be combined with the parameter values in a new equation to make projections of the dependent mortality/longevity variable.

Murray and Lopez (1996) and collaborating organizations, namely, WHO, World Bank, and Harvard University School of Public Health, employed a variation of this method to derive projections of mortality for their "Global Burden of Disease" study. Specifically, the method consisted of "regressing" four socioeconomic variables with mortality rates for all major causes for the period, 1980–1991, for 47 countries; (1) projecting the four socioeconomic variables; and (2) solving for the trend of mortality rates from the regression equations. The socioeconomic variables, all known to be related to mortality changes in the past, are: (1) per capita income; (2) average years of schooling for adults; (3) time, a proxy for the secular improvement in health during the past century; and (4) use of tobacco. In the case of tobacco use, account was taken of the time lag observed in the past between persistent tobacco use and its health consequences. A separate model was used for HIV, with adjustments for the interaction between HIV and tuberculosis. Three projection scenarios were developed using different projections of the independent variables.

Biological Determinism vs. Demographic Determinism

The following discussion is a brief reprise of the two major perspectives on projecting longevity. Some methods of projecting mortality depend wholly on the extrapolation of past series of mortality data, while others depend primarily on a theory as to how longevity and health will evolve. The first group may be described as pursuing a course of demographic/statistical determinism (Oeppen and Vaupel 2002; Vaupel et al. 1998; Wilmoth 1997), and the second as pursuing a course of biological determinism (Olshansky et al. 1998, 2001; Fries 1980; Fries and Crapo 1981). These represent two conflicting views as to the possibilities for the extension of life expectancy – one "liberal," the other "conservative."

The first approach draws very heavily on the mortality experience of the past and, in accordance with this experience, tends to be expansive in its assumptions. It points to the evidence of the continuous rise in the maximum age at death in selected countries with validated records (Wilmoth 1997), the deceleration in the rate of increase of mortality rates at the advanced ages (Horiuchi and Wilmoth 1998), and the numerous medical developments under way that may be expected to foreshadow further extensions of life expectancy and human life span. This position posits a life expectancy of 100 years and a maximum average life span of 150 years by 2065, assuming favorable circumstances. Favorable circumstances consist of the elimination of smoking, widespread pursuit of a nutritionally adequate diet and aerobic exercise, and the general availability and use of the medications that patients require. It is not clear whether this position calls for success by cell biologists in delaying the aging process.

Not all demographic extrapolative methods are so expansive, however. For example, for its projections of mortality to 2050, the Population Division of the United Nations (2007) assumes 82.5 years as the threshold value (81 male; 84 female) for the developed countries, and 84 (81, male; 87, female) for Japan, the

ranking country. The U.S. Census Bureau projects the following "middle" values for 2050 for use in its population projections:

	Total	Hispanic	Non-Hispanic black	Non-Hispanic other
Male	80.9	81.9	79.0	81.0
Female	85.3	86.3	84.3	85.3

Source: U.S Census Bureau (2009)

The Social Security Administration has also prepared several series of projections for its 2008 round. The values projected for life expectancy in 2050 in the SSA projections are:

	Low cost	Intermediate	High cost
Male	77.5	80.0	82.6
Female	81.3	83.4	85.7

Source: Board of Trustees of the OASDI Trust Funds (2008)

The biological approach offers only general guides for the future but its proponents make specific predictions about how many years life expectancy can rise in the absence of major breakthroughs that could slow aging in humans (Olshansky et al. 2001). They hold that, although aging and the life span of species are not products of evolution or specific genetic programs that evolved for that purpose, the duration of life of humans and most other living things is calibrated to genetically fixed programs that regulate biological processes associated with growth, development, and reproduction. The processes of senescence and the duration of life are believed to be inadvertent by-products of these programs. Body design and the basic physiology of humans also place limits on the duration of life (Olshansky et al. 1990, 1998). Although lifestyle modification and further advances in medical technology may be expected to continue to exert downward pressure on death rates, there are biological and biomechanical reasons why such pressure must eventually come up against a limit – even if the limit itself is not a direct product of evolution.

Supporting this line of reasoning are the overwhelming difficulties in achieving the very large declines in death rates needed for even small increases in life expectancy at currently low levels of mortality, the resurgence of several infectious diseases (e.g., tuberculosis, diphtheria), the appearance of new infectious diseases (e.g., HIV/AIDS), and the onset of an obesity epidemic among younger generations that may be expected to express itself in higher death rates at the middle and higher ages (Olshansky et al. 2005). Even success in reducing the level of some major causes of death may not add much to life expectancy because of the prevalence of comorbidities at the later ages, the multiple causes of death associated with the underlying cause of death, and the competing risks of the causes. Thus, if heart

disease is sharply reduced or "eliminated," the decline in other endogenous diseases such as stroke, diabetes, and kidney disease may be stalled and their rates may even rise.

Olshansky et al. (2001) and Fries (1980) have never set a target year for the time when life expectancy would reach a peak but they cite age 85 years as that peak (82, males; 88, females). The former argue that it is unlikely that this value will ever be exceeded, given present technology, but that it could be exceeded if it becomes possible to slow the rate of biological aging. In their view slowing the rate of aging is currently not possible but they believe that it will be achieved in time.

More recently, research carried out under the auspices of the MacArthur Network on an Aging Society by Olshansky et al. (2009) concluded that the projections of the U.S. Census Bureau and the Social Security Administration may significantly underestimate life expectancy in the United States. They forecast that by 2050 Americans may live 3.1 to 7.9 years longer than given by the government agencies (see above), resulting in life expectancies of 83.2–85.9 for males and 89.2–93.3 for females. Their findings, which in effect exceed the peak of 85 years posited earlier by the conservatively oriented analysts and could represent a new position for them, are based on the assumption that death rates will be reduced in the coming decades by accelerated advances in biomedical technology, delaying the onset and progression of the major causes of death or slowing the aging process. The projections of the Census Bureau and SSA assumed that improvements in mortality in future decades would decelerate.

The arguments on the two main sides of this issue, the essentially conservative and the essentially liberal, are all cogent but lead to very different conclusions. Some resolution of the issue, even some melding of the positions, is desirable in the interests of formulating public policy and planning major public programs, such as Social Security and Medicare. In view of the strong professional commitments of the advocates and the cogency of the arguments on both sides in this matter, I recommend taking the mean of the two proposed projections of life expectancy for 2050. This intermediate position may offset opposing biases; it is also an application of the principle of mathematical expectation. In making local population estimates demographers have had favorable experience with averaging two or more population estimates produced by independent methods and presumably having opposing biases.

Uncertainty and Its Management

Given the uncertainty of projections in general and the uncertainty of population and mortality projections in particular, it is especially desirable to present some measure of uncertainty to accompany any set of population and mortality projections. A number of devices have been used to reflect and assess the uncertainty of such projections:

1. Generating additional projections by independent methods, having presumably different and opposing biases;

2. Modeling the effect of alternative assumptions for the various components (i.e., carrying out sensitivity analyses);
3. Deriving alternative projection series with different assumptions for the basic components, designed to represent a plausible range from the "high" to the "low" series;
4. Conducting a systematic analysis of errors in past forecasts; and
5. Generating confidence limits stochastically, particularly for the middle series. In the effort to produce as realistic projections of population and mortality as possible, it is useful, in addition, to revise a set of projections frequently, tying them in with the latest current data on population changes.

Forecasts of national population are essential to the planning, implementation, and conduct of numerous federal and private programs. Therefore, however problematic their accuracy and wide their confidence intervals, they must be available for such uses. The measures of uncertainty that should accompany the projections, however derived, will provide a measure of caution in these uses. The details of applying the evaluation methods listed above are not presented here, but their general nature is suggested in the text.

Analytical utility vs. accuracy of forecasts. For many purposes, including both their practical and research uses, the most important value of projections is their analytic utility – that is, their use in sensitivity analysis – not their accuracy. For example, they provide insight into the effect of a change in one variable or component of the population system on the change in the rest of the system, such as the effect of a change in the level of the death rate or life expectancy on population size. The design of cohort-component projections readily permits testing the sensitivity of the system to a particular demographic assumption or, alternatively stated, evaluating the robustness of the assumption in its effect on the system. In addition, analytic projections can serve as an early warning system anticipating adverse demographic or socioeconomic developments if public action is not taken to change the course of the hypothesized demographic events. For example, informative analytic projections can be prepared modeling the effect of alternative degrees of control of the current obesity epidemic on future mortality rates. The effect on mortality levels and population size of partially eliminating major chronic diseases is a type of modeling in which there is considerable public interest.

Presenting alternative series. As is commonly done with state and local population estimates, the analyst can generate more than one series of projections by independent methods. If the various series can be interpreted as having opposing biases, they could reasonably be averaged. The weights used in the averaging process could be equal or unequal, as assessed by the analyst. Alternatively, different population projections can be prepared by the same method using different assumptions. As was noted above, the Actuary's Office of the U.S. Social Security Administration (SSA) and, up through the 1990s, the Census Bureau allowed for the uncertainty of population projections by developing three alternative population series with varying assumptions for the components of change (i.e., births, deaths,

and migration). Each of the three series of population projections incorporate a different set of projections of fertility, mortality, and net migration. The middle or intermediate series is offered as the most likely estimates of the future level of the U.S. population and its demographic changes, that is, as a forecast, and the others as setting plausible bounds to the middle series. Each population and mortality series have different, but unspecified, chances of realization, with the imputed probability for the lowest and highest series being far lower than that for the middle series.

However carefully conceived the assumptions about mortality and the other components of change may be, judgment is involved in their design and any given series of population projections is subject to substantial error. Commonly, the range from the low to the high series is assumed to represent a confidence interval for the middle series, but the probability level for these series cannot be closely specified. On the basis of a study of the error structure of past U.S. Census Bureau projections of the U.S. population, this range is estimated to represent approximately a 68% confidence interval (Keyfitz 1981; Stoto 1983). Greater (e.g., 95%) confidence intervals for the forecasts of the population are so wide as to provide little guidance to policymakers with respect to the likely size of the future population.

The general magnitude, if not the exact extent, of error in the middle mortality projections is suggested by the gap between the various series of mortality projections incorporated in a given set of population projections. The low, middle, and high series of values for life expectation at birth in the projection series published by the Census Bureau in 1999 can be compared in order to assess the uncertainty in the projections. Table 14.1 shows the projected figures for life expectancy at various dates from the beginning to the middle of the present century. The range of the series grows wider and wider, i.e., less and less dependable, as the projection period lengthens. After a half century the projections of deaths could be very wide of the mark and the range may not straddle the actual number. Table 14.2 shows similar figures for the projections of life expectation made by SSA in 2008. These figures widen more rapidly than the Census Bureau figures. A comparison of the intermediate series of projections of life expectation for the first half of this century, made, for the most part, by SSA in the last half of the last century, with the actual figures shows that many, if not most, of the projections are too low (Table 14.3).

Because of the age concentration of mortality, the errors in the projections of population at the older ages represent mainly errors in the mortality projections. *Ex post facto* studies comparing the intermediate forecasts of the older population for the United States with the actual numbers as indicated by census figures or the most recent projections show that the forecast errors are substantial after a few decades and become sizeable after several decades. Table 14.4 presents estimates of percent errors in selected forecasts of the Social Security area population 65 years old and over prepared by the U.S. Actuary's Office, SSA, over the last quarter century for years to 2040. They understate the elderly population for most comparison dates.

The Census Bureau and the Social Security Administration (SSA) have often been criticized on the grounds that they have tended to employ assumptions on mortality that are too conservative and that the allowance for the uncertainty

Table 14.3 Comparison of actual values for U.S. life expectation and values projected by the U.S. Actuary's Office, for 1990, 2000, and 2050

Life expectation Pub. year (Base year)	Actuarial report no.	Expectation at birth			Expectation at age 65		
		1990	2000	2050	1990	2000	2050
Male							
1957 (1954)	46	NA	71.4	71.4	NA	15.0	15.0
1966 (1965)	62	69.4	70.3	70.3	13.9	14.3	14.3
1974 (1972)	72	68.6	69.0	NA	13.4	13.6	NA
1978 (1977)	77	69.6	70.3	71.06	14.2	14.6	15.4
1983 (1981)	88	72.3	73.4	75.8	15.1	15.7	17.5
1987 (1986)	99	72.3	73.9	76.7	14.9	15.6	17.4
1991 (1989)	106	71.9	72.9	76.7	15.3	15.9	17.9
1996 (1994)	110	71.8	73.0	77.2	15.0	15.6	17.7
2008 (2004)[a]	X	71.8	74.0	80.0	15.1	15.9	19.3
Female							
1957 (1954)	46	NA	77.1	77.1	NA	17.5	17.5
1966 (1965)	62	75.6	76.4	76.4	16.8	17.2	17.2
1974 (1972)	72	76.3	76.9	NA	17.8	18.1	NA
1978 (1977)	77	77.3	78.0	80.4	18.4	18.9	20.5
1983 (1981)	88	79.8	81.0	83.8	19.9	20.8	23.1
1987 (1986)	99	79.3	80.8	83.7	19.2	20.1	22.4
1991 (1989)	106	78.8	79.9	83.1	19.0	19.6	21.7
1996 (1994)	110	78.8	79.7	83.0	19.0	19.4	21.4
2008 (2004)[a]	X	78.9	79.4	83.4	19.1	19.6	21.4
Deviations from actual							
Male							
1957	X	NA	−2.6	−8.6	NA	−0.9	−4.3
1966	X	−2.4	−3.7	−9.7	−2.2	−1.6	−5.0
1974	X	−3.2	−5.0	NA	−1.2	−2.3	NA
1978	X	−2.2	−3.7	−8.4	−0.6	−1.3	−3.9
1983	X	+0.5	−0.6	−4.2	+0.5	−0.2	−1.8
1987	X	+0.5	−0.1	−2.3	+0.2	−0.3	−1.9
1991	X	+0.1	−1.1	−2.3	–	−0.1	−1.4
1996	X	–	−0.9	−2.8	–	−0.3	−1.6
Female							
1957	X	NA	−2.3	−6.3	NA	−2.1	−3.9
1966	X	−3.3	−3.0	−7.0	−2.3	−2.4	−4.2
1974	X	−2.6	−2.5	NA	−1.3	−1.5	NA
1978	X	−1.6	−1.4	−3.0	−0.7	−1.7	−0.9
1983	X	+0.9	+1.6	+1.3	+1.0	+1.2	+1.7
1987	X	+0.4	+1.4	+0.9	+0.3	+0.5	+1.0
1991	X	−0.1	+0.5	+0.3	+0.1	–	+0.3
1996	X	−0.1	+0.3	−0.4	−0.1	−0.2	–

Source: Selected *Actuarial Studies* of the U.S. Office of the Actuary, Social Security Administration, and 2008 *Annual report of the Board of Trustees of the federal OASI and disability insurance trust funds*, 2008

Figures shown are intermediate series or averages of low and high series

NA not available, *X* not applicable, – rounds to zero

[a]Intermediate projection in *2008. Annual report of the Board of Trustees of the federal OASI and DI trust funds*

Table 14.4 Comparison of selected projections of the Social Security area population 65 years and over with the latest estimates and projections: 1980 to 2040

Population	1980	1990	2000	2010	2020	2030	2040
Publ. year (base year)							
2008 (2007)	26,196	31,981	35,474	40,303	54,082	70,437	77,353
2002 (2001)	26,149	32,036	35,516	39,481	53,150	69,408	75,177
1997 (1995)	25,870	31,719	35,188	39,499	52,230	67,881	73,178
1991 (1989)	25,832	31,711	35,559	39,905	52,576	67,402	71,569
1988 (1986)	25,832	31,718	35,480	39,718	52,026	66,123	69,650
1984 (1982)	26,364	32,570	36,184	40,574	53,273	66,340	68,847
1978 (1977)	25,675	30,708	32,960	36,045	45,409	56,871	57,375
1974 (1973)	24,969	29,265	31,034	33,629	42,766	51,383	50,347
Percent "error"							
Publ. year (base year)							
2008 (2007)	–	–	–	–	–	–	–
2002 (2001)	X	X	X	−2.0	−1.7	−1.5	−2.5
1997 (1995)	X	X	−0.8	−2.0	−3.4	−3.7	−5.4
1991 (1989)	X	−0.8	+2.4	−1.0	−2.8	−4.2	−7.5
1988 (1986)	X	−0.8	–	−1.5	−3.8	−6.1	−10.0
1984 (1982)	X	+1.8	+2.0	+0.7	−1.5	−5.8	−11.0
1978 (1977)	−2.0	−4.0	−7.1	−10.6	−16.0	−19.3	−25.8
1974 (1973)	−4.7	−8.5	−12.5	−16.6	−20.9	−27.1	−34.9

Source: U.S. Actuary's Office, Social Security Administration, Actuarial Studies nos. 72, 77, 92, 102, 106, 112, and Trustees Reports for 2002 and 2008

Note: The estimates made in 2008 are taken as the most accurate for the population in the year shown and are used as the bases of the percent errors. The estimates include an allowance for underenumeration in the censuses and for Puerto Rico, other outlying areas, and overseas citizens

Population figures in thousands

X not applicable, – rounds to zero or same as the base

of their middle, or intermediate, series of projections in the form of a range from a low to high, or low cost to high cost, series, is too narrow (Manton and Woodbury 1992, 1993; Tolley et al. 1993; Preston 1993; Olshansky 1988; Myers 1981; Alho 1991; Ahlburg and Vaupel 1990). The most recent projections of the SSA have been criticized both for being too optimistic in the face of the deceleration of the rise of life expectancy during recent decades and for being too pessimistic in the face of the levels already attained by Japan.

Analyzing errors in past forecasts. To improve on merely providing low and high projections to accompany a middle series and to describe the accuracy of the middle series more precisely, calls have been made for the two federal agencies to provide a range which can be expected to represent a specified level of confidence for the middle series. Providing confidence intervals for projections offers policymakers a measure of the probability that alternative scenarios may occur in addition to the middle or intermediate forecasts. To do this, we begin by summarizing the errors in past forecasts of national populations.

Two of the many possible parameters for measuring such errors are the percent error in the forecast of the total population (or specific age group or age ratio) and the absolute error in the rate of growth of the population. The first is represented by

$$\text{Percent error in a population forecast}: \quad r_i = (P_f - P_a) \div P_a \tag{14.6}$$

The second is represented by

$$\text{Absolute error in population growth rate}: \quad r_i = (r_f - r_a) \tag{14.7}$$

Among the other possible measures of error, we can consider the mean of the errors, the root-mean-square error, the standard deviation of the error, and the mean absolute error (Keyfitz 1981). The first measure listed averages the errors of all the population forecasts in a set and provides an estimate of the bias in the set of forecasts. Assuming, for illustration, three observations in a set,

$$r_M = (r_1 + r_2 + r_3) \div 3, \text{ or } \sum r_i \div n \tag{14.8}$$

The second measure combines both the bias and the variance in the forecasts and gives increased; weight to the larger errors:

$$r_{RMSE} = \sqrt{\{(r_1)^2 + (r_2)^2 + (r_3)^2\} \div 3}, \text{ or } \sqrt{\left\{\sum r_i^2 \div n\right\}} \tag{14.9}$$

The third measure, the standard deviation of the errors, represents (the square root of) the variance in the forecasts (with greater weight for the larger errors),

$$r_{SD} = \sqrt{\{(r_1 - r_M)^2 + (r_2 - r_M)^2 + (r_3 - r_M)^2\} \div 3}, \text{ or } \sqrt{\left\{\sum (r_i - r_M)^2 \div n\right\}} \tag{14.10}$$

and the fourth measure, a figure for the mean absolute error, or the gross error:

$$r_{MAE} = \{|r_1| + |r_2| + |r_3|\} \div 3, \text{ or } \sum |r_i| \div n \tag{14.11}$$

The analyst is advised to measure the errors separately for major age groups and to take separate account of the length of the projection period, the rate of growth, and the size of the population. Inasmuch as the projections for newborn cohorts involve projections of all the components, they are subject to greater error than the projections of older persons. Moreover, we should expect the length of the projection period to be associated with greater error. Test experience has also shown that the accuracy of national and regional projections is directly associated with their growth rates.

Uses and limitations of stochastic forecasts. As suggested, many analysts believe that it is difficult, if not impossible, to evaluate precisely the accuracy of the projections of mortality and population of the U.S. Census Bureau and SSA because these agencies have not presented their projections in probabilistic terms. They would like

to see probability statements attached to the various series analogous to statements of sampling error and believe that summarizing past errors is an inadequate basis for deriving the probability statements needed. Earlier I described the derivation of one such set of stochastically based forecasts – those prepared by Lee and Carter (1992).

The confidence intervals provided by stochastic forecasting may indicate that policies that may succeed on the basis of the intermediate projections may have a high probability of failure when likely alternatives scenarios are considered (Heller 2003:78). Furthermore, stochastic forecasting allows the user to determine the relative extent of uncertainty in different segments of the forecasts, such as age groups and components of change, and ratios among the segments, such as age dependency ratios. It is important to recognize, too, that the uncertainties associated with demographic variables may have to be considered in combination with uncertainties in the economic or other variables employed in a given application, such as when projecting the financing of the Social Security system or projecting national health-care costs (Heller 2003:79). The advantage of stochastic forecasting can be overestimated, however. Inasmuch as stochastic forecasting relies on the historical record of the variance of the population or of the principal variables (i.e., births and deaths), the error structure determined by this method for the projections of fertility, mortality, and population reflects the error structure of the past, just as any summary of past forecast errors. Stochastic forecasting depends essentially on the use of the same types of data and the same methods in the future as in the past, and it assumes the same general societal context as well. Insmuch as the last condition does not commonly apply in the long run, the projections can be viewed as somewhat unrealistic.

In sum, in designing assumptions for the future, continuing the trends of the past cannot allow sufficiently, if at all, for changes in the pace or even the direction of the projected variables. Mortality trends can change in unanticipated ways, such as is reflected in the 1918 flu epidemic, the sharp decline in several endogenous causes of death after 1970, and the AIDS and obesity epidemics of the 1980s, 1990s, and 2000s. Moreover, there is the prospect that unanticipated, significant advances could occur in the biology of aging that would contribute to far more rapid declines in death rates than now anticipated.

Projections of Health Conditions

Use and Availability of Health Projections

Health projections may be prepared both in their own right and as an intermediate step in the preparation of mortality and population projections. Forecasts (i.e., predictions) of the health status of the population are needed for many purposes, including the planning and implementation of public health programs, market planning by pharmaceutical and medical-equipment firms, and the evaluation of proposals for reform of the health system. Health projections are thought to be useful in making mortality projections because they take into account information

about the series of events leading to death, identifying and quantifying the risk factors and health conditions that contribute to death. Moreover, making health projections as a prior step in making mortality and population projections assures a reasonable consistency between these three types of projections, given their cause/effect relation. Whether health projections are useful for anticipating turning points (i.e., a change of direction or rate of change) in health, mortality, and population growth, as claimed by Myers (1981) and Manton and Woodbury (1992) is doubtful, as is any claim that calculating health projections for the purpose of making mortality and population projections would certainly increase the accuracy of the resulting population projections. There is no evidence to support such claims at the present time.

No federal agency systematically prepares projections of the general health of the U.S. population, either as an end-product or as a basis for mortality projections. It is likely that some agencies concerned with particular health conditions have made forecasts of the number of persons having the particular condition, e.g., Alzheimer's disease, diabetes, and visual and hearing disabilities. The World Health Organization and the World Bank made projections of the leading health conditions in the world on a single occasion. As mentioned in earlier chapters, in collaboration with the Harvard School of Public Health, these organizations published a comprehensive study of the "Global Burden of Disease" in 1996 that included projections of a wide range of health conditions to 2020 (Murray and Lopez 1996). In this study projections of health conditions were presented for males and females for five age groups, disaggregated by cause (107 diseases and injuries), for eight geographic/economic regions. This project also included projections of years of life lost (YLLs), years lived with a disability according to severity and duration (YLDs), and the combination of these measures called disability-adjusted life years, or the burden of disease (DALYs). More recently, Mathers and Loncar (2006) published projections of mortality and the global burden of disease from 2002 to 2030, consistent with those published earlier in the "Global Burden of Disease" study by Murray and Lopez.

Methods of Health Projections

Some methods of projecting the numbers of persons having a particular health condition, such as the prevalence-ratio method, depend on the prior availability of population projections. In such cases the population projections and the associated mortality projections are made independently of the health projections. The methods of projecting the health status of the population resemble those used for projecting mortality. The methods may be outlined as follows:

General analytic methods

- Demographic methods
  - Prevalence-ratio method
  - Event-exposure method
  - Cohort-component method; cohort-progression method

Epidemiological methods

Cause-specific methods
Reverse forecasting from mortality
Cause-age-delay method

Statistical methods

Time-series methods: ARIMA/Box-Jenkins methods; exponential smoothing
Regression/econometric/structural equation methods

Combinations of analytic and statistical methods

Stochastic epidemiological forecasting
Multivariate stochastic health process/biomedical model

The methods overlap and are not mutually exclusive. All the methods use some form of statistical or mathematical analysis to project the series of health measures; it is unusual to hold a series constant at its present level except for analytic purposes. For example, stochastic epidemiological forecasting uses time series methods for analyzing past trends and projecting them. Except for stochastic epidemiological forecasting and stochastic health process modeling, the methods listed deal with aggregate data and produce deterministic results.

Demographic methods. The list includes three types of demographic methods for generating national projections of health conditions: The prevalence-ratio method, the event-exposure method, and the cohort-progression method. All of these methods directly employ data on the health condition or disability to be projected, without resorting to data on related or prior conditions, such as the covariates or risk factors contributing to the health condition.

Demographic methods: Prevalence-ratio method. In the simplest version of the prevalence-ratio method, age-sex specific percents of persons with a chronic illness or disability in the population are projected and applied to population projections disaggregated by age and sex. The method depends on the availability of an historical series of age-sex-specific prevalence ratios of chronic illness or disability and of population projections for age-sex groups.

$$\text{Observed prevalence ratio} = (\text{number disabled}) \div (\text{midperiod population}) * 100 \qquad (14.12a)$$

$$\text{Projected prevalence ratio} = (\text{number disabled} \div \text{midperiod population}) * 100 \qquad (14.12b)$$

$$\text{Projected number disabled} = (\text{projected prevalence ratio} \div 100) * \text{projected population} \qquad (14.12c)$$

The historical series of ratios for the United States may be obtained from sample surveys (e.g., NHIS, ACS, SIPP, NLTCS) or recent censuses (i.e., 1980, 1990, and 2000). More detailed projections of this type would disaggregate the disabled

population into types of disability and categories associated with disability status, such as institutional status, education, and marital status. An illustrative measure then would be the ADL disability prevalence ratio for high school graduates.

As mentioned earlier in connection with the discussion of mortality projections, the methods of projecting a time series of prevalence ratios vary from simple mathematical extrapolation of the series of recent observations to relatively complex statistical methods. This process may involve a transformation of the original series. In the ARIMA design the annual observations are first transformed into a stationary or near-stationary series by taking first or second differences, converting the data to natural logarithms, taking moving averages, or employing some other similar method. The transformed series can then be projected by such methods as linear extrapolation, exponential smoothing,[7] or a random walk. The logarithms of the original series of age-specific prevalence ratios can be projected, for example, by linear extrapolation. The original time series can also be projected by fitting a logistic curve.

In view of the enormous uncertainty attaching to forecasts of health status, there is considerable merit in employing simple methodologies that are readily reproduced and understood by the users. A set of forecasts based on variations of the prevalence-ratio method was issued by the U.S. Interagency Forum on Aging-Related Statistics (1996). Projections were prepared of elderly disabled persons according to the severity of the disability (ADL, IADL) and institutional residence, elderly persons with selected chronic diseases and impairments, and elderly persons according to self-reported health status. Principal assumptions adopted in making these projections included holding the current prevalence ratios for age, sex, and race groups constant, and varying the detailed ratios by a 10% decline, a 5% decline, and a 5% increase every 10 years.

As with other projections, the series employing the prevalence-ratio method is not always intended to represent forecasts or predictions. The purpose of preparing the projections may be analytical rather than predictive. They may be designed to measure the consequences, for the number of ill or disabled persons, of using extreme assumptions. For example, the series may represent the highest or lowest trajectories of disability and population that are likely to occur. These types of sensitivity calculations are sometimes carried out in connection with the financing of a disability retirement system or the planning of services for disabled persons.

---

[7] In exponential smoothing the members of the series are averaged with shifting weights, the greater weights being given to the more recent ratios. Specifically, the weights for the observations decrease exponentially as the observations recede farther into the past:

$$P = aX_t + a(1-a)X_{t-1} + a(1-a)^2X_{t-2} + a(1-a)^3X_{t-3} + \ldots\ldots$$

where P is the projected ratio, a is the smoothing constant, and X_t is the past observation at time t. Determining the smoothing constant requires some experimentation but a value of 0.4 may serve the purpose. It must be between zero and 1. Once the smoothing ratios are determined by the formula, they are held constant for all future years.

Lutz and Scherbov (2003) prepared a set of alternative scenarios for the European Union to test the question whether population aging would necessarily lead to an increase in the number of persons with disabilities. For this purpose, they prepared several series of projections of the number of disabled persons for the European Union on the basis of several sets of projected age-specific disability ratios and population projections for the Union. The projections of disability ratios were generated by a method that corresponds to the age-cause-delay epidemiological method. The population projections were fixed and the age-specific disability ratios were systematically varied by the age-cause-delay method. The projected age-specific disability ratios for each future calendar year were derived by shifting the ratios to higher ages by 0, 1, 2, or 3 years of age per decade, or 0.0, 0.1, 0.2, or 0.3 year of age per calendar year. For example, according to the second assumption, the disability ratio originally for age 70 in year 2000 would become the disability ratio for age 71 in 2010; according to the third assumption, the disability ratio for age 70 in the year 2000 would become the disability ratio for age 72 in 2010.

The expectation was that, given the considerable aging of the population of Europe anticipated in the next half century and the evidence of the greater prevalence of disability among older persons than younger persons, the overall prevalence of disability would increase greatly under any of the series of disability ratios. The results show, however, that if the disability ratios are shifted to higher ages by 1, 2, or 3 years per decade, a relatively low increase, no increase, or a decline, respectively, would occur in the number of persons with disabilities in Europe in the next several decades. In other words, the aging of the population would not necessarily result in an increase in the number of disabled persons, but it may well do so.

Demographic methods: Event-exposure method. Another measure for projecting a health variable is the incidence rate, that is, an event-exposure rate. A morbidity or disability incidence rate is the ratio of persons newly afflicted with a given illness or disability during a specified period to the number of persons at risk of incurring the illness or disability at the beginning of the period. The denominator may be the entire population, or the population excluding those who have the illness or disability, at the start of the period.

To illustrate the derivation of projections with this measure, I cite the projections of disability incidence rates made by the U.S. Actuary's Office, Social Security Administration (2004), in connection with its research program on the use of stochastic methods in projecting the population. The "event" population, or numerator, is the population that becomes newly entitled to disability benefits during a year and the exposed population, or the denominator, consists of workers who are disability insured and not receiving disability insurance at the beginning of the year. Disability incidence rates for the years 1970 through 2003, for each sex, were obtained from SSA administrative records and age-adjusted on the basis of the "exposed" age distribution of the male and female population of 1996 as

the standard. Using time series analysis, the male and female series were modeled separately as ARIMA processes, on the basis of the following equation:

$$\mathrm{DIR_t} = \mathrm{DIR_t^{TR}} + .47\mathrm{dir_{t-1}} - 0.63\mathrm{dir_{t-2}} + \mathrm{E_t} \tag{14.13}$$

where DIR_t represents the disability incidence rate in year t, DIR_t^{TR} represents the disability incidence rate in a prior set of projections derived by less refined methods, dir_t represents the deviation of the disability incidence rate from the disability incidence rate in year t, and $\mathrm{E_t}$ is the random error in year t.[8] R-squared for the male and female fitted equations (i.e., the coefficients of correlation squared) were 0.89 and 0.87, respectively.

Demographic methods: Cohort-component/cohort-progression method. Suppose that you have a time series of estimates of persons who are chronically ill from a given cause for single ages for a series of calendar years, or 5-year age groups at 5-year intervals. Then, you have the data required to construct a set of age-sex specific morbidity progression ratios based on absolute numbers of persons with a chronic health condition for one or more recent periods. From these ratios you can construct some rough short-term projections by the cohort-progression method of the number of persons who will have the disease. Suppose that, in addition, you have estimates of persons becoming ill from the disease and of persons recovering from it for the same series of past years. Then, you have the data required to construct a set of age-sex specific morbidity progression ratios based on morbidity incidence rates for one or more recent periods. You also have the basic historical data needed to prepare more refined health projections by the cohort-progression method.

The simpler application of the cohort-progression method begins with the series of ratios of the number of persons at age a in year y with a disease to the number of persons at age $a-1$ in year $y-1$ with the disease, i.e., annual progression ratios, for a birth cohort:

$$^{y}\mathrm{P_a} \div {}^{y-1}\mathrm{P_{a-1}}, \text{ for a one-year cohort} \tag{14.14a}$$

$$^{y}\mathrm{P_{a,a+4}} \div {}^{y-5}\mathrm{P_{a-5,a-9}}, \text{ for a 5-year cohort.} \tag{14.14b}$$

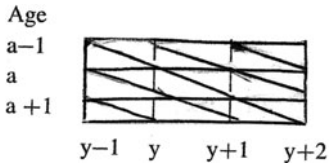
Given these ratios for the latest observed period, and assuming in the simplest case that they are held constant for the projection period, the single-age ratio, ${}^{y}\mathrm{P_a} \div {}^{y-1}\mathrm{P_a}$, would be applied to ${}^{y}\mathrm{P_{a-1}}$ of the current year to derive ${}^{y+1}\mathrm{P_a}$; ${}^{y+1}\mathrm{P_{a+1}} \div {}^{y}\mathrm{P_a}$ to the projected figure for ${}^{y+1}\mathrm{P_a}$ to derive ${}^{y+2}\mathrm{P_{a+1}}$; and so on.

$$^{y+1}\mathrm{P_a} = (^{y}\mathrm{P_a} \div {}^{y-1}\mathrm{P_{a-1}}) * {}^{y}\mathrm{P_{a-1}} \tag{14.15a}$$

$$^{y+2}\mathrm{P_{a+1}} = (^{y+1}\mathrm{P_{a+1}} \div {}^{y}\mathrm{P_a}) * {}^{y+1}\mathrm{P_a} \tag{14.15b}$$

[8] The series had been examined for stationarity and, though found to be nonstationary, was modeled without being corrected. The reason for this choice was that it was assumed that the historical series fluctuated on the basis of changes in the law rather than on the basis of demographic or socioeconomic factors.

Diagrammatically, for birth cohorts (parallelograms), we have



This example employs net cohort changes but the cohort changes could theoretically be disaggregated into components, such as deaths, new cases or disablements, and recoveries, depending on the availability of data and the health condition studied. Furthermore, the net changes could be measured in terms of ratios of rates instead of absolute numbers of cases as above. Accordingly, the prevalence or incidence ratio for age a, year y, could be related to the prevalence ratio for age a-t, year y-t, to derive a series of such ratios. This series could then be projected by one of the methods cited earlier.

Epidemiological model. Like the demographic methods, epidemiological methods use prevalence or incidence ratios, or similar measures, for diseases and disabilities from cross-sectional repeated or longitudinal surveys to estimate the past impact of a disease or health condition on a population. However, they often go further and take account of the linkages between risk factors, prior diseases, and covariates (i.e., associated and explanatory characteristics or events), on the one hand, and health conditions, on the other.

Future changes in the health status of the population may be generated by extrapolating an historical series of prevalence or incidence ratios by the age-cause-delay technique (Olshansky 1987). I consider this an epidemiological method, albeit a simple one, because it is based on an assumption as to the rate of progress in reducing chronic-disease prevalence or incidence. The age-cause-delay technique, or decelerated-aging technique, shifts each cause-specific morbidity ratio forward for a birth cohort by a given number of ages – 2 years, 3 years, ..., 7 years – for the same number of calendar years, according to the assumption made. Different numbers of ages and periods could be selected for different health conditions. One scenario would be to shift the age-specific prevalence ratios forward 3 years over a decade. Any such assumption implies a delay or postponement in the age of onset of the disease and a rise in the mean age at death and life expectancy.

The Future Elderly Model (FEM) of the Rand Corporation (Zheng et al. 2007) illustrates another type of epidemiological model employed in deriving projections of persons incurring a disease during a future period and living with it some future date. FEM is a microsimulation model that tracks a representative sample of the residents of the United States 51 years and over, disaggregated by gender, race, and education, with respect to their health and survival status, on the basis of data from the U.S. Health and Retirement Survey. First, transition probabilities for the health conditions are derived from the most recent waves of the Health and Retirement

Survey. To compute the transition probabilities, comparable data are needed at regular intervals. For FEM, data from the wave of the HRS taken in 2004 were linked with data for several prior biennial intervals.

The health status of an individual was defined by a number of self-reported health indicators, such as the presence of six chronic conditions (i.e., hypertension, diabetes, cancer, lung disease, heart disease, and stroke), five functional statuses (i.e., no limitations, IADLs only, one ADL, two ADLs, and three or more ADLs), smoking status, and obesity status. The status of each person was also defined in terms of a few economic indicators. Transition into the chronic conditions was modeled by a proportional hazards model and transition among the various functional statuses was modeled by multinomial logit regressions following a Markov process. For transition to mortality and nursing homes, logit regressions were also applied. In addition to the basic "forecasting module," an "incoming cohort module" produced estimates of the size of cohorts newly arriving at age 51 and their characteristics. Since the basic data pertain to biennial intervals in this application of the cohort-progression method, the forecasts were obtained for biennial intervals.

Stochastic forecasting of health status is a probabilistic method of forecasting, paralleling that used for mortality projections. This approach contrasts with most methods of projecting health status described above, which are discrete-state and discrete-time methods of projection. The use of stochastic simulation makes possible the production of a range of projections for which confidence limits can be determined. (See section "Mortality Projections" for further discussion of the stochastic method.)

Econometric/structural equation model. In the econometric model, regression analysis is the basis for the projections of health status. A system of simultaneous linear multivariate equations links health status with several independent variables. Health-status prevalence ratios or incidence ratios over a specified past period serve as the dependent variable to which are related several independent explanatory variables, such as health risks, biomarkers, socioeconomic status, and medical technology. Once the parameter values are determined from the various multivariate equations using historical data, one can develop projections of the dependent health variables by inserting the parameter values and projections of the independent explanatory variables into the regression equations.

Multivariate stochastic health process/biomedical model. The stochastic health process model, the final model I describe, requires biologically detailed models of change, as the title suggests. It was developed by Manton (1984), Manton and Woodbury (1992) and Manton and Stallard (1993, 1994). It models health outcomes as a multivariate stochastic process subject to exogenous influences such as public health interventions and improved health-care access, endogenous influences such as biological and genetic risk factors, and stochastic generation of DNA mutations with age. A selected set of variables from a series of health and epidemiological surveys, and vital registrations, are used to estimate the parameters

of the regression equations representing the interactions between disease, disability, and mortality. The risk factors and biomarkers include sex, diastolic blood pressure, serum cholesterol, hemoglobin, smoking, and blood sugar. The relation of age-cause specific mortality and the various risk factors is determined and then future health changes are simulated by modifying the parameters determining the age pattern of risk factors.

As Manton and Woodbury (1992) describe their method, population health is forecast as a two-sex component multivariate stochastic model of aging and mortality. One component describes the changes in multiple health-status variables (i.e., risk factors and biomarkers). The second component describes the probability of dying as a function of the values of the health-status variables stochastically generated in the first process. That is, the analysis first identifies risk factors and other variables important for mortality at the later ages (e.g., loss of lung function due to smoking; the effect of antihypertensive drugs on the age trajectory of blood pressure). Simulations based on the observed relations of risk factors to mortality are then carried out. Interactions of the risk factors ("competing risks") are taken into account, particularly the numerous interactions common in later life.

Note on Uncertainty

We have seen that it is quite difficult to gauge current health status accurately. Self-reported health obtained in sample surveys, the most common basis for measuring current health status, is subject to substantial error. Projections of health status are even more tenuous. Health projections are subject to considerable error both because of the tenuousness of the projections of health risk as well as the uncertainty of projections in general. They have some merit, however, because they take account of the current and historical estimates of the health status of the population and serve a great variety of important administrative and research uses.

References and Suggested Readings

Aging, Health, and Mortality Models

Alter,G. C., & Riley, J.(1989). Frailty, sickness and death: Models of morbidity and mortality in historical populations. *Population Studies*, 43, 25–45.

Callahan, C. M., & McHorney, C. A. (Eds.). (2003). Successful aging and the humility of perspective. A Supplement to the *Annals of Internal Medicine*, 139. Papers of the Eighth Biennial Regenstrief Conference.

Carnes, B. A., Olshansky, S. J., & Grahn, D. (2003). Biological evidence for limits to the duration of life. *Biogerontology*, 4, 31–45.

Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *The New England Journal of Medicine*, 303, 130–135.

Fries, J. F. (1987). Reduction of the national morbidity. *Gerontologica Perspecta*, 1, 50–61.

Fries, J. F. (1989). The compression of morbidity: Near or far? *Milbank Quarterly*, 67(2), 208–232.
Fries, J. F. (2003). Measuring and monitoring success compressing morbidity. *Annals of Internal Medicine*, 139, 455–459.
Fries, J. F., & Crapo, L. M. (1981). *Vitality and aging: Implications of the rectangular curve*. San Francisco: W.H. Freeman and Company.
Fuller-Thomson, E., Yu, B., Nuru-Jeter, A., Guralnik, J. M., & Minkler, M. (2009). Basic ADL disability and functional limitations rates among older Americans from 2000–2005: The end of the decline? *The Journals. of Gerontology*, 64(12), 1333–1336.
Gruenberg, E. M. (1977). The failure of success. *Milbank Memorial Fund Quarterly Health and Society*, 55, 3–24.
Horiuchi, S., & Wilmoth, J. R. (1998). Deceleration of the age pattern of mortality at older ages. *Demography*, 35, 391–412.
Inui, T. S. (2003). The need for an integrated biopsychosocial approach on successful aging. *Annals of Internal Medicine*, 139, 391–394.
Kannisto, V. (1991). Frailty and survival. *Genus*, 47, 101–118.
Martin, G. (2003). *Gerontology News* 91(5):2.
Masoro, E. J. (2001). 'Successful aging' – Useful or misleading concept? Book review. *Gerontologist*, 41(3), 415–418.
McFadden, S. H. (2002). Challenges and opportunities in the search for new models of aging. Book review. *Gerontologist*, 42, (5):705–708.
Morrow-Howell, N., Hinterlong, J., & Sherraden, M. (Eds.). (2001). *Productive aging: Concepts and challenges*. Baltimore: Johns Hopkins University Press.
Myers, G. C., & Manton, K. G. (1984). Compression of mortality: Myth or reality? *The Gerontologist*, 24(4), 346–353.
Olshansky, S. J., Rudberg, M. A., Carnes, B. A., Cassell, C. K., & Brody, J. A. (1991). Trading off longer life for worsening health: The expansion of the morbidity hypothesis. *Journal of Aging and Health*, 3, 194–216.
Robine, J. M., Cheung, S. L. K., Horiuchi, S., & Thatcher, A. R. (2008a, January 7–9) *Is the compression of morbidity a universal phenomenon*? In Society of Actuaries (Eds.), International symposium on living to 100: Survival to advanced ages, Lake Buena Vista, FL.
Robine, J. M., Cheung, S. L. K., Horiuchi, S., & Thatcher, A. R. (2008b, January 7–9). *Is there a limit to the compression of mortality*? In Society of Actuaries (Eds.), International symposium on living to 100: Survival to advanced ages, Lake Buena Vista, FL.
Rothenberg, R., Lentzner, H. R., & Parker, R. A. (1991). Population aging patterns: The expansion of mortality. *Journal of Gerontology: Social Sciences*, 46(2), S66–S70.
Rowe, J. W., & Kahn, R. L. (1987). Human aging: Usual and successful. *Science*, 237, 143–149.
Rowe, J. W., & Kahn, R. L. (1998). *Successful aging*. New York: Pantheon Books.
Schneider, E. L., & Brody, J. (1983). Aging, natural death, and the compression of morbidity: Another view. *The New England Journal of Medicine*, 309, 854–856.
Schneider, E. L., & Guralnik, J. M. (1987). The compression of morbidity: A dream which may come true, someday! *Gerontologica Perspecta*, 1, 4–10.
Siegel, J. S. (2005, January 12–14). *The great debate on the outlook for human longevity: Exposition and evaluation of two divergent views*. In Society of Actuaries (Eds.), International symposium on living to 100 and beyond. Lake Buena Vista, FL.
Strehler, B. L. & Mildvan, A. S. (1960). General theory of mortality and aging. *Science*, 132, 14–21.
Tate, R. B., Lah, L., & Cuddy, T. E. (2003). Definition of successful aging by elderly Canadian males: The Manitoba Follow-Up Study. *Gerontologist*, 43(5): 735–744.
Verbrugge, L. (1984). Longer life but worsening health? Trends in health and mortality of middle-aged and older persons. *Milbank Memorial Fund Quarterly Health and Society*, 62, 474–519.
Wilmoth, J. R., & Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human population. *Demography*, 36(4), 475–496.∗ ∗ ∗

The Aging vs. Disease Dichotomy

Blumenthal, H. T. (2003). The aging-disease dichotomy: True or false. *Journal of Gerontology: Medical Sciences*, 58A(2):138–145.
Butler, R. N. (1977). Preface. In T. Makinodan & E. Yunis (Eds.), *Immunology and Aging* New York: Plenum.
Fozzard, J. L., Metter, E. J., & Brandt, L. J. (1990). Next steps in describing aging and disease in longitudinal studies. *Journal of Gerontology: Psychological Sciences*, 45, 302–311.
Hall, D. A. (1984). *The biomedical basis of gerontology*. London: Wright-PSG.
Hayflick, L. (2001). The future of aging. *Nature*, 408, 267–269.
Holliday, R. (1988). Toward a biological understanding of the ageing process. *Perspectives in Biological Medicine*, 32, 109–121.
Kleemeier, R. W. (1965). Infinitely eliminable? *Contemporary Psychology*, 10, 53–55.
Ling, S. M., Simonsick, E. M., & Ferrucci, L. (2007). Editorial. A painful interface between normal aging and disease. *Journal of Gerontology: Medical Sciences*, 62A(6), 613–615.
Ludwig, F. C. (1980). Editorial. What to expect from gerontological research. *Science*, 209, 107.
Rattan, S. I. S. (1991). Ageing and disease: proteins as the molecular link. *Perspectives in Biological Medicine*, 34, 526–533.
Strehler, B. L. (1977). *Time, cells and aging* (2nd ed.). New York: Academic.
Williams, T. F. (1992). Aging versus disease. *Generations*, Fall/Winter 21–25.

Mortality and Health Projections

Ahlburg, D. A., & Land, K. C. (Eds.). (1992). Population forecasting. *International Journal of Forecasting*, Special Issue 8(3), 289–299.
Ahlburg, D., & Vaupel, J. (1990). Alternative projections of the U.S. population. *Demography*, 27(4), 639–652.
Alho, J. M. (1991). Stochastic methods in population forecasting. *International Journal of Forecasting*, 6, 521–530.
Bongaarts, J. (2004). Long-range trends in adult mortality: Models and projection methods. *Policy Division Research Working Paper* no.192. New York: Population Council. Reprinted in*Demography*, 42(1), 23–49 (February 2007).
Booth, H., Maindonald, J., & Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 56, 325–336.
Carnes. B.A., Olshansky S.J., & Grahn, D. (2003). Biological evidence for limits to the duration of life. *Biogerontology*, 1(4), 31–45.
de Grey, A. D. N. J., Ames, B. N., Anderson, J. K., et al. (2002). Time to talk SENS: critiquing the immutability of human aging. *Annals of the New York Academy of Sciences*, 959, 452–462.
Demeny, P. (2004). Population futures for the next three hundred years: Soft landing or surprises to come? *Population and Development Review*, 30(3), 507–517.
Fraser, G. E., & Shavlik, D. J. (2001). Ten years of life: Is it a matter of choice? *Archives of Medicine*, 161, 1645–1652.
Freedman, V. A., & Martin, L. G. (2002). The role of education in explaining and forecasting trends in functional limitations among older Americans. *Demography*, 36, 461–473.
Guarente, L., & Kenyon, C. (2000). Genetic pathways that regulate aging in model organisms. *Nature*, 408, 255–262.
Guralnik, J. M., Yanagishita, M., & Schneider, E. L. (1988). Projecting the older population of the United States: Lessons from the past and prospects for the future. *Milbank Quarterly*, 66(2), 283–308.

Hall, S. S. (2003). *Merchants of immortality: Chasing the dream of human life extension*. New York: Houghton Mifflin.

Heller, P. S. (2003). *Who will pay? Coping with aging societies, climate change, and other long-term fiscal challenges*. Washington, DC: International Monetary Fund.

Keyfitz, N. (1977). *Introduction to the mathematics of population, with revisions*. Reading, MA: Addison-Wesley.

Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review*, 7(4), 579–593.

Kunkel, S., & Applebaum, R. (1992). Estimating prevalence of long-term disability for an aging society. *Journal of Gerontology: Social Sciences*, 47(5), S273–S260.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting United States mortality. *Journal of the American Statistical Association*, 87, 659–671.

Lutz, W., & Scherbov, S. (2003). *Will population ageing necessarily lead to an increase in the number of persons with disabilities? Alternative scenarios for the European Union* (European Demographic Research Papers. 2003, No. 3). Vienna: Institute of Demography.

Manton, K. G. (1984). The application of disease-specific models for health trend projections. *World Health Statistics Quarterly*, 3.

Manton, K. G., & Stallard, E. (1993). Projecting morbidity and mortality in developing countries during adulthood. In J. Gribble & S. Preston (Eds.), *The epidemiological transition: Policy and planning implications for developing countries* (pp. 101–125). Washington, DC: National Academy Press.

Manton, K. G., & Stallard, E. (1994). Medical demography: interaction of disability dynamics and mortality. In L. G. Martin & S. H. Preston (Eds.), *Demography of aging* (pp. 217–278). Washington, DC: National Academy Press.

Manton, K. G., Patrick, C., & Stallard, E. (1990). Mortality model based on delays in progression of chronic diseases: Alternative to cause-elimination model. *Public Health Reports*, 95, 580–588.

Manton, K. G., Stallard, E., & Tolley, H. (1991). Limits to human life expectancy: Evidence, prospects, and implications. *Population and Development Review*, 17, 603–637.

Manton, K., Stallard, E., & Singer, B. (1992). Projecting the future size and health status of the U.S. elderly population [Special Issue]. *International Journal of Forecasting*, 8(3), 433–458.

Manton, K. B., Singer, B. H., & Suzman, R. M. (1993). The scientific and policy needs for improved health forecasting models for elderly populations. In K. G. Manton, B. H. Singer, & R. M. Suzman (Eds.). *Forecasting the health of elderly populations* (pp. 3–35). New York: Springer.

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease: 2002 to 2030. *PLoS Medicine*, 3(11), e442.

McNown, R. F., & Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal. of Forecasting*, 8(3), 413–432.

Murray, C. J. L., & Lopez, A. D. (Eds.). (1996). *Summary: The global burden of disease*. Cambridge, MA: Harvard School of Public Health/World Health Organization and the World Bank.

Myers, G. C. (1981). Future age projections and society. In W. M. Beattie, J. Piotrowske, & M. Marois (Eds.), *Aging: A challenge to society, Vol. 2, Part II, Social sciences and social policy* (pp. 248–260). Oxford, UK: Oxford University Press.

Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296, 1029–1031.

Olshansky, S. J. (1987). Simultaneous/multiple cause-delay (SIMCAD): An epidemiological approach to projecting mortality. *Journal of Gerontology*, 42, (4):358–365.

Olshansky, S. J. (1988). On forecasting mortality. *Milbank Quarterly*, 66, 482–530.

Olshansky, S. J., & Carnes, B. A. (1994). Demographic perspectives on human senescence. *Population and Development Review*, 20(1), 57–80.

Olshansky, S. J., Carnes, B. A., & Cassel, C. (1990). In search of Methuselah: Estimating the upper limits to human longevity. *Science*, 250, 634–640.

Olshansky, S. J., Carnes, B.A., & Desquelles, A. (2001). Still in search of Methuselath: Prospects for human longevity in an aging world. *science,* 291, 1491–1492.

Olshansky, S. J., Carnes, B. A., & Grahn, D. (1998). Confronting the boundaries of human longevity. *American Scientist*, 86, 52–61.

Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., et al. (2005). A potential decline in life expectancy in the United States in the 21st century. *The New England Journal of Medicine*, 352, (11):1138–1145.

Olshansky, S. J., Goldman, D. P., Zheng, Y., & Rowe, J. W. (2009). Aging in America in the twenty-first century: Demographic forecasts from the MacArthur Foundation Research Network on an Aging Society. *Milbank Quarterly*, 87(4), 842–862.

Preston, S. H. (1993). Demographic change in the United States, 1970–2050. In K. G. Manton, B. H. Singer, & R. M. Suzman (Eds.), *Forecasting the Health of Elderly Populations* (pp. 51–77). New York: Springer.

Siegel, J. S. (2002). *Applied demography: Applications to business, government, law, and public policy*. San Diego, CA: Academic.

Siegel, J. S., & Swanson, D. W. (Eds.). (2004). *The methods and materials of demography*. Second Edition. San Diego, CA: Elsevier/Academic.

Smith, S., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis*. New York: Kluwer Academic/Plenum Press.

Stoto, M. (1983). The uncertainty of population projections. *Journal of the American Statistical Association*, 78, 13–20.

Stoto, M., & Durch, J. (1993). Forecasting survival, health, and disability: Report of a workshop. *Population and Development Review*, 19(3), 556–581.

Tabeau, E., van den Berg Teths, A., & Heathcote, C. (2002). A review of demographic forecasting models for mortality. Introductory chapter. *In Forecasting mortality in developed countries: Insights from statistical, demographic, and epidemiological perspectives*. Dordrecht, the Netherlands: Springer.

Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society*, 162(Part 1), 5–43.

Thatcher, A. R., Kannisto, V., Vaupel, J. W., et al. (1998). *The force of mortality at ages 80 to 120*. Odense, Denmark: Odense University Press.

Tolley, H. D., Hickman, J. C., & Lew, E. A. (1993). Actuarial and demographic forecasting methods. In K. G. Manton, B. H. Singer, & R. M. Suzman (Eds.), *Forecasting the health of elderly populations* (pp. 39–49). New York: Springer.

Torri, T., & Vignoli, D. (2007). Forecasting the Italian population, 2005–2055: A stochastic approach. *Genus, LXIII*, (1–2): 93–118.

Tuljapurkar, S., & Boe, C. (1998). Mortality change and forecasting: how much and how little do we know? *North American Actuarial Journal*, 2, 13–47.

Tuljapurkar, S., Li, N., & Boe, C. (2000). A universal pattern of mortality decline in G7 countries. *Nature*, 405, 789–792.

United Nations (UN). (2003). *World population prospects: The 2002 revision. Vol. III. Analytical report*. New York: United Nations Department of Economic and Social Affairs. Also www:unpopulation.org

United Nations (UN). (2007). *World population prospects: The 2006 revision*. New York: CD Extended Data. Population Division, Department of Economic and Social Affairs, United Nations.

U.S. Census Bureau. (2004). Accessed July 2007, from www.census.gov

U.S. Census Bureau. (2009). U.S. Population projections: Methodology. United States population projections by age, sex, race, and Hispanic origin: July 1, 2000–2050. Accessed July 2009, from www.census.gov/population/www/projections

U.S. Interagency Forum on Aging-Related Statistics. (1996). *Projections of health status and use of health care of older Americans*. Occasional Paper from the National Center for Health Statistics, Centers for Disease Control and Prevention.

U.S. Actuary's Office Social Security Administration. (2004). A stochastic model of the long-range financial status of the OASDI program. *Actuarial Studies, No. 117. By Cheng, A.W., Miller M.L., Morris, M., Schultz, J.P., et al.*

U.S. Social Security Administration, Actuary's Office. (2008). Accessed April 2008, from www.ssa,gov/OACT/index.html

U.S. Technical Panel on Assumptions and Methods. (2003). *Report to the Social Security Advisory Board, 2003*. Washington, DC: U.S. Government Printing Office.

Vaupel, J.W., Carey, J. R., Christensen, K., Johnson, T. E., et al.(1998). Biodemographic trajectories of longevity. *Science*, *280*(5365),855–860.

Wilmoth, J. R. (1997). In search of limits. In K. W. Wachter & C. Finch (Eds.), *Between Zeus and the Salmon*: The biodemography of aging. *Washington, DC: National Academy Press*.

Wilmoth, J. R. (2000). Demography of longevity: past, present, and future trends. *Experimental. Gerontology*, 35, 1111–1129.

World Health Organization. (2004). *World health report*. Geneva, Switzerland: World Health Organization.

Zheng, Y., Girosi, F., & Goldman, D. (2007). *The future elderly model*. Unpublished document. Santa Monica, CA: Rand Corporation.

Part V
Applications to Health Policy and Health Administration

Scope of Part V

The final chapters of the book focus on some of the issues that arise in connection with government responses to the findings on health and longevity described in earlier chapters, that is, the uses to which governments put this information. The scope of the subject matter ranges from the choices governments make regarding the provision of health care in the face of limited resources to the preparation by local governments of estimates and projections of health requirements, and analyses of the health data of small areas. It encompasses the choices that individuals, the public, business, and governments make in dealing with some of the vexing ethical and legal questions that arise regarding health and longevity. Chapter 15 relates to health policy, and considers how much health care should be provided, who should pay, and what the relative roles of government, the health provider, and the private citizen should be in managing an individual's health care. Chapter 16 gives a short introduction to applied health demography and deals mainly with the problems of securing and analyzing the data needed for the implementation and evaluation of health programs. Underlying all health policies and programs are legal and ethical issues that fundamentally reflect the values of a particular society. Technological developments and scientific progress continuously force society to confront new legal and ethical issues with respect to living and dying. The exposition of some of these issues and possible solutions is the subject of Chap. 17.

Chapter 15
Health Policy

Official health policy is reflected in a government's laws, tax policies, and public programs relating to health as well as in the public statements of senior government officials regarding the role of government in health affairs. National health policies differ in the extent of the involvement of the national government in health programs, the share of the national budget allocated to health programs, the level of health expenditures per capita, the areas of health publicly supported, and the nature of the national health system. We may assume that all national governments nominally pursue a policy of disease prevention and health promotion, although this policy may only be implicit in public actions taken or programs pursued, and governments may differ widely in how invested they are in such a policy.

The UN report on *World Population Policies* 2003 (UN 2004) lists the issues of most significant demographic concern to the More Developed Countries and the Less Developed Countries. Health is not included for the More Developed Countries. Their main concern relates to low fertility and its consequences, including population aging and the shrinking of the working-age population. For the Less Developed Countries, however, high mortality is the most significant concern. Over 80% of the Less Developed Countries consider infant and child mortality, maternal mortality, and HIV/AIDS as their most urgent population problems. Inasmuch as health situations, the structure of the health systems, and the economic/political/legal conditions differ greatly among the More Developed Countries and the Less Developed Countries, it is useful to consider health policy issues separately for these two groups of countries.

More Developed Countries

Providing for the health care needs of an aging, developed country is a complex, large-scale, societal problem. According to the Medicine/ National Research Council (1985), the two most relevant issues of health care in an aging society

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_15, © Springer Science+Business Media B.V. 2012

are (1) how to understand, prevent, diagnose, and treat the effects of chronic illness and (2) how to design public and institutional policies to meet the nation's needs for long-term care. The Institute maintains that long-term care has to be incorporated fully into the health programs of an aging society. Another goal should be to extend the years of active life well into advanced age, so as to reduce the years of chronic illness, disability, and dependency. The Institute of Medicine/National Research Council (1985) has enumerated the following steps to be taken toward the goal of achieving maximum functional independence of the population:

- Frequent detailed surveys of the physical and mental health and functional status of the population
- Examination of the social and economic costs of chronic illness and dysfunction
- Increased research on distinguishing the effects of disease from the effects of aging
- Greater effort to conceptualize and measure active life expectancy
- Increased research on chronic diseases and dysfunctions and increased efforts at rehabilitation and reversal of the effects of these diseases
- Accelerated research on dementias
- Study of the resources for long-term care and related services for chronic illness and dysfunction
- Studies of the quality of nursing-home care and the outcomes of various types of long- term care

In sum, these steps call for regular measurement of the health status of the population, research on chronic diseases of later life, particularly dementias, and evaluation of the resources for long-term care, their quality, and the efficacy of their use. Biogerontologists may note that these steps do not emphasize research on the basic causes of sickness and death at the molecular level, geriatricians may note that disease prevention is not a focus of this plan, and medical sociologists may note the absence of attention to socioeconomic inequalities in health.

Health Promotion/Disease Prevention vs. Treatment/Cure of Diseases

For the most part, public policy has favored the treatment of diseases rather than their prevention. For example, Medicare – the national public health insurance program for the elderly – and private insurance and health-management companies have tended to require prior indications of specific illness for reimbursement of health claims. Many diagnostic tests and preventive procedures are not covered unless this requirement is met.

"Healthy People 2010," a comprehensive nationwide agenda for health promotion and disease prevention of the American people promulgated by the U.S. Department of Health and Human Services, Public Health Service, serves as a

roadmap for improving the health of the population of the United States during the first decade of the twenty-first century. The goals of Healthy People 2010 are twofold, "to increase life expectancy and improve the quality of life" by 2010 and "to eliminate health disparities among different segments of the population," i.e., the sexes, socioeconomic groups, and race/ethnic groups in the United States (U.S. DHHS 2000). Masses of data have been assembled to implement the agenda of Healthy People 2010 and used to construct more than 800 health measures, which are associated with 467 objectives. These official efforts are intended to bring about improvements in the health of the American population while at the same time eliminating all differences in health. Progress toward these goals would be carefully monitored.

These goals taken in combination are hardly realistic or achievable. There are several barriers to the program's success. One general problem is that a program designed to achieve one goal may contribute to the failure of the other. The goals may compete with and offset one another. More specifically, a program designed to achieve a greater measure of health for the general population risks not only an increase in group health inequalities but also the possibility that the goals with respect to the reduction of specific diseases and disabilities may conflict with each other, given the competing risks of death. For example, meeting targets for reductions in, say, the death rate from heart disease could result in patients' spending a more extended period of time with chronic heart disease. Or success in reducing the cancer death rate could indirectly contribute to a rise in the heart disease rate. Furthermore, it is possible that the health status of all segments of the population could improve while the gap between the health status of the most healthy segment of the population and the least healthy segment could widen. We have a disturbing analogy in the data on the income status of the U.S. population. Between 1985 and 2005 the nominal median household income of the American people rose but income inequality widened at the same time. All income segments of the population shared in the general improvement but the income gap between the richest segment of the population and the poorest segment increased. Finally, the goal of eliminating health inequalities between socioeconomic groups would appear to be an impossible dream inasmuch as it is premised on the virtually unrealizable goal of eliminating all socioeconomic differences in the society, i.e., differences in educational level, occupational status, and income.

The American health system tends to emphasize cost protection for the providers of health services and products. At present it supports and encourages the use of drugs as continuing treatment, partly because its orientation is on treatment, not on prevention, and partly because sale of costly drugs is more remunerative than more invasive, comprehensive, and long-term treatments. Preventive programs are more cost-effective in the long term, however. The concern is that the preoccupation with drugs will forestall any effort to move into a preventive mode on the part of health services users and providers.

Adding More Years vs. Improving the Quality of Years Lived

With the dramatic increases that have occurred in life expectation in the last century, it has been important to determine whether these added years are years with an improved quality of life or the opposite, and if the latter is the case, what steps should be taken to improve the outlook for a better quality of life. In Chap. 14 attention was called to the many discomforting and activity-limiting conditions, mainly non-lethal chronic diseases and sensory impairments, that mar the quality of life of untold millions of Americans, not only in the later years but, for some, much of their lifetimes. Among these discomforting and functionally limiting conditions that cry out for improved treatments and cures are osteoarthritis, osteoporosis, insomnia, chronic back pain, irritable bowel syndrome (IBS), vision and hearing loss, edentulism (i.e., loss of all natural teeth), peripheral vascular/arterial disease (PVD or PAD) and the associated intermittent claudication (e.g., pain in calf on walking), benign prostatic hyperplasia (BPH) and the associated nocturia, irritable bladder syndrome, urinary incontinence, and chronic fatigue syndrome. None of these conditions is lethal in itself and only a few are disabling to the extent of being or causing a personal care disability (i.e., ADLs). Public health policy gives little attention to these conditions and health insurance companies tend to dismiss claims for many of them as simply indicators of normal aging or as unjustified complaints, and not real illnesses. Foremost among the more serious conditions that are viewed in the way described are the mental illnesses, especially depression, which destroy or limit the productive capacity of many millions of persons in the United States and the other countries of the world in addition to causing intense suffering.

The general public policy issue is whether to devote public resources mainly to extend years of life or to improve the quality of the years already gained. Some believe that, given the high level of life expectancy and average maximum life span already attained, too little effort is being devoted to improving the quality of life of those afflicted with the various conditions I have enumerated as well as others like them. With a modest amount of effort in these areas, the benefits in improved well-being and in worker and general productivity would be considerable.

Consider the case of peripheral vascular disease (PVD). It is associated with intermittent claudication, a painful condition of the lower leg that is usually experienced while walking and that results from blockages of the arteries in the thigh, knee, or calf. The disease is quite prevalent in the older population, affecting more than 20% of the U.S. population over age 75, and is expected to become far more common as the population grows older. The symptoms of leg pain and impaired mobility seriously affect functional capacity and quality of life. Further, it is a precursor of, and causally associated with, more serious illnesses such as heart disease and stroke. Persons with PVD may be “asymptomatic” (i.e., have no perceived symptoms such as pain or breathlessness) in the early stages or even later, so that millions of older Americans are afflicted with PVD, some in the advanced stages, without being aware of it. A specific policy question relating to this situation is, under what conditions is it cost-effective to test a person for the disease? Then,

assuming a clinical finding of PVD, what should the attending physician advise those who are asymptomatic when the only dependable treatments are expensive and invasive ones?

Similar questions arise with respect to many other conditions, for example, abdominal aortic aneurysm, avascular necrosis of hip, carotid atherosclerosis, and thyroid disease (mostly hypothyroidism), all of which may be asymptomatic or masquerade with symptoms of other diseases, and go undetected by the person. To screen every older person for these conditions by employing the latest technology (e.g., magnetic resonance imaging, angiography, bone scan, special blood tests, Doppler ultrasound, etc.) would be prohibitively expensive, even if the equipment were always available. Moreover, physicians would be reluctant to recommend such tests for preventive/diagnostic purposes only, in the absence of symptoms, because insurance companies tend to deny coverage under these conditions. What balance between risk factors, symptoms, probability of identifying positive cases, and costs should be considered appropriate for testing patients?

Measuring the Cost-Effectiveness of Medical Interventions and Comparative Effectiveness Research

A basic question to be asked is whether the benefits achieved in health care in the United States are worth the costs on average, as measured by some broad indicator such as increases in life expectancy. Cutler and Richardson (1998) and Cutler et al. (2006) maintain that, taking the health-care system as a whole, they are worth the costs. Even if we accept their finding, though it is strongly disputed by some, we may still question whether some health-care practices are worth the cost as compared with others for dealing with specific conditions. In order to answer the question whether we are getting, or would get, more health improvement for the money spent or allocated, from the various preventive, diagnostic, and therapeutic procedures that are being used or contemplated for use, than another possible procedure, we need appropriate measures of the cost-effectiveness of these procedures. Cost-effectiveness analysis has been applied to surgical procedures and similar interventions but, as shown by Weinstein (2005), they can also be applied in guiding the use of new diagnostic equipment and technologies, such as fMRI machines vs. Doppler ultrasound in detecting PVD, DNA- or RNA-typing of infectious agents, or genomic or proteinomic markers in patients. When money is spent on one particular procedure rather than another less costly or more costly one, then, how can we compare the alternative procedures in terms of their health outcomes and their relative cost-effectiveness?

Cost-effectiveness analysis is all the more important at a time of sharply rising health care costs and increasing use of technology because, as will be indicated, the costs of medical technology is expected to be the major component of the rise in health care costs in the future. Cost-effectiveness analysis for various medical technologies requires a combination of the element of costs and measures of health outcomes. Costs refer to the net resources expended in employing the health

procedure, measured in dollars. The cost element includes the total costs of the medical professionals, hospitalization, medications, medical devices, and nursing home care. Health outcomes may be measured in terms of years-of-life added, improvements in the quality of life, or QALYs (i.e., quality-adjusted life years). To calculate QALYs, one needs to assign relative weights to the various possible health conditions that may be experienced after receiving the health intervention, and then take the cumulative product of these weights and the durations of the health conditions. (See Chap. 8 for a further explanation of QALYs.)

Use of a medical technology can increase or decrease costs. A technology may decrease costs by saving enough money to offset the cost required to pay for it. This may be true of either preventive, diagnostic, or therapeutic interventions, and if the procedure yields a health benefit in QALYs, then it ought to be favored for use. Similarly where a procedure costs money and yields a poorer health outcome, then it should be rejected. The problematic area relates to the procedures that cost money and yield a health benefit. The formula that Weinstein (2005) employs for measuring cost-effectiveness of a medical procedure is:

$$\frac{\text{Net gain in health outcome(QALYs)}}{\text{Net increase in health care costs(\$)}} * 1{,}000{,}000 = \text{QALYs per million dollars} \tag{15.1}$$

(The reciprocal of the above ratio, that is, dollars per QALY, has been conventionally used as a cost-effectiveness measure and numerous studies have employed it.) By this type of analysis Weinstein found, for example, that use of beta-blocker medication after a myocardial infarction[1] for patients who are at high risk for subsequent cardiac death is a very cost-effective intervention. More than 200 QALYs are gained for $1 million spent. On the other hand, coronary angiography, a diagnostic test for heart disease that maps the flow of blood through the principal coronary arteries, is not a very cost-effective intervention when used for the purpose of identifying low-risk patients who can benefit from coronary revascularization after a myocardial infarction. It can buy less than 10 QALYs for $1 million spent. The dividing line when an intervention becomes cost-effective was set at 10–20 QALYs gained.

Weinstein (2005) finds that in the United States some highly beneficial, low-cost procedures are significantly underutilized, and that other procedures are overutilized, as shown by the health benefits they provide in relation to their costs. For example, colonoscopies, an invasive screening procedure for colon cancer, and flu vaccinations for persons over age 65 are very cost-effective interventions, but are not very popular and are underutilized. This finding suggests a reallocation of funds

[1]Myocardial infarction is the medical term for a heart attack. It involves necrosis (i.e., death) of heart muscle and other tissues of the heart resulting from a blockage of the blood supply. Coronary angiography involves catheterization of the coronary arteries to determine the extent of blockage of the arteries, and the computer recording of the results. Coronary revascularization/coronary angioplasty is a therapeutic procedure involving catheterization of the coronary arteries designed to treat the blockage in the coronary arteries.

between those spent on less cost-effective medical interventions and those spent on more cost-effective ones. U.S. federal and state government agencies have resisted the use of cost-effectiveness analysis, but it is widely used in the other developed countries. Although United States spends 16% of its gross domestic product on health care, according to Weinstein, it is not getting as much health improvement for its money as it could.

Comparative effectiveness research (CER) differs from cost-effectiveness research in that it involves weighing the risks and benefits of medical procedures without regard to the costs. The goal of CER is to compare the effectiveness of diagnostic or treatment practices for specific categories of patients rather than all patients having a certain condition. There are a number of procedures for which information on comparative outcomes are sorely needed. They include, for example, the effectiveness of different devices for treating hearing loss, the effectiveness of different drugs and surgical procedures when used alone or in combination in the treatment of cancers of various types, and the effectiveness of different procedures and medications for preventing falls by the elderly. CER takes into account the risks of mass testing that may lead to further invasive follow-up testing (and anxiety). This, in turn, may show that only a few cases were really serious and could have benefited from early treatment. Again, the assessment determines risks vs. benefits, apart from costs. So it does not involve putting a value on any particular life that would have been lost by discontinuing the mass testing. CER puts group interest above individual interest.

Evidence has shown that mass screening, as for breast cancer by mammogram and prostate cancer by the PSA blood test, can be grossly inefficient, even if effective. The U.S. Preventive Services Task Force pursued its mission properly in recommending against routine annual mammograms for women younger than 50. This decision was evidence-based and consistent with the public's interest, if not the desires of many physicians. The outcry against this decision was politically motivated and ill-advised. The goal of CER certainly is not to ration care or reduce the standards of medical care, as some believe. Rather it is intended to improve these standards by providing evidence regarding comparative effectiveness of procedures to physicians so that they can make informed decisions and provide more personalized medicine.

Allocation of Resources Between Research on Specific Diseases and on Research on Aging Processes

The gerontological research community is divided in its view as to the proper allocation of funds and resources between research on specific chronic diseases of later life and research on the fundamental processes of aging. The advocates of the former approach, primarily medical researchers, believe that the bulk of research funds should be allocated to research on specific major chronic diseases, so that

further substantial reductions in the death rates from these diseases, and perhaps even their eradication, can be accomplished. These advances could come about by improving the management of the diseases first, then postponing their age of onset, and finally eliminating them. In this way, additional decreases in death rates from cancer, heart disease and stroke, and reductions in diabetes, Alzheimer's disease, and some other major diseases would be achieved. This would result in increases in life expectation at each age and the eradication of many of these diseases could be anticipated. Moreover, the quality of life of persons afflicted with these diseases would be enhanced if they could face them with the hope of improvement and recovery.

On the other hand, the advocates of the molecular/biological approach do not choose to focus their efforts on the conquest of any particular chronic disease. They prefer to focus their efforts on understanding the fundamental processes of aging, i.e., determining why our cells age, become senescent, and die, and then finding ways to modulate these processes in order to extend the period of healthy life. This approach acknowledges that the molecular events that initiate, maintain, and intensify disease processes remain to be discovered, but it is believed that underlying them all is the fundamental process of aging itself. The advocates of this view further suggest that, even if a small delay in the biological process of aging – 3–7years for example – could be accomplished, the resulting health benefits would be enormous. Thereby, the age of onset and age progression of all fatal and non-fatal diseases of aging would be postponed simultaneously. Such research would contribute to treatment protocols for chronic diseases as a group by slowing the rate of biological aging. Current research suggesting the role of inflammation in the onset of several chronic diseases of later life is consistent with this view. Moreover, a number of studies with animal models have demonstrated that interventions to slow aging are plausible in humans.

The difference in the interests and goals of the two groups is more than theoretical; it bears a direct relation to the way research funds on health studies are allocated. The second group of researchers insists that relatively too little money has been made available for research on the fundamental processes of aging as compared with that allocated for research on the treatment of specific chronic diseases such as cancer, heart disease, and Alzheimer's disease. This argument is played out in the competition for funds appropriated by the Congress to the National Institutes of Health (NIH). In the past, NIH support for research in biogerontology, i.e., the study of the biological processes of aging, has been miniscule compared to support for research in geriatrics, i.e., study of the treatment of particular diseases of older age (Hayflick 2000), and this remains the case.

The fiscal year 2006–2007 proposed budget for the National Institutes of Health has, in fact, reduced allocations for research on the various major diseases over the previous year, but does not dedicate more money to biogerontological research and to the budget of the National Institute on Aging for such research. Funds have been reduced for research in the diagnosis and treatment for many of the leading causes of death and sickness, e.g., breast, lung, and prostate cancer, diabetes, arthritis, osteoporosis, diabetes, stroke, and mental illness. The reduction of funds means

Table 15.1 Proposed research funding at NIH, for selected causes of death, per death and per case: 2006–2007

Cause of death	Research funding[a] (billions)	Deaths[b]	Funds per death	Cases[c]	Funds per case
Total	28.6	2,426,264	11,787	NA	NA
Heart disease	2.5	631,636	3,957	24,107[d]	103.7
Cancer	7.8	559,888	13,931	15,820[d]	493.0
Cerebrovascular disease	0.338	137,119	2,465	5,642[d]	59.9
Lower respir. diseases	1.2	124,583	9,632	77,305[e]	15.5
Diabetes	1.0	72,449	13,802	14,012[f]	71.4
Influenza and pneumonia	3.8	56,326	67,464	NA	NA
Alzheimer's disease	0.645	72,432	8,904	4,500	143.3
Nephritis, nephritic syndrome, and nephrosis	0.448	45.344	988	3,017[f]	148.5
HIV/AIDS	2.9	12,044	240,783	405,926[g]	7,144.2

NA, Not available
[a]*AARP Bulletin*, July-Aug 2006
[b]U.S. NCHS Deaths: Final data for 2006. *National Vital Statistics Reports, 57*(14), Table 15.10, 2009
[c]Cases of the disease in the population. In thousands
[d]For adults 18 years and over; for 2006. NCHS (2007)
[e]*Statistical Abstract of the United States, 2006*, Tables 186 and 187. Data for 2003
[f]*Statistical Abstract of the United States, 2006*, Table 184. Data for 2003
[g]*Statistical Abstract of the United States, 2006*, Table 180. Data for 2003

fewer research grants for these specific types of diseases. Funds for some diseases, such as pneumonia, influenza, and heart disease, would go up somewhat, however.

The question arises, nevertheless, are research funds being allocated in proportion to the distribution of deaths by cause or the prevalence of the conditions by cause? The answer is that there is little relation between funds allocated and the number of deaths or the number of cases (Table 15.1). Heart disease, the most common cause of death, receives only one-quarter as much per death as cancer. Alzheimer's disease receives about the same as chronic lower respiratory diseases per fatality although the latter are twice as frequent. HIV/AIDS receives twice as much per fatality as several other major causes of death combined. Mental illness would fare superbly by this method of evaluation since it results in few deaths (unless suicide is linked to it), but when research funds per case of illness is considered, mental illness fares dismally. HIV/AIDS research appears to be the most heavily and disproportionately supported of all the diseases, given both the numbers of deaths and cases in the population.

The new developments in genomics, epigenetics, and proteinomics may soon make much of the argument moot. With the addition of gene therapy, genomic profiling, and stem cell research to the diagnostic and therapeutic armamentarium of NIH and their use by researchers supported by NIH, the approaches of geriatrics and biogerontology now have some common ground for tackling many health conditions. Two new programs at NIH will jointly attack the genetic variations in a group of patients with certain common diseases and monitor the personal environmental exposures of patients, to determine how the genetic-environmental interaction results in disease.[2] Inasmuch as 75% of the country's health care costs are associated with only a handful of common chronic diseases, including those under study, NIH believes that these programs can contribute both to advancing fundamental research on aging cells and research on specific diseases, while contributing to the control of skyrocketing health care costs.

Inequality in Health Risks and Treatment

In Chap. 7, I set forth data regarding the extent of the differences in the levels of morbidity and mortality between the sexes, races, and socioeconomic classes – representing differences in risks – and considered the factors explaining the differences in these risks. Here I explore the relation of the inequality of risks to the inequality of treatment and the implications of the inequality of risks and treatment for health policy.

---

[2] In pursuance of a merger of the two views as to the model way to attack disease, the President's budget for fiscal year 2006–2007 calls for an appropriation of $40 million a year to advance a multi-year Genes and Environment Initiative (GEI). This program will combine genetic analysis with the development of new environmental monitoring technology, to accelerate research on common diseases such as asthma, arthritis, and Alzheimer's disease. The program will analyze genetic variation in groups of patients with specific diseases, and monitor environmental exposures that interact with genetic variations and result in specific diseases. In addition, a partnership between private industry and government researchers, the Genetic Association Information Network (GAIN), would determine genetic contributions to seven common diseases. These initiatives are expected to identify the many genes in an individual that, taken together, contribute to the risk of illness.

Both the GEI and GAIN programs focus on genetic analysis of single nucleotide polymorphisms (SNPs) that normally occur within the three billion DNA base pairs constituting a person's genome. About one tenth of one percent of the SNPs alter the function of the gene while the combination of many slightly altered genes may significantly increase the risk of developing a specific disease. The goal is to identify the many genes in an individual that, in combination, contribute to an increased risk of illness. Having determined the genetic basis of these common diseases, researchers can develop targeted treatments that either prevent them from occurring or treat them effectively after they occur. (*NIH Record, 58*(4), 3, February 24, 2006).

Gender Inequality

Gender inequality in matters of health remains a matter of concern even though age-specific death rates are consistently lower for women than for men and life expectancy is much higher for women. In spite of advantages in survival and longevity, women suffer more than men from serious non-lethal health conditions and a substantial part of their advantage in longevity is spent in a disabled state. More pertinent to the present discussion is the evidence that the health of women, particularly their quality of life, could be greatly improved if they were treated more equitably in health matters. They are treated inequitably in both obvious and subtle ways. This is true in the More Developed Countries, but it is especially true in the Less Developed Countries (see Chap. 11 and below).

Many illustrations could be given of differences in the physiological reactions of men and women to various adverse stimuli. We have seen that the neuroendocrine responses of the sexes are different. Men and women differ in their responses to pain and analgesics. Biologically women have a lower pain threshold than men but have been socialized to experience pain at a higher threshold, recovering more quickly from pain and coping more effectively with it (U.S. NINDS 2001). Women are more prone to nicotine and alcohol addiction than men. More generally, women respond with greater stress to pharmacological agents. Women experience more strokes than heart attacks while men experience more heart attacks than strokes.

Consider gender differences in the symptoms of, responses to diagnostic testing for, and treatment of heart disease. Women's symptoms for heart disease can be different from those of men. Women may suffer from the "common" symptoms of heart disease (e.g., chest pain), but this is often not the case according to the research study, Women's Ischemia Syndrome Evaluation. In the case of a heart attack, women do not always experience the chest pains, shortness of breath, pain radiating down an arm, and sweating that men do, but may simply experience distressing fatigue, a backache, and moderate chest discomfort. Physicians have tended to dismiss these symptoms as fleeting, not serious, and in particular, not symptomatic of a heart condition.

Heart disease in women may not be diagnosed correctly by the conventional tests, which often are indeterminate. Coronary angiography for heart disease is a less accurate diagnostic procedure for women than for men. It does not reveal heart disease in many, if not most, of the women who have it because the women do not suffer from stenosis of the main coronary arteries but rather from blockage of the small subsidiary arterioles ("coronary microvascular syndrome"). The latter condition is difficult to detect. Men suffer from this syndrome as well but it makes up only a minority of the cases of heart disease in men. This situation suggests performing further tests such as an MRI scan of the heart or administering a paper test to the women calling for responses to a performance scale. Finally, women are less likely to receive the same aggressive treatment for heart disease as men and, as a result, are less likely to survive heart surgery.

"Old-time" male physicians tend to underrate the importance for women of such gynecological health practices as breast self-examinations and pap smears

and such gynecological conditions as PMS and menopause. Male psychiatrists are not in the best position to understand the special emotional and mental health problems troubling women. This type of inequitable treatment is likely to be reduced as women increasingly become practicing physicians and women patients seek them out.

It is recognized now that a wide range of diseases, including heart disease, lung cancer, and autoimmune disorders, affect women in very different ways from men. This is true also for some medications, such as statins used in the treatment of high cholesterol. The immediate implication of this sample of differences between women and men is that women require a different kind of attention, evaluation, and treatment than men for a host of symptoms and health conditions. They suggest the need for employing diagnostic and therapeutic procedures adapted to women and for including women separately in clinical trials (Pinn 2003). Until recently women had not been included in the various clinical trials and epidemiological studies, and few research studies had dealt with health conditions peculiar to women. Hence, it was not known whether the results of health research pertained to women as well as men. It was assumed that they did, however, and this assumption has sometimes proved to be wrong. In sum, because of differences in physiology, hormonal responses, drug reactions, and life course experiences (i.e., pregnancy, childbirth, and menopause), women respond differently to risk factors, diagnostic testing, and therapeutic procedures. Women must, therefore, be considered separately from men in medical research, diagnostic testing, and treatment protocols.

An important step in this direction has been taken. A Women's Health Initiative (WHI) was established by the National Institutes of Health in 1991 to address the most common causes of death, disability, and impaired quality of life in post-menopausal women. The conditions investigated were the cardiovascular diseases, cancer, and osteoporosis. The WHI was a 15-year program and one of the largest preventive investigations carried out by NIH. By 2006, masses of data are available for analysis.[3]

Class and Race Inequality

As explained in Chap. 7, in the United States blacks experience greater health risks and unequal access to health care as compared to whites. Health inequality is true also for the upper socioeconomic classes as compared to the lower socioeconomic classes, as identified by income, wealth, educational attainment, or occupation. Class differences in health are virtually universal, even in countries with universal health insurance. Health inequalities between the races and between the socioeconomic classes prevail over most of the life course, only tending to diminish and even disappear at the oldest ages. To the extent that inequitable treatment among

[3]The WHI had three components. They were: (1) A randomized clinical trial testing promising but previously unproven techniques of prevention; (2) an observational study to identify predictors of disease; and (3) a study of community programs intended to develop healthful behaviors.

racial/ethnic and socioeconomic groups contributes to inequalities in health status and it is avoidable, it must be viewed as unjust. Inequitable treatment explains only part of the differences, however. Inequalities in health in a society largely reflect the other inequalities in the society, such as inequalities in income, wealth, education, social status, environmental conditions, social participation, and the degree of autonomy. It is virtually impossible, therefore, for public programs to eliminate or even reduce health differences markedly without substantial progress in reducing socioeconomic differences. Reducing health differences in the United States, however, has become increasingly difficult because economic inequality between the race/ethnic groups and between socioeconomic groups has been widening and educational inequality, while narrowing (except for Hispanics), remains substantial.[4]

Marmot (2004), the distinguished British epidemiologist, maintains that investing in child care and education for the disadvantaged members of a population, improving deteriorated and hazardous urban environments, and providing social support for the elderly are the best ways to reduce socioeconomic differences in health in a society. While he regards differences in the quality of jobs, in wealth, and in social status as inevitable in societies today, he argues that the degree of these differences varies widely from one society to another and, to the extent that they are reduced, public health would improve. Marmot argues further that, above a certain level of poverty, the most effective health policy is one directed not at modifying the medical system with massive expenditures but at implementing programs that reduce inequalities in income and education, provide decent housing and safe working and living environments, and protect children and the elderly. Marmot contrasts the situation in England and the United States. In England, since the late 1990s, fiscal policy has been used to reduce income inequalities arising from labor market forces, whereas in the United States tax policies have tended to widen economic differences between the highest income group and the lowest income group. In England a higher overall level of health is achieved with less expenditures per capita, and race and socioeconomic group differences are smaller.

Is there environmental racism? An example of a possible difference in risk among the races is the phenomenon called by critics environmental racism. The term refers to the allegation that environmental hazards such as waste dump sites, nuclear power plants, chemical plants with toxic by-products, and other locally unwanted land uses (LULUs) are selectively located in racial-minority neighborhoods and that, therefore, society is supporting a racist practice. Specifically, questions have been

[4]The U.S. National Institutes of Health (NIH) has a major program of research designed to "eliminate" health inequalities among racial and ethnic minority communities in the United States. The National Center on Minority Health and Health Disparities, a unit in NIH, supports research on health inequalities among racial and ethnic groups with the goal of eliminating health inequalities among them. This goal is also included in the Department of Health and Human Services' plan entitled, "Healthy People, 2010."

raised with respect to the adverse effects on the health of racial minorities of the allegedly selective location of such facilities in rural areas where poor blacks live.

Environmental equity would require that, in determining the location of undesirable facilities, no particular population group should bear a disproportionate burden with respect to the number and types of undesirable land uses or adverse health effects. Several studies have concluded that racial discrimination plays an important role in the siting of LULUs. However, a national study of the location of hazardous waste treatment, storage, and disposal facilities (TSDFs), using census-tract-level data, conducted by Anderton et al. (1994), concluded that racial discrimination was not a significant factor in their location. According to this study, when tracts containing TSDFs are compared with tracts that do not contain TSDFs, no significant difference in the percentage of the population that is black was found. The one variable that was found to be significant with respect to TSDF location, after the effects of other independent variables were removed, was the percentage of the population in manufacturing occupations. Other demographic categories than race and ethnicity may be affected by decisions regarding use of the environment. For example, the location of undesirable facilities may place an unequal burden on poor, less educated, and less skilled persons, whether black, Hispanic, or white, and on the rural poor, in particular.

A community may be greatly dependent on a particular industry to provide jobs, either because unemployment is high or because it would be high if the local plant reduced or terminated its operations. As a result, though the environment of the people in the community may be adversely affected, they may support the expansion of industrial operations in the area because of the economic opportunity it provides. Furthermore, the community may have met national pollution standards, and the government may have been offering incentives to industry for economic development and job expansion in the particular area. The question remains whether the operation is or will be a health hazard, and whether some racial minority will be adversely affected to a disproportionate degree in comparison with other racial groups in a more extensive area. For example, in 1996 St. James parish in Louisiana had been facing these issues as a Shintech plastic complex was proposing to establish a petrochemical plant in the parish at a location largely populated by blacks and low-income households. The black community was divided on the issue for the reason given earlier. This area had already been cited in government reports as having hazardous waste and chemical facilities that disproportionately affected many black communities. In 1998 the Shintech company withdrew its plan to build the plant when the community voiced its opposition and when it appeared that the U.S. Environmental Protection Agency was about to disallow their proposal.

During the Clinton Administration the Environmental Protection Agency developed a rating system for assigning risk scores for industrial air pollution to every square kilometer in the United States. This rating system may be adapted to infer racial and ethnic differences in exposure to industrial air pollutants. As reported in the *Washington Post* on December 14, 2005, these scores indicate that in 19 states blacks were more than twice as likely as whites to live in neighborhoods where air

pollution posed the highest health hazard and that in 12 states Hispanics were twice as likely as whites to be living in such areas. The scores also indicate that residents in neighborhoods with the highest pollution scores also tend to be less educated, less affluent, and more likely to be unemployed than residents in other neighborhoods in the country.

Allocation of resources for research on equalizing class health risks. As described in Chap. 7 and again above, health risks vary greatly among socioeconomic and racial/ethnic groups in the United States. These inequalities exist even where these groups have equal access to health care, as among Medicare (ages 65 and over) and Medicaid program participants in the United States. Even in the countries of northern and western Europe with well developed systems of universal health care, substantial inequalities among socioeconomic groups persist. They are even larger in the United States among the socioeconomic groups for persons under age 65 that lack health insurance.

Several agencies in the U.S. Department of Health and Human Services, particularly PHS ("Healthy People 2010"), NIH, and AHRQ, have announced programs designed to eliminate these inequalities. While I consider this to be an impossible goal – it would require the elimination of all educational, income, wealth, occupational, and housing differences in the population – the goal of reducing inequity in access to and quality of health care among socioeconomic groups is an attainable goal. One proven way of reducing these inequities is to establish a system of universal health care, as in the other industrialized counties, to extend health services in underserved areas, particularly rural areas, and to restructure the delivery of health care to make it more patient-centered. Even with universal health care systems, the member countries of the European Union are still living with health inequalities among socioeconomic groups and the European Union is studying ways to reduce these inequalities further.

In 2007 the American Cancer Society (ACS) announced a change in the primary focus of its efforts and in the expenditures of its funds, from research on new treatments and cures for cancer to efforts to mobilize the population to live more healthfully and to utilize more fully the existing knowledge regarding the relation of lifestyle and behavior and good health. A corollary of this new focus of the Society's efforts, not specifically articulated, is to reduce socioeconomic and racial/ethnic inequalities in health, including inequities in access to health care. This is a private "policy" decision, but it is useful to compare it with the health policy of the federal government. The goals just described are included among the goals of HHS's "Healthy People 2010." At the same time, NIH has been pursuing a policy of research in the treatment of a whole array of diseases as well as limited research in the fundamental processes of aging. It would appear that ACS and the Federal Government are now both proceeding on a broader front than in the past in their efforts to advance the health status of the U.S. population.

Success in the new thrust of the ACS can have tremendous dividends. These were discussed in Chaps. 6 and 7. Several years could be added to a population's life expectancy at birth if everyone consumed a healthful and nutritious diet, followed a regular exercise regimen, added to or removed body weight to bring it into the

normal range, stopped smoking, and took regular complete physical examinations with appropriate follow-up. This program defines an ideal goal that may be difficult to achieve but in any case it defines a highly desirable trajectory.

One difficulty with the new policy of the National Cancer Society is that it has implications for the relation between the various socioeconomic and racial/ethnic groups that are troublesome. For the most part, high-income and educated people are familiar with the guidelines for a healthy lifestyle and are looking for new treatments and cures for diseases, to be achieved by research. Their interests then are bypassed by the new ACS policy in the interests of bringing the rest of the population "up to par." To the extent that ACS maintains a balance between the goals of equalizing lifestyle and behaviors among groups and engaging in research on treatments, few supporters of ACS would fault their modified program.

Factors and Trends in the Costs of Health Care

Past Trends in Costs of Health Care

The costs of health care in the United States have been increasing sharply for many years. Good health care is not within the budget of masses of Americans. A family of four persons with income at the median level in the United States must spend about one-fifth of its income for health insurance. A family of four with income at the poverty level would have to pay about 50% of its income on health insurance. Tens of millions of Americans are not insured and most of those who are insured have insurance that covers only a limited share of the costs of care. As a result, many forego the kind and scope of treatment they need, or fail to secure them until an illness has progressed to the point of extreme financial cost to the system. Even if a family tried to avoid illness by following the guidelines on diet, weight, exercise, and other lifestyle and behavioral practices, it can still be difficult financially to comply. Some aspects of a healthy lifestyle are relatively cheap, such as exercise, but others are costly and may not be affordable for some families.

Life expectancy at birth in numerous other countries around the world, including Canada, Costa Rica, Cuba, France, Great Britain, Greece, Israel, Japan, Spain, Sweden, and New Zealand, is about the same as or higher than in the United States. Yet, all of these countries are less wealthy than the United States as measured by gross domestic product per person and they spend far less per capita on health care than the United States. Currently the United States spends nearly twice as much per capita as the other developed countries.

In 2008 national health expenditures in the United States amounted to $2.3 trillion and the per capita expenditure for health services and supplies amounted to $7,681 (Table 15.2). Personal health expenditures grew at an average annual rate of 8.8% since 1980 although the annual rate of increase fluctuated greatly above and below this average level from year to year while showing a general tendency to decline. The rate of increase in health-care costs has regularly exceeded the rate of increase in gross domestic product (except for a few years between 1995 and 2005,

Table 15.2 National health-care expenditures, total and as percent of gross domestic product, for the United States: 1980–2005, and projections, 2010 and 2015

Year	Total expenditures[a]	Annual average percent change[b]	Percent of gross domestic product	Per capita expenditures[c]
1980	253.4	13.0	9.1	1,100
1990	714.1	10.9	12.3	2,814
2000	1,352.9	6.9	13.6	4,789
2005	1,982.5	6.9	15.7	6,701
2006	2,112.5	6.6	15.8	7,071
2007	2,239.7	6.0	15.9	7,423
2008	2,338.7	4.4	16.2	7,681
Projections				
2010	2,776.4	6.7	17.2	8,985
2015	3,874.6	6.7	19.2	12,062

Source: U.S. Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group, www.cms.hhs.gov/NationalHealthExpendData

[a]In billions of dollars

[b]1980 and 1990, annual average percent change for preceding 10 years; 2000–2008, percent change for preceding year; 2010 and 2015, annual average percent change for preceding five years

[c]Based on estimated July 1 resident population

when the annual percent increases in health care costs were relatively low). As a result, health-care costs as a share of the gross domestic product have been rising more or less steadily. It was 9.1% in 1980 and 16.2% in 2008.

As projected by the Centers for Medicare and Medicaid Services, the costs of health care will reach $3.9 trillion in 2015. This projection assumes a 7.2% annual average rate of increase in health-care costs over the decade, 2005–2015, as compared with an actual increase of 6.9% in 2005. The rate of growth of health-care costs is expected to exceed the rate of growth of GDP in the 2005–2015 period, and hence health-care costs will continue to rise as a share of GDP, reaching 19% in 2015 (Table 15.2). For the long term, Medicare costs – a leading component of the health-care bill – are projected to rise between 2005 and 2050 from 2.5% to 9.2% as a share of GDP.

In 2008 nearly two-thirds of the personal health-care dollar was spent for hospital care and physician and clinical services (Table 15.3). Hospital care alone accounted for nearly two-fifths of the total. Other professional services, prescription drugs, and other medical products accounted for over one-quarter of the personal health-care bill. Nursing-home care and home-health care made up only one-tenth of the total cost. This distribution of health spending has changed little since 2000.

Factors Driving Health-Care Costs

In past decades cost inflation, including both general inflation and excess inflation in the health sector of the economy, and increased per capita demand for health care, driven by developments in medical technology, have been the major factors in the

Table 15.3 Amounts and per capita costs of personal health-care expenditures according to type of expenditure, for the United States: 2000 and 2008

	(Amounts in billions of dollars)					
	2000			2008		
Type of expenditure	Amount	%[a]	Per capita[b]	Amount	%[a]	Per capita[b]
Total	1,139.2	100[b]	4,037	1,952.3	100[b]	6,412
Hospital care	416.9	37	1,477	718.4	37	2,359
Physician and clinical services	288.6	25	1,023	496.2	25	1,630
Other professional services	101.1	9	358	166.9	9	548
Other personal health care	37.1	6	131	68.1	3	224
Prescription drugs and other medical products	169.8	15	602	299.6	15	984
Nursing home care	95.3	6	338	138.4	7	455
Home health care	30.5	3	108	64.7	3	212

Source: U.S. Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group. www.cms.hhs.gov/NsationalHealthExpendDataNationalealthExpendstatistics. Calculations of per capita expenditures by author

[a]Percent distribution

[b]Based on estimated population as of July 1

Table 15.4 Factors contributing to the increase in personal health-care expenditures in the United States: selected years, 1965–2008

	Percent increase				Percent of total increase			
Year	Total expenditures	Population	Prices[a]	Medical-care utilization[b]	Total	Population	Prices[a]	Medical-care utilization[b]
2008	4.6	0.9	3.1	0.6	100	20	67	13
2007	5.9	1.0	3.4	1.4	100	17	58	25
2006	6.5	1.0	3.5	2.0	100	16	53	31
2005	6.8	0.9	3.6	2.3	100	13	53	33
2000	6.7	1.0	3.3	2.4	100	15	50	36
1995	6.0	1.1	3.6	1.4	100	18	59	23
1990	11.7	1.2	6.8	3.7	100	10	58	32
1985	10.2	1.0	7.3	1.9	100	10	72	18
1980	15.9	1.1	11.6	3.2	100	7	73	20
1975	14.5	0.9	11.0	2.6	100	6	76	18
1965	8.4	1.2	2.8	4.3	100	15	33	52

Source: Unpublished data provided by Anne Martin, Office of the Actuary, U.S. Medicare and Medicaid Services. Related publication: Hartman et al. 2010

[a]Medical prices, that is, the sum of economy-wide inflation and excess medical inflation

[b]Calculated as a residual. Includes improvement in the quality and growth in the quantity of health-care services delivered per capita (i.e., use of new technology and any other changes in health-care services delivered per person)

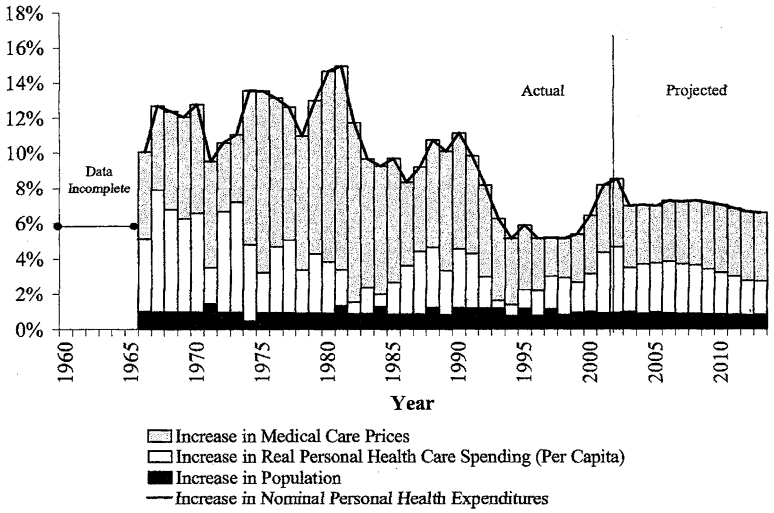


Fig. 15.1 Factors Contributing to Growth in Nominal Personal Health-Care Expenditures, for the United States: Actual Growth Rates, 1966–2001, and Projected Growth Rates, 2002–2013 (Note: To make the component factors additive, the percentages for each component were calculated as continuous growth rates (i.e., assuming exponential compounding) rather than as discrete annual changes. The former are only slightly smaller than the latter and the difference is not detectable on the chart; Source: U.S. Congressional Research Service, using data from the U.S. Census Bureau and the Office of the Actuary, U.S. Centers for Medicare and Medicaid Services)

rise in health-care costs in the United States (Table 15.4). The increase in health-care prices was the number one factor in rising health-care costs while the increase in per capita demand was second. Several authors in a special edition of *Health Affairs* devoted to *Health and Costs of the Future Elderly* (Vol. 24, Suppl. 2, 2005) point to the increasing use of medical technologies as an important driving force behind rising costs of health care in the past (e.g., Shekelle et al., 2005; Lubitz, 2005). Population growth has been a minor factor (Siegel 2002). Analysis of data compiled by the Centers for Medicare and Medicaid Services by the U.S. Congressional Research Service (2008) and the California HealthCare Foundation (2008) confirm these conclusions. In general, they found that, for the years from 1960 to 2002, over half of the increase in personal health-care costs was accounted for by price changes (including both general inflation and excess medical inflation) and that approximately one-third has been associated with technological advances in medicine (i.e., rise in per capita demand) (See Table 5.4 and Fig. 15.1.).

The effect of population growth and aging has been small – under 15% in most years since 1961 – but it has been rising somewhat over the last few years, from 14% in 2005 to 20% in 2008. Price increases also accounted for a larger share of the total increase in personal health-care expenditures in the years since 2000 – from 50% in 2000 to 67% in 2008 – while utilization (i.e., per capita demand) accounted for a smaller share – from 36% to 13%. Hence, in spite of the recent rise in its role, population still remains a minor factor in the rise in health-care costs.

Prospects for Costs of Health Care

There are two widely divergent views as to the prospects for health care costs in the United States and the Medicare system. The first view, which I label pessimistic, is that health-care costs will continue to rise at a fast pace and that this increase in costs will have severe and adverse consequences for the future delivery of health care and for the U.S. economy. The second view, which I label optimistic, is a minority view that focuses on the reductions in disability ratios in the last few decades, the expectation that this trend will continue, and the likelihood that this change will result in the reduction of health-care costs.

Pessimistic view. The first view is set forth in many of the papers in the issue of *Health Affairs* cited (Vol. 24, Suppl. 2, 2005, *Health and Costs of the Future Elderly*). The entire issue was devoted to the prospects for spending on Medicare in the next quarter century, and most of the authors agreed that foreseeable improvements in health will cost the American public more money rather than save money. They believe that future technological changes in medicine will drive up expenditures for Medicare. This factor and others, including increases in per capita demand, inflationary pressures, and demographic changes, will drive Medicare costs up even if progress is made in reducing the incidence of the chronic diseases of later life and disability. Population increases and changes in age structure, as well as general inflation and excess inflation in the health care industry, as in the past, are expected to combine with new developments in and increasing use of medical technologies in causing a rise in health costs. According to this view, neither new medical developments nor improved functional status of the population at the older ages are likely to relieve the budget pressure on Medicare. These analysts generally agree that improving people's health will not reduce spending for health care (see, for example, Vladeck 2005 and Lubitz 2005). Furthermore, if health is not improved but life is prolonged (i.e., assuming morbidity expansion), Medicare costs would be even greater since sick persons who survive for long periods are more expensive to maintain than sick persons who die early.

The increase in the projected costs of health care will result in part from the use of many new expensive medical technologies, as the following illustrative list suggests: Improved defibrillators, pacemakers, and other cardiac-assistive devices; robotic nanotechnology, including proton beam therapy for cancer treatment; new diagnostic technologies, including new and improved imaging devices, internal wireless information devices, and protein signature identification; improved heart by-pass and heart-and-other-organ transplant surgery; implementation of gene therapy on a broad scale; and widespread application of regenerative medicine and tissue engineering.

According to this view the costs of health care are expected to rise sharply in the next several decades, especially after 2011. The advent of the large baby-boom cohorts that were born in 1946–1964 to age 65, the age of eligibility for Medicare, in 2011–2029 will partly explain the expected continuation of the rapid rise in health-care costs. The costs will tend to rise simply because there will be many more people

in the ages of greatest expenditures on health care, and the new elderly are likely to live longer than their predecessors. As we saw, most analysts predict that life expectancy will increase in the next several decades though they differ greatly on the extent of the increase.

Goldman et al. (2005) conclude from their simulation studies that, while better health (e.g., reduction of chronic diseases), will accompany the use of improved and new medical technologies, these changes will be associated with increased per capita demand for health care and are likely to cause a great increase in medical spending. Use of medical technologies will be a major force in rising health costs in the future, therefore. An older population with improved functional status could have lower annual health-care costs but, assuming greater longevity, it will also have more years in which to accumulate health-care costs and hence a greater likelihood of high long-term costs.[5] Moreover, there would still be the high costs of the last year before death. Although disability ratios among the elderly declined during the last few decades of the last century, the prospective trend is uncertain and, in fact, they have risen in the first several years of the 2000s.

During the 2011–2029 period, demographic changes are expected to have a much greater effect on health-care costs than in prior years but developments in medical technology are expected to play an equally important role. Cutler (2005) has estimated that about half of the future increase in health-care costs will be the result of demographic trends and half will result from the excess of medical-cost growth over GDP growth.

The rising trend in obesity recorded in the last few decades is expected to contribute to the higher annual and lifetime spending for health care. Whether obesity diminishes the length of life is unclear, but more money will have to be spent on maintaining persons with the condition over a longer time because persons with obesity spend fewer years than non-obese persons in a disability-free state. There has also been a rise in the prevalence of some chronic diseases, including diabetes, chronic obstructive pulmonary diseases, and Alzheimer's disease, among pre-elderly persons that may be expected to result in a rise in other associated chronic conditions and in the costs of health care in later life. Accordingly, the implementation of measures to prevent or delay the onset of the leading chronic diseases is likely to have only modest effects on the solvency of the Medicare program (Gruber 2009). Given the poor or limited adherence of the U.S. public to recommendations regarding lifestyle and behavioral changes relating to diet, weight, exercise, smoking, stress, and blood pressure, it is not likely that all or even most people will act in healthy ways in the future.

[5]When mortality rates fall, there are two opposing effects on health-care costs. First, more people are alive at the older ages and many of these will suffer from the chronic diseases from which they were saved or other chronic diseases for many years. This factor tends to make health-care costs rise, Second, at each older age a smaller percentage of the population is close to death and hence fewer are spending gigantic sums of money for end-of-life illness, This factor tends to make health-care costs fall. These two factors tend to offset one another to an unknown extent, but it is likely that the former effect is dominant.

In sum, most reasonable scenarios regarding future health-care costs show that improved health will not necessarily mean a reduction in the costs of health care. Nor is there any evidence that health prevention reduces health-care costs in the current technological environment. The health situation may even stagnate in the present or other form. Consequently, health-care costs are not likely to fall in future years.

The proponents of the above position concede, however, that it may be possible to achieve cost savings in several ways. For example, medical technologies could be developed that do more cheaply what is done now by more expensive means. With greater coverage of the population by health programs, the share of untreated persons in the population could fall and the general health situation could improve sharply. Moreover, most of the new insured persons would draw little on the system because they are in the "healthier" ages. Greater prosecution of fraud in Medicare and greater control on patients' decisions as to what treatments to secure, particularly treatments not supported by comparative-effectiveness guidelines, are other important cost-control programs. Medicare spends more than $300 million extra per year for additional care needed as a result of medical error or adverse events, such as postoperative blood infection and errors in medication dosage (C. Zhan et al. 2006).

Optimistic view. A second, more optimistic, view regarding future health-care costs anticipates large increases in human longevity, continuation of reductions in chronic diseases and disability, the compression of morbidity, and as a result, substantial reductions in the costs of maintaining the Social Security system, Medicare, and Medicaid. This view considers these three programs as one broad interacting financial system. Manton et al. (2006a,b) have set forth this position in its most comprehensive form. They argue that, as a result of the considerable reductions anticipated in future levels of chronic diseases and disability, the share of non-disabled years among total expected years will greatly increase in the future. Healthy life expectancy at age 65 was estimated at 8.8 years in 1935 and 13.9 years in 1999. It is projected, in their scenario one, to rise to 16.4 years in 2022 and 20.8 years in 2080. This projection assumes that the 1982–1999 rate of disability decline continues to 2022 and that there is an 0.8% annual decline between 2022 and 2080. Non-disabled life expectancy as a share of total life expectancy is projected to rise from 79% in 1999 to 85% in 2022 and 88% in 2080.

Fries (2003) is a philosophical ally of the Manton et al. position. He concludes that compression of morbidity occurred in the decades of the 1980s and 1990s by comparing the larger reductions in disability ratios with the smaller ones in mortality rates. Furthermore, he has noted that various longitudinal studies of health enhancement programs of the elderly population show reductions in health risks, postponement of disability, improved health status, and decreased medical-care utilization of at least 10% per year. Fries is clearly suggesting that compression of morbidity of a population implies improved health status for its members and that these changes lead to savings in health-care costs.

Age Variations in the Costs of Illness

Age variations in expenditures on health care roughly mimic the age pattern of age-specific death rates. Increasing age through adulthood is accompanied by an increasing need for health services and products, and direct expenditures for health services and products rise with increasing age. Persons change their savings and consumption patterns in older age, tending to save less and to shift a greater share of their expenditures into health services and products. For example, in 2003, health-care expenditures as a share of total expenditures, per consumer unit, rose steadily from 2.4% for the age group 18–24 years to 15.4% for the age group 75 years and over (U.S. Bureau of Labor Statistics 2003).

Age variations vs. time to death. Numerous studies have demonstrated that health care costs of patients tend to be concentrated in their last year of life, regardless of their age (Yang et al. 2003; Miller 2001; Zweifel et al. 1999; Lubitz and Riley 1993). Controlling for time-to-death reduces the effect of age greatly and results in a (theoretical) decrease in hospital costs even for the oldest old (Seshamani and Gray 2005). Seshamani and Gray found that costs in the last quarter of the last year of life are seven times the quarterly costs three years from death. Participants in the U.S. Medicare program who die in a given year account for 28% of all Medicare expenditures but constitute only 6% of Medicare enrollments. Time to death has a large impact on hospital costs even five years from death. Hence, it can play an important role in determining future demand for health services, allocation of resources for health, and costs of health programs.

Costs of illness vs. costs of death. With the event of illness, there are direct and indirect costs. The total costs include both the direct costs of treating the illness and the indirect costs represented by the value of the production lost for the duration of the illness. Generally, the indirect costs will greatly exceed the direct costs. With the event of death, the direct costs of illness are terminated. The indirect costs of death, incurred first during the period of illness, are "continued" after death in the form of the value of the production lost during the period of years corresponding to the decedent's "life expectancy" at the age of death (i.e., years of worklife expectancy plus informal productive years minus dependent years).

Whether the death of a severely ill person is more economical for a society than keeping him or her alive at all costs is sometimes raised in debates relating to the costs of health care. From a purely economic point of view, the value of an individual life depends on whether the person is economically active or in unpaid work on which a market value can be placed (e.g., homemaker and caretaker services).[6] When the person is still economically active, the market value of the person can be balanced against the costs of any illness. This value depreciates when he or she experiences a prolonged illness. Under typical conditions following retirement, the death of a person afflicted with a degenerative, disabling illness is financially less costly

[6]A refined calculus assigns an economic value to the services of parents and grandparents in managing a home for others, rearing children and grandchildren, and caring for dependent elders.

for a society than prolonging the life of the person. This is not just an economic question, however, but an ethical or moral question as well. It becomes a practical issue for serious debate when the individual is at an extremely advanced age, is quite ill, and requires expensive and extensive medical treatments consuming valuable resources that could be made available to others. (See further discussion in Chap. 17.)

In sum, it is doubtful that any money is saved by extending human life to higher and higher ages since the ongoing ills of the survivors have to be treated over a longer period, use of complex medical technologies that are quite expensive may have to be administered, and all survivors generate the costs of the final year of life when costs are at a maximum. Considering societal money costs alone, then death is cheaper than life at advanced ages with serious chronic illness, and it may be cheaper than life without such conditions. Again, I want to emphasize that human life at the later years has value in itself, especially to the families involved, apart from the issue of societal costs and cost savings. Finally, however, if morbidity can be compressed up to the time just before death at a very advanced age, the costs of health care would be minimized as compared with the costs of a prolonged period of major illness or long-term care. This should be the goal of health research in later life.

Life years lost by premature death and their costs . The life years lost by premature deaths and the costs of the premature deaths are matters of interest to public health administrators and economists as well as to society and the families affected. Table 15.5 presents illustrative estimates of life years lost by premature deaths and the costs of the premature deaths, separately for males and females, age groups, and selected causes of death. They are based on deaths in the United States and U.S. life tables for 2002. Life years lost by death are computed on the basis of the life expectancy at the age of death of each individual decedent, according to sex and cause of death. A short-cut computation for life years lost by death for males 5–9 years of age in year 2002 is as follows:

$$\begin{aligned} &\text{Deaths, males, } 5-9 * \text{life expectancy, males, at age 7.5 years} \\ &\quad = 1{,}702 * 67.8 = 115{,}396 \text{ years} \end{aligned} \tag{15.2}$$

This estimate of life years lost is based on a standard life table that does not allow for prospective changes in death rates; hence, the estimate is probably biased downward as compared with an estimate based on the corresponding generation life table.

Life years per death represent the average years lost for each death. The measure can be computed by dividing the total life years lost by the number of deaths, or simply noting the life expectancy used in the calculation of total life years lost. The age variation in the figures of life years lost per death reflect, in reverse, the variation in the ages at which the deaths are concentrated. Since deaths are fewer and life expectation is higher at the younger ages, life years lost per death are greater at the younger ages. Observe that the ratio for cancer for women, 17.9, is higher than the ratio for heart disease, 11.8; this difference reflects the fact that cancer tends to occur at younger ages, on the average, than heart disease. Similarly, the ratio for adverse effects (e.g., violence), 30.7, is much higher than both of these causes.

Table 15.5 Estimated years of life lost by premature death and the present money value of the years lost, by age and sex and by cause of death, for the United States: 2002

Characteristic	Number of deaths (1000)	Life years lost[a]		Cost of death[b]	
		Total (1000)	Per death	Total (mil. dol.)	Per death (dol.)
Male	1,199	21,137	17.6	340.827	284,264
Under 5 years	19	1,379	74.4	19,341	1,044,188
5 to 14 years	4	282	67.3	5,312	1,265,460
15 to 24 years	24	1,404	57.5	36,464	1,493,433
25 to 44 years	87	3,626	42.0	111,931	1,296,567
45 to 64 years	250	6,584	25.4	140,248	541,322
65 years and over	800	7,862	9.7	27,530	34,138
Heart disease	341	4,762	14.0	61,487	180,367
Cancer	289	5,481	19.0	81,952	283,805
Cerebrovascular disease	63	763	12.2	8,041	128,410
Accidents and other adverse effects	69	2,444	35.3	61,311	886,215
Other	438	7,687	17.6	128,036	292,642
Female	1,244	18,080	15.3	134,274	107,933
Under 5 years	14	1,147	79.8	11,834	823,603
5 to 14 years	3	215	72.7	2,934	993,950
15 to 24 years	9	543	62.9	9,948	1,152,744
25 to 44 years	46	2,116	45.8	41,257	893,658
45 to 64 years	110	4,798	43.5	55,154	500,156
65 years and over	1062	10,261	9.7	13,147	12,383
Heart disease	361	4,203	11.8	18,373	50,715
Cancer	267	4,807	17.9	36,920	136,557
Cerebrovascular disease	100	1,146	11.5	4,788	47,279
Accidents and other adverse effects	35	1,081	30.7	16,548	441,578
Other	469	7,638	16.1	58,276	120,899

Source: Unpublished data reprinted with permission of the Institute for Health and Aging, University of California at San Francisco. Primary source of principal data, including deaths and life tables for calculating life years lost: U.S. NCHS. See also U.S. Bureau of the Census, *Statistical Abstract of the United States: 2006.* Table 116

Number of years person would have lived if he/she did not die prematurely. Cost of death: Value of lifetime earnings lost by persons who died prematurely

[a]Life years lost based on life expectancy at each age

[b]Cost estimates based on the person's age, sex, life expectancy at the age shown in year of death, age-specific labor force participation ratios, annual earnings, value of homemaking services, and a 3% discount rate by which to convert the potential aggregate earnings lost in future years to present worth

Money value of lives lost prematurely. The common method of determining the indirect costs of premature death uses the so-called human capital approach. It involves making a projection of what an individual would have earned in his or her remaining lifetime had he or she lived. Although it is impossible to predict

the future employment and earnings history of an individual, values of future lifetime income or earnings can be estimated for aggregates of persons on the basis of their current demographic, socioeconomic, and financial characteristics. It is apparent that the future lifetime earning capacity of a man 30 years of age is likely to be much greater than that of a man 55 years of age. Similarly, we may expect a large difference in prospective earnings for a college graduate and a high-school dropout, or an accountant and a office clerk. It is not possible, however, to secure accurate estimates of lifetime earnings for many of the characteristics that determine the earnings prospects of an individual, such as occupation and continuity of lifetime work experience, particularly in cross-classification with other determinative characteristics for which data are available, such as race and educational level.

Short-cut method. A rough method of calculating future lifetime earnings at a particular age for a designated group (e.g., working males who dropped out of high school) is to sum the mean annual earnings, for this group, for each successive age from the current age to the age corresponding to the expected working life for this group:

$$\mathrm{L_a} = \sum_{\mathrm{n=a}}^{\mathrm{n=a+ewx}} \mathrm{Y_n} \tag{15.3}$$

where $\mathrm{L_a}$ is total lifetime earnings at age a, n is the shifting age from the starting age a to each successive age until retirement age $a + ewx$, ewx is expectation of working life at age a, and Y_n, is the mean annual earnings at each age. Hence, using a figure for expected working lifetime of 26.0 years, the future lifetime earnings of a 30.0-year-old male high-school dropout could be approximated as the sum of mean annual earnings for such persons at each age from age 30 through age 56 (the sum of exact ages 30.0 and 26.0). Such estimates are too rough to be of value in cases, for example, where the effect of losses to the economy resulting from premature death is being analyzed or where liability for death is being adjudicated, so that a more refined method needs to be considered.

Refined method. A more refined and realistic method of calculating the money value of a life lost by premature death involves extension of the above formula to allow for additional demographic, socioeconomic, and financial variables in addition to age, sex, and mean annual earnings, such as survival rates, age-specific labor force participation or employment, and a discount rate by which to convert the potential earnings lost over future years to present worth. An illustrative formula, such as that used by the U.S. Bureau of the Census in its publications and the Institute for Health and Aging of the University of California, San Francisco, in constructing Table 15.5, is as follows:

$$\mathrm{L_a} = \sum_{\mathrm{n=a}}^{\mathrm{n=a+wx}} \mathrm{y_n p_n e_n (1+x)^{n-a+1/2} \div (1+r)^{n-a+1}} \tag{15.4}$$

where lifetime earnings at age a – (L_a) – is calculated as the cumulative product of (1) mean annual earnings for persons at each age n – (y_n) – for specialized aggregates of persons with certain demographic and socioeconomic characteristics starting from a designated age a and continuing to the end of life, (2) the survival rate from age a up to each successive age until age $a + wx$ – (p_n), and (3) the employment ratio at each age (e_n). With this formula, the calculations are carried up the age scale as far as earnings data are available for employed persons ($a + wx$). This formula assumes that the individual saved at age a is still subject to mortality at higher ages. At each age n, this product is adjusted upward by an assumed (fixed) factor for the annual increase in earnings due to rising worker productivity – (1+x) – and adjusted downward by an assumed (fixed) discount rate representing the excess of the inflation rate over the interest rate – (1 + r).[7]

With all the efforts at refinement, such a measure has severe limitations, especially when applied to an individual, such as in the adjudication of a liability claim. Aggregate data and average risks are often applied to individuals in these cases. The essential component, cross-sectional earnings data for each age, is compiled as data for a synthetic cohort but interpreted as data for a real cohort; that is, mean annual earnings are "frozen" at the level of the base year. Other rates and ratios are also held constant at the base-year level (i.e., employment ratios and survival ratios), and fixed, model allowances are made for other factors (i.e., productivity, inflation, and interest rates). The allowances that must be made for productivity, inflation ratios, and interest rates require the analyst to anticipate the state of the future economy. In short, the accuracy of the estimates of lifetime earnings depends on how closely these current data reflect conditions that would have been experienced in future decades by the deceased worker.[8]

---

[7] The factor 1 + x allows for real growth in productivity. The exponent assumes productivity growth for the number of years to age n, starting at age a. The factor for the discount rate (1 + r) converts future earnings to their present values. The exponent assumes that the discount rate applies to each year until age n, given starting age a.

This formula for the calculation of the lifetime income lost assumes that the decedent would have had an expense-free life. In a further refinement of the formula, mean annual earnings at each age should be reduced for expected taxes and living expenses. Optimally, allowance should be made for all costs incurred after retirement up to burial:

$$L_a = {}^{n=a+wx}\sum_{n=a}(y_n - t_n - l_n)p_n e_n(1 + x)^{n-a+1/2} \div (1 + r)^{n-a+1} \qquad (15.5)$$

where t_n represents taxes, l_n living expenses, and $a + wx$ the highest age for which data are available on employed workers. One could also take into account the health and disability status of the person. A group in the aggregate will spend a number of years in a disabled state, some before and/or others after retirement. To the extent that expenditures for private-home care and nursing-home care are taken care of by the family, some adjustment should be made in the estimates for money that would be spent on such care.

[8] It is normally assumed that persons do not begin working for pay until age 18 and cease working when they reach age 65 or so. Alternatively, the mean age of retirement according to current tables of working life or other sources can be used to establish this age. Generation life tables or generation table of working life would be much more realistic than calendar-year life tables, but they are subject to considerable uncertainty because of the need for projecting both the probabilities of dying and the probabilities of working.

Issues in Health Insurance and Effects on Mortality and Morbidity

Health Insurance Coverage

The vast majority of elderly persons in the United States participate in the national health insurance program called Medicare, most children have health coverage under the State Childrens Health Insurance Program (SCHIP), Medicaid, or private plans, and military personnel and veterans enjoy federal health insurance. At the same time, tens of millions of Americans do not have health insurance, and the number and share of the uninsured in the total population are rising rapidly. In 2006 an estimated 47.0 million persons in the United States, or 15.8% of the total population, lacked health insurance (Table 15.6). The total of 47 million uninsured includes several million illegal alien residents, and other millions who choose not to participate although health insurance is offered through their place of employment or otherwise. Many young people do not feel that health insurance is necessary and/or do not choose to spend their money for this purpose even though a substantial share of them can secure highly subsidized health insurance from their employer. Many of the uninsured are not offered health insurance in their place of employment. The majority of these families cannot afford to buy health insurance in individual plans on the basis of their income or net worth (Bernard et al. 2009). According to the Current Population Survey's Annual Supplement, there was an increase in the number uninsured for the 5-year period, 2001–2006, amounting to 7 1/4 million, or an annual average increase of 3.3%.

To the figures on the uninsured, we need to add the persons who are underinsured, that is, persons who have health insurance but pay a substantial share of their income for health insurance and still may not have coverage for many medically necessary conditions. According to the Commonwealth Fund Biennial Health Insurance Survey (Doty et al./Commonwealth Fund 2008), 14%, or 25.2 million, of the under-65 population was underinsured in 2007.[9] The number grew rapidly in the preceding years; it was only 15.6 million in 2003. In addition, most persons who nominally have adequate health insurance lack meaningful dental insurance, eye-care insurance, and mental health insurance. An estimated 108 million persons have no dental insurance. This number includes three-quarters of the population 65 years and older. While attention has often been drawn to the number lacking health insurance, the situation regarding the underinsured and the widespread lack of dental, optical, and mental health insurance has been generally ignored.

[9]The estimate of the number of underinsured can vary widely depending on the coverage standard set. The Commonwealth Fund defined underinsured as having health insurance all year but falling in one of the following groups: (1) medical expenses totalling more than 10% of family income; (2) for families with incomes below 200% of the poverty level, medical expenses totaling more than 5% of family income; or (3) health insurance plan deductibles that are more than 5% of family income.

Table 15.6 Health insurance coverage, by selected characteristics, for the United States: 2006 (Persons in 2007 for coverage in 2006; numbers in millions)

Characteristic	Total persons	Covered by health insurance					Not covered by health insurance	
		Total[a]	Private	Government[ab]			Number	Percent of population
				Total	Medicare	Medicaid		
Total	296.8	249.8	201.7	80.3	40.3	38.3	47.0	15.8
Age (years)								
Under 18	74.1	65.4	47.9	22.1	0.4	20.1	8.7	11.7
18–24	28.4	20.1	17.0	4.0	0.2	3.3	8.3	29.3
25–34	39.9	29.2	25.8	4.5	3.4	0.5	10.7	26.9
35–44	42.8	34.7	31.5	4.4	3.0	0.8	8.0	18.8
45–64	75.6	64.9	57.5	11.3	4.7	5.2	10.7	14.2
65 and over	36.0	35.5	21.9	34.0	33.8	3.4	0.5	1.5
Race								
White alone	237.9	202.4	167.6	62.6	34.4	26.5	35.5	14.9
Black alone	37.4	29.7	20.0	12.5	4.1	8.5	7.7	20.5
Asian alone	13.1	11.1	9.3	2.6	1.2	1.5	2.0	15.5
Hispanic origin[c]	44.9	29.6	19.4	12.2	2.8	9.6	15.3	34.1
Household income								
Less than $25,000	55.9	41.9	18.3	31.7	15.7	18.5	13.9	24.9
$75,000 or more	109.8	100.5	95.7	12.7	6.3	3.9	9.3	8.5
Below poverty line	36.5	24.9	8.5	18.9	4.7	15.4	11.5	31.5

Source: U.S. Bureau of the Census, *Current Population Reports*, P60–233, and internet www.census.gov. Primary source: Current Population Survey, March 2007 Annual Social and Economic Supplement

[a]Includes other government insurance, not shown separately. Persons are counted only once in the totals even though they may have more than one type of insurance

[b]Government health insurance includes Medicare, Medicaid, military health programs, State Childrens Health Insurance Program (SCHIP), and state health plans

[c]Persons of Hispanic origin may be of any race

Coverage varies greatly for different ages, races, and income classes. The availability of Medicare, Medicaid, SCHIP, and government health programs for military personnel greatly affects the variations in health insurance coverage among these groups since about one-quarter of the population has government insurance alone, or government insurance in addition to private insurance. For example, in 2006 nearly all persons 65 years of age and over (94.4%) had public coverage through Medicare (93.9%) or Medicaid (9.4%), and most of them (60.8%) had private insurance as well (Table 15.6).

Considering age variations further, according to data for the 5-year period, 2001–2006, non-coverage is greatest among the 18–24-year age group (29%), and then falls steadily with advancing age until it is negligible at ages 65 and over (1.5%), as expected. The great majority (88%) of children under 18 have health coverage, whether through Medicaid (27%), SCHIP (8%), Medicare (0.5%),

or private insurance (65%), with some children participating in more than one program, as these figures suggest. There are also substantial differences between the race/Hispanic origin groups. One-fifth of blacks and one-third of Hispanics, but only one-seventh of whites, lack coverage (Table 15.6).

Household-income classes show even wider variations in non-coverage, from only 8% of persons in households with incomes of $75,000 or more to one quarter of the persons in households with incomes less than $25,000. However, most (51%) uninsured persons live in households with incomes between $25,000 and $75,000. Nearly one-third (32%) of persons below the poverty line lack health coverage, but over one-half of poor families have some form of government-sponsored health insurance. These figures indicate that most of the "poor" are covered by public health insurance, the "rich" are covered by private insurance, and persons of moderate means are financially most pressed to participate in health insurance.

Studies by Collins et al./Commonwealth Fund (2006a, b) confirm these findings. They found that 41% of working-age persons with moderate to middle incomes ($20,000 to $40,000 per year) lacked health insurance for at least part of the year in 2005. This figure represents a dramatic increase over the figure in 2001, when just over one-quarter (28%) were uninsured. About one-quarter of the uninsured are lower-paid working people earning less than $25,000 annually.

The increase in the cost of health insurance has been outstripping the increase in household incomes and the increase in other health-care costs, This fact largely accounts for the steep upward trend in the number of uninsured persons. For many persons the cost of health insurance amounts to as much as one-quarter of their incomes. Most persons lacking health insurance live in households that simply earn too little money to buy health insurance even though someone in the household is working. In addition, increasing numbers of employers are dropping health coverage for their employees or are offering plans that are too expensive for them. Only about 60% of workers are now covered.

As a result, many of the uninsured secure medical attention at the emergency departments of hospitals but usually for a more severe condition that could have been treated earlier and at far lower cost. The costs for any free emergency services are absorbed into the health-care costs of the insured and into public taxes.

Many middle-class persons who lack health insurance are burdened by medical debts. According to Doty et al./Commonwealth Fund (2008), in 2007 an estimated 41%, or 72 million, of all adults under age 65 reported medical-bill problems or accumulated medical debt. In addition, 7 million persons aged 65 years and over had medical-bill problems or accumulated medical debt. A much larger share of the uninsured and underinsured was among these than the insured. High out-of-pocket spending on health matters and slow growth in real incomes have contributed to the rise in medical debt as well as personal bankruptcies and sacrifice of health care for other needs.

Some Considerations in Redesigning the U.S. Health Insurance System

The U.S. program of health insurance and health care is patently dysfunctional, whether judged from the view of numbers of persons excluded, numbers underserved, and numbers under threat of losing their health insurance through loss of their jobs, excessive and spiraling costs, the unsatisfactory quality and scope of care and services, inadequate health outcomes, fragmentation of the services within the system, the degree of waste in health-care delivery, the high rate of iatrogenic, nosocomial, and pharmacological errors, socioeconomic inequalities in access, quality of care, and outcomes, and the lack of a modern electronic medical information system. The United States spends about twice as much per person on health care as the industrialized countries (i.e., median for OECD countries) while, according to the principal indexes of health and longevity, it lags behind them. Questions can be raised with respect to the effectiveness of employer-based insurance, the voluntary nature of participation in health insurance programs, the basis of financially supporting the system, ways of achieving universal health insurance, the effect of a market-based system on the quality of health care, the merits of a single-payer system, and the role of privatization.

Whatever system is devised, it may be reasonably argued that, as a result of the lack of universal health insurance, much necessary health care is denied to many persons and families. Accordingly, this would appear to be a required feature of a new system. An insurance system is not financially viable in the long term if the segment of the population that is the most healthy – young adults – can choose not to participate. Medicare has proved to be a model as a one-payer system with modest administrative costs (3%)[10]; but with an enrollment that is composed of the least healthy age segment of the population, is surviving an increasing numbers of years to older ages, and seeks the most technologically advanced care, the system is financially unsustainable. Persons 65 years old and over spent on the average $8,776 per person as compared with $7,071 for the population as a whole (Kaiser Family Foundation 2009). In sum, sustainability is dependent on universal coverage and the general pooling of participants, and universal coverage is dependent on mandating participation.

An employer-based system alone cannot be sustained or justified. Many small businesses will not participate under the present circumstances; coverage is too expensive for many employers and employees. Many low-paid employees, the unemployed, and nonworkers, especially disabled nonworkers, are in effect excluded. Proposals have been made to bring more employers into the system. They include forcing employers to provide health insurance or contribute to a national or state fund that would cover uninsured workers, providing grants to employers to set up or participate in HMOs (i.e., health maintenance organizations), and offering employers tax rebates or exemptions, vouchers, and other incentives.

[10]This claim is disputed by supporters of a private health-insurance system. They maintain that Medicare's administrative costs are artificially low and should include such other costs as those of collecting the taxes and the increased taxation required to operate the program.

The private component of the U.S. health-care system offers a multitude of health care management plans that are confusing to the consumer, control closely the decisions of physicians with respect to the services that can be offered, and have relatively high administrative costs (about 13%). The U.S. health-care system allows private-sector health organizations great freedom to make their own determinations with regard to medical interventions and "medical necessity." The proponents of a privately insured health-care system maintain that it allocates costs in accordance with benefits and is therefore more efficient than a public system.

A system of national health insurance covering the entire population, with a single (i.e., government) payer, even with provision of health care by private health-care providers, is neither politically feasible nor acceptable to the American public. Most insured Americans say they like the health coverage they have even though it is too expensive. A single-payer system would involve a massive change in the structure and financing of the U.S. health-care system. It is opposed by the insurance industry and most doctors, as they had opposed Medicare nearly a half century earlier. A single-payer system could also alienate many patients and their doctors because close limits on the amounts and types of services would have to be established to keep costs from skyrocketing. Doctors would have to deny certain services to patients because the services were not medically justified, some services did not conform to guidelines on comparative effectiveness, and, in general, costs had to be controlled.

A system of voluntary personal medical accounts, which grants a health allowance to each person and so imposes a degree of responsibility on the health consumer, has been proposed, but this will be inadequate to solve the problem of making health insurance affordable. A reasonable and appropriate solution is to tax employer-sponsored health benefits, but this is not politically feasible because it would be considered a tax increase. Another, very different, possibility is mandatory universal health insurance with government, private plans, and the individual sharing the costs. Some states have adopted their own health care-systems following these principles, using managed health-care organizations that monitor the types of services offered, and setting fees for each type of service, in collaboration with employers and the Medicaid and Medicare programs. The states of Massachusetts and Vermont have initiated such programs and other states (e.g., California) are considering them as well. A system of universal health insurance with mixed public and private management may be politically and economically feasible at this time for the nation. Proposals for redesigning the U. S. health system have been set forth in AARP (2008) and the Alliance for Aging Research (2003). Now let us consider how some foreign countries manage their health needs.

Comparison of National Health Systems

National health insurance has been widely instituted abroad – in Canada and the countries of Europe, even the countries of Latin America and parts of Asia. Every industrial country in the world has a comprehensive system of national health

insurance except the United States. Proposals for national health insurance in the United States often invoke critical comments about the systems in other countries such as Canada, Great Britain, Switzerland, the Netherlands, and Australia. France's system is arguably the best system in the world, but the critics choose to overlook it. The foreign systems are criticized for requiring long waits for medical appointments and not providing the same specialized care as is available in the United States (Oneill and Oneill 2008). Physicians may not be available at night or on weekends and, because of low compensation, there may be a shortage of specialists. In the face of tightening budgets, these countries have had to whittle away somewhat at various services.

On the other hand, the countries with national health insurance systems secure better outcomes at lower costs than the United States. They all have life expectancies higher than, or nearly the same as, the United States, and lower or nearly the same infant mortality rates. The United States spends more on health care per capita than all other nations by far but ranks 31st in life expectancy and 40th in childhood survival to age five.[11] Many countries in Latin America, such as Mexico, Argentina, Cuba, and Costa Rica, earmark half or less of the share of their GDP for health care as the United States and yet have life expectancies at or only a little below that of the United States.

The systems differ somewhat from one another, but they all have in common a greater involvement of the national government in the provision of health care, and they usually have a greater coverage of health costs through taxes, than the United States. None of these other countries has an uninsured population, as the United States does. They differ generally on where they fall on the public-private continuum. United States is at one extreme of the western industrialized countries and Great Britain is at the other. Great Britain has a single-payer system funded by general revenue. It is a compulsory system of national health care in which the individual can obtain free medical attention from any participating hospital and doctor, and the cost is covered by the national government and local taxes. A small charge is levied for some services. Most physicians and nurses are government employees and receive salaries with specified terms and conditions of employment. Most hospitals are publicly owned or accountable to the government. Waiting lists are common, especially for an appointment with a specialist. There is rationing for certain expensive procedures for the terminally ill. If Britons choose, however, they can secure health service at their own expense from their private doctor. This gives patients access to higher quality care without having to wait as long. Because of severe budgetary problems various features of the British system have been privatized in recent decades.

Canada and France fall midway between the United States and Great Britain with reference to the public-private continuum. Canada has a system of national health insurance in which the federal government pays the costs, the provinces administer

[11]This ranking includes several small city-states but the U.S. ranking would still be very low if they were not included.

the program, and the delivery of health services is private. Health services are provided mostly by private, fee-for-service doctors and by private not-for-profit hospitals. The French system is financed largely by social security payroll taxes, and health services are provided mostly by private fee-for-service doctors, as in Canada, and a combination of public and private for-profit hospitals. All of the population is covered under a compulsory universal scheme, and benefits include the services of general family physicians and specialists, laboratory tests, diagnostic procedures, prescription drugs, dental services, sick days, and maternity leave and care. Patients are reimbursed largely but not completely; they must pay 20% of hospital expenses.

The Swiss are required to buy health insurance from private insurance companies and, if they cannot afford to do so, they can secure a subsidy from most cantons. The Swiss also pay for about one-third of their health-care services by themselves. The Dutch government, on the other hand, funds persons to purchase their own health insurance, to a large extent, from private insurance companies, and patients pay only 8% of their health-care bill. In both these countries, Switzerland and the Netherlands, private insurers are paid more for participants with health problems through various risk-adjustment systems (e.g., establishment of a risk-equalization fund in the Netherlands), but the governments rule on the minimum benefit packages and the prices the private companies can charge.

Recent research provides evidence of the superiority of the British system over the U.S. system in health outcomes although the United States spends twice as much per person on health care as England does (Banks et al. 2006). Health data for a nationally representative sample of white persons 55–64 years of age were examined for both countries. The results showed that middle-aged whites in England are considerably healthier than middle-aged whites in the United States. Americans have more heart disease, cancer, lung disease, diabetes, and other diseases. The poorest third of the English had less cancer, hypertension, and diabetes than the richest third of Americans. The health differences between the two countries cannot be fully explained by differences in smoking, obesity, race, or socioeconomic status.

Effects on Mortality and Morbidity

An important question is whether having a universal system of health insurance makes a difference in the health and longevity of the insured individual and the insured population. A definitive answer to this question cannot be given, but suggestive evidence is provided by various data. First, there is the superior record of life expectancy and child mortality for countries with universal health systems noted above. Next, one may explain the sharp fall in several of the chronic degenerative diseases of later life beginning in the late 1960s in the United States, at least in part, by the availability of Medicare and Medicaid after 1965. It is reasonable to assume that persons lacking health insurance receive less of the care they need and hence are at greater risk of illness and earlier death. Collins et al./Commonwealth Fund (2006a, b) found that persons without health insurance were much more likely than those who were insured to forego recommended health screenings such as

Table 15.7 Percent of recommended care received, by type and mode of care, for the United States: 2003

Type or mode of care	Percent
Total	54.9
Type of care	
Preventive	54.9
Acute	53.5
Chronic	56.1
Mode of care	
Medication	68.6
Laboratory testing or imaging	61.7
Surgery	56.9
Counseling or education	18.3

Source: McGlynn et al. (2003), Tables 3 and 4. Reprinted with permission of the Massachusetts Medical Society

colonoscopies and mammograms, were less likely to have a regular physician, and if they had a chronic condition such as diabetes or asthma, were more likely to skip medications for their conditions and end up in an emergency room of a hospital.

Many persons do not receive the care health providers have recommended to them or follow the prescribed treatment. Lack of health insurance is a likely common cause. Table 15.7 sets forth the percent of recommended care received for various types and modes of care, independent of the availability of health-insurance coverage. On average only about 55% of recommended care is being received, whether the care recommended is preventive or therapeutic, including surgical procedures. Patients who go untreated often suffer from more serious and more expensive health problems later.

Long-Term Care Financing

Long-term care (LTC) is expected to emerge as a major domestic public-policy issue in the next several years as the baby boom cohorts reach old age. A substantial and growing share of its members will require support by others as disability overtakes them in their more advanced years. As the need and costs of long-term services increase, political pressure will mount for policy-makers to design programs and financing to deal with this issue. The current and prospective facts are clear. Currently about 10 million Americans require some form of long-term assistance with Activities of Daily Living (ADL) and, as the population 65 years and over more than doubles over the period 2008–2040 from 38 million to 80 million, the population needing long-term care is also expected to grow sharply, possibly also doubling (Polivka 2008.)

Informal aid, that is, care by unpaid caregivers, mainly wives and daughters, constitutes 70–80% of all long-term assistance and was worth over $350 billion in 2006, or four times the amount of privately paid formal care (Johnson et al. 2007). The informal caregiver population will probably contract in the future as family size

falls and high levels of divorce disrupt family lines of caregiving. Hence, the share of the formal part of LTC – both home-based and community-based services (HCBS), and nursing-home care – and the costs of paid LTC services will grow sharply. About 75% of all expenditures on LTC currently support nursing home care and 25% covers LTC services in the community, mainly in-home services (Tritz 2006; Polivka 2008). It is clear that the increasing need for low-cost long-term care for many Americans will become a pressing issue that the federal government will have to confront.

The evidence for policy-makers is quite cogent in calling for a shift of public financial support from nursing homes to HCBS. Part of the change needed is an expansion of LTC insurance, which is now purchased by only 11% of the elderly and should be made more affordable through tax incentives and inclusion of LTC insurance in more group insurance plans. It is desirable both from a financial and personal point of view to extend HCBS. The disabled prefer home-based and community-based services. Such services are more cost-effective than nursing-home care and, in the few states where this shift has been instituted, it has proved successful. Resistance to such a shift comes from the nursing-home industry and state Medicaid offices, while the aging service agencies in most states are too weak to prevail. Additional details on the relevant research, the financial aspects of the issue, and the experience with implementing the move toward HCBS are presented in Polivka (2008).

Demographic and Economic Factors Affecting U.S. Entitlement Programs

Aging of the Population

The demands on the U.S. entitlement programs for the elderly, namely Medicare, disability insurance, and Social Security retirement, are affected by several demographic and socioeconomic factors that will play a major role in the future costs of these programs. Among these factors the age structure of the population merits prime attention. The age structure of a population reflects the net effect of the components of population change – mortality, fertility, and migration – on a population over a period of time. I have already considered these demographic relationships in Chap. 12, but review this material briefly here.

The populations of the United States and other industrial countries have been rapidly aging over the last century; that is, the share of the total population 65 years and over has been rising at a fast pace. Concomitantly, the share of the population 18–64 years of age – traditionally taken to represent the principal working-age population – has been changing very little. As a result, there are relatively far fewer persons of primary working age now in relation to the number in the principal retirement ages than there were several decades ago.

As explained in Chap. 12, the historical decline in the birth rate has been the principal factor in the long-term aging of the population. Declines in the birth rate since the early 1960s have reinforced this historical trend. In the years since the late 1960s, however, sharp declines in death rates at the older ages have become a dominant factor in the aging process. The effect of these factors has been reinforced by the declines in mortality rates at the middle ages, which have brought more survivors to old age. With the advent of the huge baby-boom cohorts of 1946–1964 and the changes in mortality noted, we can expect a massive increase in the number of persons 65 years of age and over in the two decades after 2011. The combination of these developments with the historical and recent declines in the birth rate will cause a sharp rise in the share of elderly persons in the total population. These same factors explain the decline in the share of the population below age 65.

Economic Dependency

As a result of the changing age structure of the population during past decades and the relative stability of the labor force participation ratios of the working-age population, a shift in the relative number of persons in the labor force and the number of elderly persons not in the labor force – the principal beneficiaries of the entitlement programs – is emerging in the United States and the other more developed countries. The numbers of workers who must provide the funds needed to support the present pay-as-you-go Social Security (SS) retirement system, disability insurance (DI) program, and Medicare, are projected to decline relative to the number of persons who must be supported. This relation can be expressed either as the ratio of elderly nonworkers to all workers, or as the ratio of beneficiaries from the SSA trust funds to its contributors. The formulas are:

$$\text{Elderly economic dependency ratio} = \text{Persons not in labor force 65} + \div \text{Total labor force 16} + \tag{15.6}$$

$$\text{OASDI beneficiary ratio} = \text{Beneficiaries of all ages} \div \text{Contributors of all ages} \tag{15.7}$$

The ratio of persons 65 and over not in the labor force to all persons in the labor force (16 years and over) was 14 (per 100) in 1950 and 21 in 2004, and is expected to rise to 37 by 2050 (Table 15.8 and Fig. 15.2.). The corresponding economic support ratios, the reciprocal of the economic dependency ratios, are 7.1, 4.7, and 2.7; these represent the number of workers for each nonworking elderly person. The dependency ratios based on SSA data are 6 (per 100) in 1950, 30 in 2004, and 49 in 2050 (Table 15.8 and Fig. 15.3). Today 3.3 contributors are supporting each beneficiary, but in 2050 only 2.0 contributors will be supporting each beneficiary. These figures indicate a tremendous increase in elderly dependency in the next half century.

Table 15.8 Elderly economic dependency ratios and beneficiaries/contributors ratios, for the United States: 1950–2004, and projections, 2010–2050

Year	Not in labor force 65^+ per 100 in the labor force 16^+[a]	Beneficiaries per 100 contributors[b]	Contributors per 100 beneficiaries[b]
1950	14.0[c]	6.1	1639.3
1975	20.7	31.1	321.5
1990	22.1	29.5	339.0
2000	21.6	29.1	343.6
2004	21.2	30.2	331.1
2010	21.3	31.5	317.5
2020	25.9	38.8	257.7
2030	33.1	46.0	217.4
2040	35.9	48.5	206.2
2050	36.8	49.1	203.7
Percent change,			
1950–2004	+51	+395	−80
2004–2050	+74	+63	−39

Source: Toossi, M. (2005. November). Labor force projections to 2014: Retiring boomers. *Monthly Labor Review, 128*(11), 25–44, Table 10 Toossi, M. (2006, November). A new look at long-term labor force projections to 2050. *Monthly Labor Review, 129*(11), 19–39, Table 6 U.S. Social Security Administration, *The 2006 annual report of the board of trustees of the federal old-age and survivors insurance and disability insurance trust funds*, Washington, DC: U.S. Government Printing Office

[a]Labor force estimates from the Current Population Survey; intermediate series of projections from 2010 to 2050. Estimates are adjusted to include the institutional population in the "not in labor force, 65+" population and the Armed Forces in the "labor force, 16+"

[b]SSA beneficiaries include both recipients of OASI and DI

[c]Estimated

Disability Retirement and Disability Measures

Disability retirement poses a special strain on the financial solvency of a retirement system. Disabled workers exit the labor force prematurely, typically well before the regular age for retirement. Then they begin drawing benefits from the system and cease contributing to it. They may receive generous benefits at an early age. Moreover, the system presents a strong temptation for workers to misrepresent themselves as disabled so that they can retire early.

Various types of dependency ratios and rates may be defined in terms of the disabled population and the worker population. A disability-worker dependency ratio for a given year can be calculated as the ratio of the number of disabled workers receiving benefits under Social Security to the number of workers covered by the Social Security system in that year (per 100):

$$\frac{\text{Disabled beneficiaries}}{\text{Covered workers}} * 100 \tag{15.8}$$

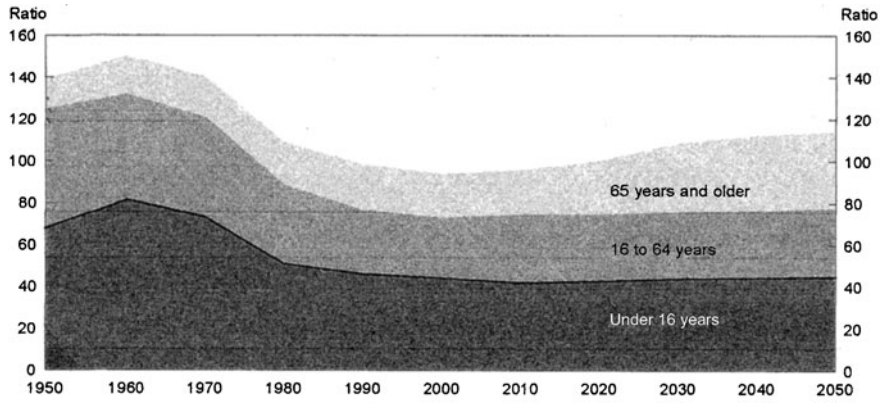


Fig. 15.2 Economic dependency ratios for broad age groups: United States, 1950–2000, and projections, 2010–2050 (Note: An age-specific economic dependency ratio is the number of persons not in the labor force in an age group per hundred persons in the labor force. Source: Toossi, M. (2006, November). A new look at long-term labor force projections to 2050. *Monthly Labor Review, 129*(11), 19–39, Table 6)

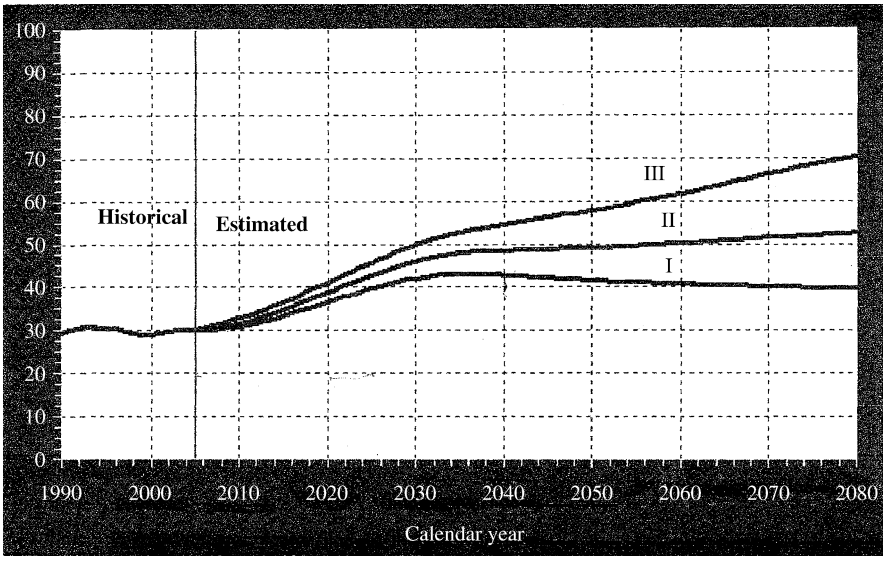


Fig. 15.3 Number of OASDI beneficiaries per 100 covered workers: United States, 1990–2005, and projections, 2005–2080 (Note: The three series represent alternative projections based on alternative assumptions regarding fertility, mortality, and migration. Series I is the low-cost series, with assumptions of high fertility and small reductions in mortality, Series III is the high-cost series, with assumptions of low fertility and large reductions in mortality, and Series II is the intermediate series; Source: U.S. Social Security Administration. (2006). *The 2006 Annual Report of the Board of Trustees of the Federal Old-age and Survivors Insurance and Disability Insurance Trust Funds*, Washington, DC: U.S. Government Printing Office, Figure IV.B2. See also Table 15.8 in this chapter)

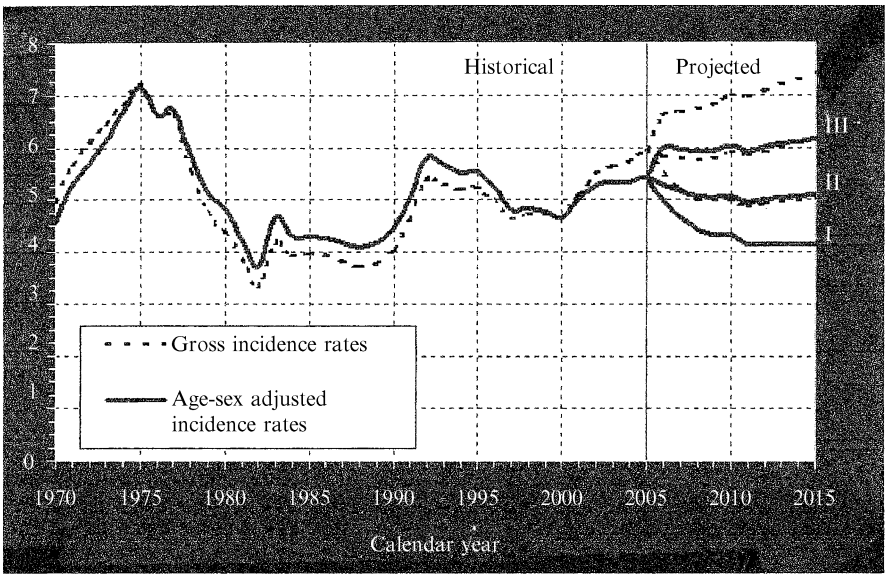


Fig. 15.4 DI Disabled-Worker Beneficiary Incidence Rates: United States Social Security Area, 1970–2015 (Notes: Number of annual disabled-worker beneficiaries per 1,000 covered workers exposed to disability: 1970–2005, and projections 2005–2015. Figures show both gross, or unadjusted, incidence rates and age-sex adjusted incidence rates; standard population is age-sex disability-exposed population on January 1, 2000; Source: U.S. Social Security Administration, *The 2006 annual report of the board of trustees of the federal old-age and survivors insurance and disability insurance trust funds*, Washington, DC: U.S. Government Printing Office, Figure V.C3)

This measure dropped between 1980 and 1990, but has been rising steadily since the latter date. The figure was 5.5 per 100 in 2005 as compared with 3.1 in 1990. Another measure is the disability incidence rate. The Actuary's Office, Social Security Administration, calculates disability incidence rates as the number of new disability awardees per 1,000 disability-exposed population:

$$\frac{\text{New disability awardees}}{\text{Total covered workers} - \text{total disabled beneficiaries}} * 1000 \qquad (15.9)$$

Figure 15.4 shows the trend of disability-worker incidence rates from 1970 to 2005, with projections to 2015. (The rates in this figure are age-sex adjusted on the basis of the disability-exposed population as of January 1, 2005.) The rate fluctuated between 4.5 and 6.0 for the years 1990–2005 and is projected to be about 5.2 in the year 2027.[12]

[12]Disability is defined under Social Security as "an inability to engage in substantial gainful activity due to medically determinable physical or mental impairment severe enough to satisfy the requirements of the program" and pertains to individuals who have not yet attained normal retirement age. The disability-exposed population is the disability-insured population that is not currently entitled to disability-worker benefits.

Other Socioeconomic Factors

At present there is little basis for believing that expected increases in longevity and in years of healthy life will induce many workers to continue working longer than their predecessors did and will thereby contribute to restoring a more favorable balance between workers and dependents. For over half a century the average age at retirement was falling (by about 5 years from 1950 to 2005), while life expectancy was rising (by 12 years) and, presumably, the nation's health was improving (Gendell 2008; Gendell and Siegel 1992). With reduced Social Security benefits on early retirement, perhaps a company pension, some savings, and occasional part-time work, most workers have decided that they would rather have the additional leisure beginning in their early 1960s than work longer. (Some biodemographers question whether such an early retirement age would continue if the rate of biological aging was slowed.)

Improved health reduces the need to leave the labor force early as disabled, but disability incidence rates have not changed much in the last few decades. The decision on the part of able-bodied persons whether to continue in the workforce or not is determined essentially as a tradeoff between the value placed on increased leisure and the ability to afford retirement, not primarily by one's perceived longevity. The principal factors affecting the decision to retire include the income earned on the job, other income received, availability at affordable health insurance, the worker's expectation of future economic return under retirement, and the value placed on additional leisure. Less determinative factors are the rising cost of living and the worker's prospective longevity. The last two factors have made the task of saving enough to afford to retire increasingly challenging, but they are not the principal factors in the decision to retire.

Some workers are now choosing to continue to work or to return to the labor force because they anticipate a more difficult economic future, particularly given the recession of 2008–2009. They may lack a company pension, the value of their pension may have shrunk, they may lack sufficient savings to retire comfortably for the longer period expected, or their investments may have lost value (Mermin et al. 2007). In addition, the sharp increases in the cost of health insurance have made it increasingly difficult for former workers to afford the premiums if they are not employed, or they are not offered employer-sponsored insurance at a subsidized price, prior to qualifying for Medicare. Moreover, companies have been reducing or eliminating health coverage of their retirees. As a consequence, labor force participation ratios of elderly persons have been rising in the years since 1995, even while the median age at retirement (about age 61) remains near the age of reduced benefits under Social Security (i.e., age 62). (See Gendell 2008.)

The social security systems of the United States and other western countries have built-in incentives that encourage older workers to retire early (Gruber and Wise 2004). These incentives intensify concerns about the financial viability of national retirement systems. This situation can be remedied in the United States by eliminating the reduced-benefit feature of OASDI or making it much less attractive, tying the normal age of retirement more closely to increasing life expectancy or perhaps to healthy life expectancy, and removing other incentives to early retirement.

Raising the retirement age. Earlier I described two widely divergent views regarding the control of the future costs of Social Security and Medicare. According to the second view, that supported by Manton et al. (2006), the health gains achieved and anticipated justify raising the normal age of retirement beyond the present mandated 67 years in 2027 to age 70 or more. Their specific proposal is to raise the SSA normal retirement age to 70 years in 2005–2006, 72.0 years in 2022, and 77.0 years in 2080, on the basis of current estimates and projections of disability-free life expectancy. They note that increasing the normal retirement age to these advanced ages would provide the 8.8 years of Social Security benefits, on average, to persons in the non-disabled state that the original beneficiaries received when OASDI was initiated in 1935. They believe also that the Medicare and Medicaid programs may have been responsible for the recent large increases in healthy life expectancy, and should be improved by enhancing the quality of care and expanded by lowering the age of eligibility for Medicare to 62 years or even 60 years. The latter step would reduce the share of the U.S. population lacking health insurance and increase the pace of improvement in health at the later ages. In short, Manton et al. maintain that eliminating several years of retirement by tying in the age of eligibility for full Social Security benefits to historical trends in life expectancy, and lowering the age of eligibility for Medicare, would assure solvency of the three systems until late in the twenty-first century.

The weakness in this proposal is that many persons may choose not to work or could not work until such late ages. The legal option to retire early under reduced benefits has to be eliminated if workers are to be given strong inducements to retire at higher ages. The proposals to eliminate the reduced-benefits option and to use official life tables as a guide for determining the age of full retirement have been made earlier (Siegel 2002:602), but Manton et al. (2006) provides a more complete technical justification and a concrete proposal. The application of the proposal to the sexes, the races, and other groups that have life expectancies different than the general population requires further exploration.

Since 1935, when the Social Security program was enacted, life expectancy at age 65 has increased by nearly six years, and since 1965, when the Medicare program was enacted, life expectancy at age 65 has increased by nearly four years (Table 15.9). One variation of the plan to use shifting ages as the basis for receiving full benefits under Social Security is to preselect a fixed number of years until death (e.g., 15 years). The age in the life table corresponding to that fixed number of years would be selected as the age for full benefits under Social Security. For example, as shown in Table 15.9, in 2003 the age with a life expectation of 15 years was 69.9 years whereas in 1980 it was 67.3 years. These ages differ for the sexes, the races, and other groups. For example, in 1980 blacks had a life expectation of 15 years at age 65.7, which differed from the all-races figure by 1.6 years (Table 15.10). By 2003, when blacks had a life expectation of 15 years at age 68.7 years, the race difference had fallen to 1.2 years (Table 15.10). As indicated, some resolution of the issue of group differences in life expectancy at the higher ages would have to be made if this procedure is to be implemented.

Table 15.9 Age corresponding to life expectancies of 10 and 15 years, based on period life tables for the United States: 1940–2005

Year	$e_x = 10$ years	$e_x = 15$ years	e_0	e_{65}
1935			61.4[a]	12.5[a]
1939–1941	70.0	61.4	63.2	
1949–1951	71.7	63.1	68.1	
1959–1961	72.5	64.0	69.9	
1965			70.4	14.6
1969–1971	73.7	65.0	70.8	
1979–1981	75.9	67.3	73.9	
1989–1991	76.9	68.4	75.4	
1999–2001	77.0	68.9	76.8	
2005	78.6	70.3	77.8	18.7
Increase,				
1965–2005			7.4	4.1
1935–2005			16.4	6.2

Source: Based on U.S. decennial life tables, 1939–1941 to 1999–2001, and U.S. annual life tables for 1965 and 2005, published by NCHS
[a]Estimated

Table 15.10 Age corresponding to life expectancies of 10 and 15 years based on period life tables for whites and blacks in the United States, 1950–2005

	$e_x = 10$ years		$e_x = 15$ years		e_0	
Year	White	Black	White	Black	White	Black
1949–1951	71.7	73.6[a]	63.2	61.8[a]	69.0	60.7[a]
1959–1961	72.4	73.9[a]	64.1	62.9[a]	70.7	63.9[a]
1969–1971	73.6	74.6	65.1	63.4	71.6	64.1
1979–1981	75.9	75.7	67.4	65.7	74.5	68.5
1989–1991	76.9	76.1	68.5	66.0	76.1	69.2
1999–2001	77.0	76.1	69.0	67.0	77.4	71.7
2005	78.4	77.9	70.3	68.7	78.3	73.2
Increase,						
1950–2005	6.7	4.3	7.1	6.9	8.3	12.5

Source: Based on U.S. decennial life tables, 1949–1951 to 1999–2001, and U.S. annual life tables for 2005, published by NCHS
[a]For total nonwhite population

Raising fertility. Some "obvious" demographic solutions are not likely to be beneficial in confronting the solvency problem. The course of raising fertility to bolster the number of workers is not a viable one. First, a few decades would be required to make this solution effective. Next, women in the industrial countries are choosing to have between one or two children in their lifetime and, although some are having larger numbers, the average number of babies born per woman (1.6 in the MDC and 2.0 in the United States) is below replacement level (2.1).

There is little prospect that this trend will be reversed and, although women report that they consider more children to be ideal, they are not acting on this ideal. A substantial increase in the proportion of women with three or more children would be necessary to turn the tide, but Western women do not want this many children. According to Morgan (2003); Bumpass (1990), and others, low fertility will persist indefinitely in the West. While there may be short-term fluctuations in the birth rate, fertility is likely to remain around replacement level for the foreseeable future. Hence. there is little prospect that fertility will rise enough to become an important factor in the solution of the social-security financial problem. From extensive experience in Europe, it can be concluded that pronatalist policies and programs cannot be depended on to raise fertility sufficiently to change the worker/dependent ratio much.

Expanding the volume of immigration. Expanded immigration is often mentioned as a solution to the support problem in the United States. Theoretically a sufficient volume of immigrants with a youthful age structure would help reduce the dependency problem, but currently the numbers of immigrants are not adequate for this purpose. The volume of immigration would need to be vastly increased to change the dependency/support balance significantly. A doubling of the volume of immigration would not greatly improve the situation (Center for ImmigrationStudies 2007). Moreover, additional immigration would have too many undesirable side-effects. It would exacerbate the present intolerable social and economic burden at the local level (e.g., health, crime, social welfare, transportation, recreation); it would impose a further ecological strain on areas that are already suffering from crowding, pollution, and other environmental and "esthetic" ills; and it would be a potential drain on the social security system at a later date. Rather, we should, on these and other grounds (e.g., national security), be considering a reduction in the volume of legal immigration, if not a moratorium on it.

Increasing labor productivity. It may be argued that the proposals previously made for dealing with the expected severe imbalances in the financial support of the Social Security system, disability insurance system, and Medicare smack of demographic determinism (Friedland and Summer ca. 2000). Some have suggested that the problem could be greatly relieved and even solved by sufficient increases in labor productivity. Labor productivity is the output of a worker per hour. An increase in labor productivity would tend to raise workers' production, lead to an increase in hourly compensation, and hence contribute to tax revenues and the Social Security and Medicare Trust Funds.

This mechanism can be helpful but it cannot be depended upon to have a sufficiently ameliorative effect on the financing of the social security and health systems. The annual percent increase in output per hour in the business sector of the United States economy averaged 2.2% between 1980 and 2004 and was almost always positive, but it fell well below the average in several years (U.S. BLS 2005). Real hourly compensation in the business sector has roughly moved in tandem with worker productivity but at a substantially lower level (1.2%). Labor productivity may complement other more manageable factors in achieving solvency

in the entitlement programs, but history gives little confidence that it will be able to overcome the additional costs resulting from demographic and other pressures on the entitlement programs.

Broadening coverage. Not all workers are required to contribute to the Social Security Trust Fund. For example, state and local employees with their own retirement systems may not be covered under Social Security if these governments choose not to participate. More than three-fourths of all state and local employees are covered under Social Security, but the largest state, California, has opted out. Broadening compulsory coverage to include all employees, both public and private, would widen the base of contributors sufficiently to ease the burden on the national retirement program. In addition, as suggested earlier, lowering the minimum age of elderly health insurance to age 60 or 55 would bring in many more healthy people. The younger group would be a financial bonus to the system since it would contribute to the pool of premium-payers, would have fewer health problems, and so would make fewer claims.[13]

Less Developed Countries

The current situation in the Less Developed Countries (MDC) with respect to aging and dependency is quite different from that in the More Developed Countries (LDC). As we have seen, the MDC are well on the way to becoming aged populations with very high shares of elderly persons and very high dependency ratios. Up to the end of the last century, most LDC not only had extremely low percentages of persons aged 65 and over, but they were not aging. The numbers of elderly were growing rapidly, however. As explained in Chap. 12, recent declines in fertility and mortality in many of these countries have bought about temporary age-structural transitions with reduced percentages of children along with low percentages of elderly and low elderly dependency ratios (CICRED 2005). However, the current "window of opportunity" will come to an end in a few decades. Impending changes in age structure will soon raise elderly dependency ratios and lead to imbalances of workers and elderly dependents, creating serious problems in supporting the elderly and in providing health care for them. The current but transitory demographic

[13] A number of financial, actuarial, and managerial changes have been proposed to lessen the risk of insolvency of the entitlement programs. They include raising the maximum amount of payroll disbursements that is taxable, raising the payroll tax rate, reducing benefits and annuities, requiring electronic record-keeping throughout the health-care system, and monitoring disability retirements more carefully so as to return more persons to employment or regular retirement. All of these proposals would bring in more money with some assurance except electronic medical record-keeping. The latter change is not likely to save any money for Medicare and Medicaid, but it could improve the quality of health care. Further evaluation of these proposals is outside the scope of this book.

situation may be viewed as an opportunity for the LDC to restructure their social security systems to accommodate to their prospective needs (Mason and Lee 2006).

Like the governments of the MDC, the governments of the LDC have implicit or explicit policies supporting the reduction of mortality. The LDC face a dilemma in their population policies, however. Lower death rates contribute to higher growth rates, but many LDC have policies of reducing growth rates through support of family planning programs. It is not uncommon for governments that have an active policy of lowering childhood and maternal mortality to couple this policy with a policy of lowering growth rates through fertility reduction. The programs of lowering mortality and lowering fertility are inextricably linked in programs of reproductive health, which seek to help women plan the number and spacing of their births, lower maternal mortality, and reduce infant and child mortality.

It is often maintained that women in the LDC tend to have more children than they consider ideal in order to make up for excessive infant and child mortality, and that an important method of effecting a reduction in fertility in the LDC is to reduce infant and child mortality. There is little formal evidence to support this hypothesis when individual-level data are analyzed. Reductions in mortality, including infant and child mortality, occur in the context of broader economic and social changes. These changes may lead simultaneously to declines in childhood mortality, declines in mortality at higher ages, and reductions in the number of children women have. Alternatively, success in family planning programs may lead to declines in infant and child mortality and maternal mortality, and these may be brought about both by reductions in the number of children and improvements in their spacing. (For an opposing view, see Hossain et al. 2007).

Some Problems Facing the Health Systems of LDC

As explained in Chaps. 6 and 11, infectious and parasitic diseases have long occupied the attention of health authorities in the LDC. The chronic diseases of later life, however, have emerged in the last quarter of the last century as important causes of illness and death in these countries. Because the latter conditions had always been considered afflictions of the more affluent countries, only limited financial and political support for programs aimed at their prevention or control has been given by their governments to health agencies. Now, like the omnipresent, endemic infectious diseases, the chronic degenerative diseases are becoming significant components of the health profile of the LDC and adversely affecting the economies of these populations. This has yet to be widely recognized, however. The heavy burden of both the infectious diseases and the chronic degenerative diseases is sharply reducing worker productivity, and hence the gross domestic product, of these populations. The effects spill over internationally since investors abroad get less return on their investments in these countries and *their* economies are adversely affected.

Malnutritution, especially undernutrition, along with drug addiction, alcoholism, smoking, and even obesity, presents the LDC with a massive health-cost burden. The World Bank has estimated that undernutrition is costing the LDC 3% of their annual gross domestic product (GDP). More specifically, the African and South Asian economies together are losing as much as $36 billion a year, or 3% of their 2003 GDP, for this reason. Inasmuch as pregnant women and children under 2 years old are particularly vulnerable to the ill effects of undernutrition, the World Bank has recommended that more resources be devoted to these groups and less to older children in school-feeding programs. Malnutrition not only compromises the future physical development of the infants, but their intellectual development as well.

The sexual practices that have caused the spread of HIV/AIDS and other sexually transmitted diseases through the LDC, especially sub-Saharan Africa, are having deleterious consequences in all aspects of life in these countries. Infected workers have had to withdraw from the labor force in large numbers as their illness progressed and their productivity at work declined. Their productivity at home has also declined. Other members of the household have had to devote time and energy to care-giving activities, so that the economic and social responsibilities in the home have had to be reassigned. All this is taking an immense toll on the economy of these countries.

Paying for Health Care

In much of the world some variation of a public-private partnership pays for such health care as is available. In the LDC and the countries in transition the systems for the provision of health care and payment for it vary widely, but in most LDC user fees are charged. These are co-payments that may be large in some countries and small in others. In Africa the provision of health care is typically a public-private partnership and involves such user fees (Callahan and Wasunna 2006). In the countries of Eastern Europe, nominally health care is a free public service, but in fact patients "bribe" health providers to secure such services as are available. (In effect, the combination of private health insurance and Medicare, and deductibles and co-payments, serve to make the system in the United States a type of public-private partnership.) In some countries a free-market system is in effect. In Singapore the free-market system involves a medical savings program, participation in which is mandatory. In India the bulk of care is provided privately, and as a result a huge gap in the health care of the rich and poor has developed. In China the once widespread informal program of health care in the countryside ("barefoot doctor" program) has disintegrated and no formal program has replaced it. Traditional medicine and western medicine are practiced side by side in the two parts of the country and a large health gap has developed between urban and rural areas in access to adequate health care.

The Global Health Funding Crisis

A number of problems, both general and specific, define the global health funding situation, particularly in the LDC. Some have been mentioned. The most general one is the need to rebuild the health-care systems of the LDC, including the health infrastructure. Others are the emigration of skilled personnel, the glaring shortage of health workers, and hence the need for training a sufficient numbers of new health personnel; the failure to develop and manufacture new drugs appropriate for the LDC, and hence the serious lack of necessary drugs; the rise in chronic degenerative diseases together with the persistence of the endemic infectious diseases; making access to reproductive health universal; and the dwindling funding role of the WHO in relation to the role of other funding agencies. Underlying these "health" problems are the extreme poverty and illiteracy and the limited human rights and education of women and girls in many of the LDC.

Funding Problems

The United States and other industrialized countries have pledged money for health assistance, particularly for HIV/AIDS, to the LDC, but there is general recognition that the money pledged for combating global epidemics is insufficient to prevent a global public health crisis. For example, since its inception in 2002, the Global Fund to Fight AIDS, Tuberculosis, and Malaria (GFATM) has received pledges of only $5.4 billion through 2008 of the $10 billion requested (*OECD Observer parris* 2004).

To deal with the problem of rising chronic diseases, major investments need to be made in primary-care treatment focusing on people at elevated risk for cardiovascular diseases and diabetes. Specifically, international efforts have to be made to (1) support the distribution of the various medications that deal with chronic health conditions – antihypertensives, statins, aspirin, and smoking-control products, (2) curb the availability of tobacco, a principal agent in causing sickness in these populations, and (3) invest in research for new cheap medications useful in the LDC for treating the chronic degenerative diseases (Yach et al. 2003). The two major strategies adopted internationally, the Framework Convention on Tobacco Control (FCTC), adopted in 2003, and the Global Strategy on Diet, Physical Activity and Health, adopted in 2004, have yet to be implemented. The FCTC, which is designed to control the use of tobacco by increasing excise taxes, instituting marketing restrictions, and establishing smoke-free places, was expected to take effect in early 2005.

The LDC suffer from a lack of late-model drugs and vaccines. According to the *OECD Observer parris* (2004), between 1975 and 1997 only 13 of the 1,223 new medicines commercially produced were designed to treat tropical diseases and no new class of TB drugs has been produced since 1966. Diseases common only in poor countries do not generally induce pharmaceutical companies to develop new drugs.

To deal with this problem, the *OECD Observer parris* suggests that incentives for the pharmaceutical firms be created by having bodies such as GFATM make pre-commitments to purchase newly developed medicines and vaccines.

To deal with the problem of emigration of skilled personnel, the *OECD Observer parris* (2004) maintains that a multilateral solution based on ethical recruitment protocols and compensation for countries losing skilled personal is urgently needed. The UK government's 2001 Code of Practice for the international recruitment of healthcare professionals strongly condemns recruiting in the LDC except under strict conditions. This is a short-term solution. A long-term solution must be based on creating incentives for local staff to stay in their home countries, and this requires resources that many countries lack.

Reproductive health. Funding has dropped by 30% since the mid-1990s for international family planning programs in the Less Developed Countries (Bongaarts and Sinding 2009). A number of factors have contributed to this change, including the diversion of attention and funds to the AIDS epidemic and the persistent opposition of conservative governments and institutions, in particular the Bush administration and the Vatican. The Obama administration is poised to change this situation. Early in 2009, President Obama signed an executive order removing restrictions that prevented U.S. agencies from giving aid to international organizations that promote or perform abortions with their own funds. The restrictive policy was originally promulgated by President Reagan and it resulted not only in restricting the free speech of international family planning organizations but also in impeding comprehensive reproductive health services to women in the LDC, including the delivery of contraceptives.

The Obama administration has announced that it unequivocally supports the worldwide consensus that reproductive health is a basic human right and that universal access to it is critical not only for individual health but also for family well-being and economic development. This is consistent with the principles laid out in the UN International Conference on Population and Development (Cairo 1994) and with the UN Millennium Development Goals. To implement this policy, the Administration is being called on by a coalition of nonprofit organizations to work with the U.S. Congress to fund international family planning programs at $1 billion in order the reverse the decade of limited funding that has just passed.

In spite of the widely held belief that family planning programs are not cost-effective, the World Bank estimates the cost of family planning at $100 per life-year saved, which compares favorably with other health interventions. Family planning is much more cost-effective than antiretroviral treatment of AIDS (Bongaarts and Sinding 2009). Furthermore, large-scale national family planning programs have, for the most part, been quite successful and explain a large share of the decline in fertility and infant mortality in the LDC in the last several decades. For example, in Mexico, where the use of modern contraceptives nearly doubled, infant mortality fell by 70% between 1970 and 2005. Similar results have been observed in Bangladesh, Egypt, Thailand, and other countries (Population Council 2010).

Funding Agencies

In the LDC, funds for health programs are provided by governments for their own people, private citizens in these countries, the UN and WHO, the United States government, and a variety of other agencies. The other agencies include other international organizations such as the World Bank, for-profit pharmaceutical companies, and nonprofit nongovernmental organizations such as private foundations, charitable organizations, public service organizations, and even a nonprofit pharmaceutical manufacturing company. There is a growing involvement of the World Bank in health programs. Major U.S. private philanthropic foundations have entered the world health arena, testing innovative strategies in an effort to tackle global health problems. The Bill and Melinda Gates Foundation, for example, provided funds to a nonprofit company manufacturing drugs, and these drugs are sold by the company at not-for-profit prices in the LDC. The John D. and Catherine T. MacArthur Foundation, the William and Flora Hewlitt Foundation, and the David and Lucille Packard Foundation jointly granted nearly $100 million in 2008 to support UN's fourth and fifth Millennium Development Goals, two goals promoting reproductive health, including the improvement of maternal health and the reduction of child mortality (see below). I have already noted the major involvement of the U.S. government in AIDS relief in the LDC. The current U.S. program is called "The President's Emergency Plan for AIDS Relief," or PEPFAR.

The actions of these agencies have shifted the balance of influence in international public health away from the World Health Organization and have raised questions about the latter's role. With an annual budget of only $6.6 billion, WHO is constrained financially. Furthermore, it has no enforcement power; that is, it cannot require member countries to accept its policies or programs. The future role of the WHO in managing and resolving the health problems of the world is in question.

International Goals and Assessment of Progress

Millennium Development Goals

In September 2000 a gathering of Heads of State adopted the UN Millennium Declaration, which was elaborated into the United Nations System's Millennium Development Goals (MDGs). The implementation of these goals and success in achieving them could ameliorate greatly the health conditions in the LDC. The goals include the eradication of extreme poverty and hunger, reductions in the mortality of young children and maternal mortality, and halting the spread of

HIV/AIDS, malaria, and tuberculosis. The eight Millennium Development Goals as promulgated by the UN are (WHO 2000; CICRED 2005):

1. Eradicate extreme poverty and hunger
2. Achieve universal primary education
3. Promote gender equity and empower women
4. Reduce child mortality
5. Improve maternal health
6. Combat HIV/AIDS, malaria, and other diseases
7. Insure environmental sustainability
8. Develop a global partnership for development

All 191 UN member states and all the world's leading development institutions have pledged to meet these goals by 2015. The goals were to be implemented in terms of 16 targets and 48 indicators. Three of the 8 goals (nos. 4, 5, and 6), 8 of the 16 targets, and 18 of the 48 indicators are directly related to health. Health is also an important contributor to several other goals and is a beneficiary of these other goals (nos. 1, 2, and 3). The MDGs do not encompass all aspects of health, making no specific reference to effective health systems, reproductive health, or noncommunicable diseases. Nevertheless, the MDGs may be interpreted as encompassing a wide range of health outcomes that development should achieve, such as reducing the number of women dying in childbirth, increasing the number of children surviving the early years of life, managing the HIV/AIDS pandemic, assuring access to necessary drugs, and achieving better health as a step in the reduction of poverty.

Assessment of Progress at Mid-Decade

WHO has issued a mid-term report on the progress toward achieving the health goals of the Millennium Development Goals (World Health Organization 2005). WHO's summary of its report is as follows:

The MDGs are far from being met. While some LDC have made impressive gains, many more are falling behind. Progress is particularly slow in many countries of sub-Saharan Africa. With respect to Goal No. 4, reduction of child mortality, nearly 11 million children under age 5 still die every year. Some reduction has taken place in child mortality, but in 16 countries, 14 of which are in Africa, levels of child mortality are higher in 2005 than in 1990. In some LDC, because girl babies are not favored, female infanticide is still practiced. With respect to Goal No. 5, improvement of maternal health, more than 1/2 million women still die in pregnancy and childbirth every year, despite increases in the rate of attended deliveries in some areas, i.e., Southeast Asia and North Africa. Maternal death rates are 1,000 times higher in sub-Saharan Africa than in high-income countries. With respect to goal No. 6, combating HIV/AIDS, malaria, and other diseases, the situation is bleak in many countries, despite success in some others. As we saw, the HIV/AIDS global pandemic has reversed life expectancy in several African countries.

As stated, the money pledged for combating global epidemics and failing health systems by the United States is considered to be insufficient to prevent a global public health crisis. The goals of rebuilding the health systems of the LDC and providing them with adequate human and physical resources cannot be achieved without a substantial increase in financial resources. Failure to respond to the global health funding crisis will allow potentially catastrophic diseases to ferment unchecked, especially the multi-drug resistant diseases.

The 2003 WHO *World Health Report* pointed out the urgency of rebuilding the health systems of the LDC, but noted that this would require a substantial increase in resources. The report notes further that current health spending in most low-income countries is insufficient for the achievement of the MDGs and calls for doubling of aid from $US50 billion to $US100 billion a year for the MDGs as a whole, total spending of $US20 billion per year on HIV/AIDS by 2007, and a fivefold increase on donor spending on health. The chair of the WHO Commission on Macroeconomics and Health, Jeffrey Sachs (2003) reported that it would require an extra $25 billion annually from the rich countries to handle the health problems of the LDC. That is only about one thousandth of their annual income.

References and Suggested Readings

More Developed Countries

Health Promotion and Disease Prevention

Hayflick, L. (2000). The future of aging. *Nature, 408*, 37–39.

Institute of Medicine/National Research Council. (1985). *Health in an older society*. Washington, DC: National Academy Press.

Pinn, V. (2003). Sex and gender factors in medical studies: Implications for health and clinical practice. *Journal of the American Medical Association, 289*(4), 397–400.

United Nations. (2004). *World population policies 2003*. New York: United Nations.

U.S. Department of Health and Human Services. (2000). *Healthy people 2010: Understanding and improving health*. Washington, DC: U.S. Government Printing Office.

U.S. National Center for Health Statistics. (2001). *Healthy people 2000 final review: National health promotion and disease prevention objectives*. Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2007). Summary health statistics for U.S. adults: National health interview survey, 2006. By J. R. Pleis and M. Lethbridge-Çejku. *Vital and Health Statistics 10*(235) Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (2009). Deaths: final data for 2006. By M. Heron, D. L. Hoyert, S. L. Murphy, J. Xu, et al. *National Vital Statistics Reports 57*(14) Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Institute of Neurological Diseases and Stroke (NINDS). (2001). *Pain: Hope through research* (NIH Publication No.01–2406). Bethesda, MD: NINDS.

U.S. Public Health Service. (1979). *Healthy people: The surgeon general's report on health promotion and disease prevention*. Washington, DC: U.S. Public Health Service.

Wallace, S. P. (2000). American health promotion: Where individualism rules. Book review section. *Gerontologist*, 40(3), 373–377.

Health, Social Programs, and Retirement

AARP. (2008). *Building a sustainable future: A framework for health security*. Washington, DC: AARP.

Alliance for Aging Research. (2003). *Redesigning healthcare for an aging nation*. Washington. DC: A Congressional Forum Sponsored by the Alliance for Aging Research.

Anderton, D. L., Anderson, A. B., Oakes, J. M., & Fraser, M. (1994). Environmental equity: The demographics of dumping. *Demography, 31*(1), 229–248.

Banks, J., Marmot, M., Oldfield, Z., & Smith, J. P. (2006). Disease and disadvantage in the United States and England. *Journal of the American Medical Association, 295*(17), 2037–2045.

Bumpass, L. (1990). What's happening to the family? Interactions between demographic and institutional change. *Demography, 27*(4), 483–498.

Burkhauser, R. V., Couch, K. A., & Phillips, J. W. (1996). Who takes early Social Security benefits? The economic and health characteristics of early beneficiaries. *Gerontologist, 36*(6), 789–799.

Center for Immigration Studies. (2007, August). By Camarota, S.A. 100 million more: projecting the impact of immigration on the U.S. population, 2007 to 2060. Backgrounders and Reports. Washington, DC: Center for Immigration Studies. http://cis.org.

Feder, J. (2008). *Our troubled health care system. Why is it so hard to fix? Policy Brief 37*. Syracuse, NY: Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University.

Friedland, R. B., & Summer, L. (Eds.). (ca. 2000). *Demography is not destiny*. Washington, DC: National Academy on an Aging Society, See esp. pp. 51–54 and Chap. V.

Fries, J. F. (2003). Measuring and monitoring success in compressing morbidity. *Annals of Internal Medicine, 139*(5, Pt. 2), 455–459.

Gendell, M. (2008). The comeback of older workers in the United States. *Monthly Labor Review, 131*(1).

Gendell, M., & Siegel, J. S. (1992). Trends in retirement age by sex, 1950–2005. *Monthly Labor Review, 115*(7), 22–29.

Gruber, J., & Wise, D. A. (Eds.). (2004). *Social security programs and retirement round the world: Micro-estimation* (A National Bureau of Economic Research Conference Report. Chicago: University of Chicago Press.

Manton, K. G., Gu, X., & Lamb, V. L. (2006a). Change in chronic disability from 1982 to 2004/2005 as measured by long-term changes in function and health in the U.S. elderly population. *Proceedings of the National Academy of Sciences, 103*(48), 18374–18379.

Manton, K. G., Gu, X., & Lamb, V. L. (2006b). Long-term trends in life expectancy and active life expectancy in the United States. *Population and Development Review, 32*(1), 81–106.

Marmot, M. G. (2004). *The status syndrome: How social status affects our health and longevity*. New York: Henry Holt.

McGlynn, E. A., Asch, S. M., Adams, J., et al. (2003). The quality of health care delivered to adults in the United States. *New England Journal of Medicine, 348*(26), 2635–2645.

Mermin, G. B. T., Johnson, R. W., & Murphy, D. P. (2007). Why do boomers plan to work longer? *Journal of Gerontology: Social Sciences, 62B*(5), 286–294.

Morgan, S. P. (2003). Is low fertility a twenty-first century demographic crisis? *Demography, 40*(4), 589–604.

Oneill, J. E., & Oneill, D. M. (2008). Health status, health care, and inequality: Canada vs. the U.S., *Forum for Health Economics and Policy, 10*(1), (Frontiers in Health Policy Research), Article 3.

Serow, W. (ca. 2000). (n.d.). *Demographic aspects of health care access in the United States: 1990–2020* (Working Paper PI-93–6). Tallahassee, FL: Pepper Institute on Aging and Public Policy, Florida State University.

U.S. Bureau of Labor Statistics (BLS). (2005, June 2). *Productivity and costs, news*. USDL 05–964. www.bls.gov/lpc (See also U.S. Census Bureau, *Statistical Abstract of the United States: 2006*, Table 623).

Costs of Health Care

Bernard, D. M., Banthin, J. S., & Encinosa, W. E. (2009). Wealth, income, and affordability of health insurance. *Health Affairs, 28*(3), 887–896.

California HealthCare Foundation. (2008). *Health care costs, snapshot* (p. 101). Oakland, CA: CHCF. Cited in AARP Public Policy Institute, *Fact Sheet* 144, October 2008.

Collins, S. R., Kriss, J. L., Davis, K., et al. (2006a). Squeezed: Why rising exposure to health care costs threatens the health and well-being of American families. New York: Commonwealth Fund.

Collins, S. R., Davis, K., Doty, M. D., et al. (2006b). Gaps in health insurance: An all-American problem. New York: Commonwealth Fund.

Cutler, D. M. (2005). The potential for cost savings in Medicare's future. *Health Affairs, 24*(Suppl. 2), W5-R77–80, *Health and Costs of the Future Elderly*.

Cutler, D. M., & Richardson, E. (1998). The value of health: 1970–1990. *American Economic Review, 88*(2), 97–100.

Cutler, D. M., Rosein A.B & Vijan, S. (2006, August). The value of medical spending in the United States, 1960–2000. *New England Journal of Medicine*, *355*, 920–927.

Doty, M. D., Collins, S. R., Rustgi, S. D., & Kriss, J. L. (2008). *Seeing red: The growing burden of medical bills and debt faced by U.S. families* (Issue Brief). New York: The Commonwealth Fund.

Goldman, D. P., Shang, B., Bhattacharya, J., et al. (2005). Consequences of health trends and medical innovation for the future elderly. *Health Affairs, 24* (Suppl. 2), W5-R5–18, *Health and Costs of the Future Elderly*.

Gross, M. J., & Schwenger, C. W. (1981). *Health care costs for the elderly in ontario: 1976–2026. Occasional Paper 11*. Toronto, ON: Ontario Economic Council, See esp. Chap. 5.

Gruber, J. (2009). *Universal health insurance coverage: Progress and issues* (Policy Brief No. 41/2009). Syracuse, NY : Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University.

Hartman, M., Martin, A., Nuccio, O., & Catlin, A. (2010). Health-spending growth at a historic low in 2008. *Health Affairs, 29*, 147–155.

Johnson, R., Toohey, D., & Wiener, J. (2007). *Meeting the long-term needs of the baby-boomers: How changing families will affect paid helpers and institutions*. Washington, DC: Urban Institute.

Kaiser Family Foundation (KFF). (2009). *Health care costs: A primer. Key information on health care costs and their Impact*. Washington, DC: Kaiser Family Foundation.

Lubitz, J. (2005). Health, technology, and medical care spending. *Health Affairs, 24*(Suppl.2), W5-R81-R85, *Health and Costs of the Future Elderly*.

Lubitz, J., & Riley, G. F. (1993). Trends in Medicare payments in the last year of life. *New England Journal of Medicine, 328*, 1092–1006.

Miller, T. (2001). Increasing longevity and Medicare expenditures. *Demography, 38*(2), 215–226.

Polivka, L. (2008, September 24). *The experience of public and private long-term care financing and Medicaid reform*. Presented at the Workable Solutions for Long-Term Care Forum, The Heritage Foundation. Tampa, FL: School of Aging Studies, University of South Florida.

Riedel, M., & Hofmarcher, M. M. (2003). Austrian health expenditures exhibit an age profile. *Vienna Yearbook of Population Research 2003, 1*, 197–213. Vienna: Austrian Academy of Sciences.

Tritz, K. (2006). *Long-term care: Trends in public and private spending* (CRS report to Congress). Washington, DC: Congressional Research Service.

Seshamani, M., & Gray, A. (2005). *Effect of age and approaching death on hospital costs*. (Unpublished study). Oxford, UK: Health Economics Research Centre, University of Oxford.

Shekelle, P. G., Ortiz, E., Newberry, S. J., et al. (2005). Identifying potential health care innovations for the future elderly. *Health Affairs, 24*(Suppl. 2), W5-R67-R76, *Health and Costs of the Future Elderly*.
Siegel, J. S. (2002). *Applied demography: Applications to business, government, law, and public policy*. San Diego, CA: Academic, esp. Chap. 13.
U.S. Bureau of Labor Statistics (BLS). (2003). Consumer expenditure survey, annual. www.bls.gov/cex (See also U.S. Census Bureau, *Statistical Abstract of the United States: 2006*, Table 129.)
U.S. Congressional Budget Office. (2008, January 31). *Growth in health care costs*. Washington, DC: Congressional Budget Office, Statement of Peter R. Orszag, Director, before the Committee on the Budget, United States Senate.
Vladeck, B. C. (2005). Accounting for future costs in Medicare. *Health Affairs, 24*(Supp.2), W5-R94–96, *Health and Costs of the Future Elderly*.
Weinstein, M. (2005). Spending health care dollars wisely: Can cost-effectiveness analysis help? Policy Brief 30. Syracuse, NY: Center for *Policy Research*, Maxwell School of Citizenship and Public Affairs, Syracuse University.
Yang, Z., Norton, E. C., & Stearns, S. C. (2003). Longevity and health care expenditures: The real reasons older people spend more. *Journal of Gerontology: Social Sciences, 58B*(1), S2-S10.
Zhan, C., Friedman, B., Mosso, A., & Pronovost, P. (2006). Medicare payment for selected adverse events: Building the business case for patient safety. *Health Affairs, 25*(5), 1386–1393.
Zweifel, P., Felder, S., & Meiers, M. (1999). Ageing of population and health care expenditure: A red herring. *Health Economics, 8*, 485–496.

Less Developed Countries

Bongaarts, J., & Sinding, S. W. (2009). A response to critics of family planning programs. *International Perspectives on Sexual and Reproductive Health, 35*(1), 39–44.
Callahan, D., & Wasunna, A. A. (2006). *Medicine and the market: equity and choice*. Baltimore: Johns Hopkins Press.
CICRED (Committee for International Cooperation in National Research in Demography). (2005). Policy implications of age-structural changes. CICRED Policy Paper, No. 1. Paris: CICRED.
Hossain, M. B., Phillips, J. F., & Le Grand, T. K. (2007). The impact of childhood mortality on fertility in six rural thanas of Bangladesh *Demography, 44*(4), 771–784.
Mason, R., & Lee, R. (2006). Reform and support systems for the elderly in developing countries: capturing the second demographic dividend. *Genus, LXII*(2), 11–36.
Mosley, W. H., Bobadilla, J. L., & Jamison, D. T. (1993). The health transition: Implications for health policy in developing countries. In D. T. Jamison, W. H. Mosley, A. R. Measham, & J. L. Bobadilla (Eds.), *Disease control priorities in developing countries* (pp. 673–699). Oxford, UK: Oxford University Press.
Parris, B. (2004, May). False economies: A global health crisis *OECD Observer parris*, 243.
Population Council. (2010, January 8). pubinfo@popcouncil.org
Sachs, J. D. (2003, October 29). *Achieving the millenium development goals: Health in the developing world*. In Speech at the second global consultation of the commission on macroeconomics and health, Geneva, Switzerland.
United Nations. (2004). *World population policies 2003. E.04.XIII.3*. New York: United Nations.
World Health Organization. (1999). *Aging and health: A global challenge for the 21st century*. In Proceedings of a WHO symposium, Kobe, Japan, November 10–13, 1998. Geneva, Switzerland: World Health Organization.
World Health Organization. (2000). WHO and the millennium development goals. Geneva, Switzerland: World Health Organization. Internet: www.WHO.org.int

World Health Organization. (2003). *World health report*. Geneva, Switzerland: World Health Organization.

World Health Organization. (2005). *The score at half time. Millennium development goals report: 2005*. Geneva, Switzerland: World Health Organization. Internet: www.WHO.org.int.

Yach, D., Leeder, S. R., Bell, J., & Kistnasamy, B. (2003, January 21). Global chronic diseases. *Science, 307*(5708), 317.

Chapter 16
Applied Health Demography

Applied demography is that branch of demography concerned with applications of the methods, materials, and perspectives of demography to the problems of government, business, and other private organizations. When these problems relate to the health of the population or the health of some specific segment of the population, such as the labor force, they may be viewed as issues in applied health demography. Most of the applications of applied health demography relate to small areas so that this term is often used to include any analysis of geo-referenced small-area health data. The term small areas relate to such areas as metropolitan areas counties, and their subdivisions. Others, coming from a different perspective, call the field spatial epidemiology.

Area Population Estimates and Projections for Public Health Programs

Often for the solution of problems in applied health demography, it is necessary to obtain current population estimates and population projections for small areas. These are used to calculate various types of vital rates and measures of risk and to determine the need for health programs. Some examples of tasks that involve preparation of population estimates and projections in applications of health demography are:

Identifying various high-risk population groups by type of risk and neighborhood
Determining whether a public hospital system should be established in a city
Deciding on the optimum location of a new mental health clinic
Designing public health emergency services for a city
Determining the need to add a specialized department to a hospital

Applied demographers and spatial epidemiologists collaborate with one another and with other types of analysts in dealing with such issues. Some of the applications of

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_16, © Springer Science+Business Media B.V. 2012

health demography are concerned with demographic and socioeconomic segments of a population, such as the labor force, an occupational group, poor persons, and householders. Projecting the health costs of the retirees of a company for the next decade illustrates such an issue. Another is determining the size and socioeconomic characteristics of the population lacking health insurance in the counties of a state.

Establishing a New Health Facility or Expanding an Old One

The growth of certain industries is regulated by various government jurisdictions on the basis of estimates and projections of the need for their products and services by consumers. State laws restrict the entry of new companies into specified areas where the service is already provided by other companies, and limit the expansion of old companies in providing new services. Such regulation aims to control competition – to restrict or expand it, as seems desirable – to prevent the construction of publicly subsidized facilities that may be underutilized, to promote orderly, efficient growth, and to prevent unnecessary duplication of expensive services.

The requirement of state approval applies especially to industries calling for large and expensive facilities such as hospitals and nursing homes. The establishment of a new hospital or expansion of a major division of an old hospital is restricted by state law generally for the purpose of protecting existing hospitals from competition in the same service area where public demand for the service is not growing sufficiently to recommend it. Before moving into a new area or expanding in an old area, then, a hospital or nursing home must first obtain approval from the appropriate state regulatory agency to take this action. The state regulatory agency typically grants approval or denies it on the basis of estimates and projections of community need. In estimating and projecting the level of need, the prospective output of current providers may be compared with the prospective need for the products or services to determine whether a new facility is warranted.

Estimates and projections of population size and a few demographic characteristics (e.g., age and sex) for some designated geographic area – the presumed service area – are almost always required to evaluate the need for expansion of specified health services in the area. The service area is defined geographically in terms of land parcels such as counties, census tracts, health districts, or radial distances such as miles from a facility, and is defined demographically in terms of total population and possibly age-sex distribution. For the establishment of a new hospital, a relatively limited geographic area may be satisfactory. For the establishment of a medical specialty within a hospital or a medical-specialty hospital, however, the service area should be much larger since the new department or hospital may draw patients from a considerable distance away, e.g., across an entire metropolitan area.

Another indicator of need in addition to total population size, called a measure of utilization, reflects how well the services of the existing hospitals in the service area are being utilized. A common measure of utilization is the number of discharges

from a hospital per capita, especially in relation to the number of discharges the hospital could handle. Alternatively, the occupancy ratio of hospital beds, that is, the percent of beds that are occupied, may be used as such a measure.

For each type of industry, some unit of measurement of supply is employed. For health facilities, the unit of measurement is usually the number of beds per 1,000 population. Population is converted into beds per 1,000 population on the basis of the record of experience in the community relating bed requirements for given numbers of persons. For example, if three hospital beds are required for every 1,000 persons in a service area and the population of a service area is 48,000 persons, the number of hospital beds required in the area is 144:

$$3 \div 1000 = .003$$

$$.003 * 48{,}000 = 144$$

An existing hospital may request a judicial review of the matter if another hospital company is planning to enter the same "market area" or an established hospital is planning to extend or add a new department. Its legal brief would state the basis of its opposition to the establishment of the new hospital or the new department. The states vary in the type of judicial review required. In such a legal contest, the plaintiff organization must establish its "standing" to bring the suit, that is, it must demonstrate that there will be injury to its business or the quality of its service and that the proposed facility will have a direct role in causing the prospective injury. Usually, the plaintiff argues that the proposed facility would be located too close to one of its facilities, and that the prospective demand for hospital services based on the projected size and characteristics of the population of the service area would not require the additional facility. Further, the plaintiff may argue that its own facility can be readily expanded to fill any prospective demand and that public funds would be wasted in supporting the new facility. The defendant may note that, given the uncertain accuracy of the plaintiff's population projections, the plaintiff should prepare and present several alternative projections, perhaps also for a variety of proposed service areas.

An illustration of a legal contest in which demographic issues were central to the arguments of the contestants and the judge's decision is the suit of Maury Regional Hospital and Williamson Medical Center of Maury and Williamson Counties, Tennessee, respectively, against the Hospital Corporation of America (HCA). HCA had already secured from the State Health Services and Development Agency (HSDA) a Certificate of Need (CON) approving the building of a new hospital designated Spring Hill Hospital, which would straddle the county lines. The Judge ruled in September 2009 to throw out HCA's CON for Spring Hill Hospital, principally on the grounds that the population projections used by the defendant were grossly overstated and that those of the plaintiff were far more reasonable. She faulted the defendant's unsophisticated method of making its projections and found, on the basis of discharge and bed-occupancy data, that the facilities of the present hospitals were greatly underutilized and there would be a duplication of services if

a new hospital was built. Previously in 2007 an administrative law judge had ruled against the granting of a CON to HCA to build Spring Hill Hospital on the same grounds, but HCA appealed in Chancery court and lost again.[1]

Note on Methods of Estimation and Projection of Subnational Population

Methods of preparing population estimates and projections for subnational areas are described in other publications and it is not practicable to review these methods in detail here. The reader is referred to these publications.[2] The principal methods of preparing current population estimates for subnational areas are:

1. Sample surveys
2. Component method based on school enrollments
3. Component method based on income tax records ("administrative records method")
4. Ratio-regression and other multivariate regression methods
5. Housing-unit method
6. Combinations of these and other methods

[1] While underlying all such contests is the matter of competition for the limited demand in a specified service area, the actual legal issue may be quite different and the need for population projections may be secondary. For example, when Suburban Hospital of Bethesda, Maryland, wished to expand its services to introduce a cardiac surgery program in 2004, it had to apply for a "certificate of need" from the Maryland Health Care Commission, an agency of the State of Maryland, and the certificate was awarded. Other hospitals in the Washington, DC Metropolitan Area legally contested the issuance of the certificate. MedStar Health, which owns the Washington Hospital Center in Washington, DC, contested the award of the certificate before the Maryland Court of Appeals. It did so, not on the grounds of destructive competition or lack of need for the additional services, but on the grounds that the Maryland Health Care Commission acted illegally in reworking the health plan a year earlier to allow new centers of heart surgery to open if one of the region's centers fails to meet an annual threshold for the number of surgeries performed.

A year earlier the Commission had changed the criterion from one essentially of need for additional hospital services to one permitting new centers to enter the field if the existing centers failed to meet an annual threshold for surgeries performed. This is a quite different criterion from the "need" criterion. The state regulation permits the commission to authorize a new center after conducting a review of such factors as availability, access, cost, and quality of the planned services. In its decision, the court said that the Commission had acted within its authority to rework the criterion for the state health plan in 2004 from need to underperformance of other hospitals, and hence it ruled in favor of Suburban Hospital's petition.

[2] See especially Siegel (2002), Chaps. 9, 10, and 11; and Siegel and Swanson (2004), Chaps. 20 and 21; and Smith et al. (2001). The U.S. Census Bureau prepares population estimates for a limited number of areas, which it publishes in the *Statistical Abstract of the United States* and on its web site www.census.gov. Several private companies prepare up-to-date local population estimates and projections as part of their product-line.

Component methods combine census data for the local area with separate allowances for each of the components of population change, namely births, deaths, and migration. Because internal migration data are usually lacking, the latter component has to be estimated indirectly, using an indicator variable such as school enrollment or tax returns. In regression analysis, population size or a variant of it for the local area is related mathematically by a regression equation to a series of independent indicator variables for the area, such as housing units, tax returns, school enrollment, and births. In the housing-unit method, the population is assumed to change in accordance with the change in the number of housing units or some variant of it.

The leading methods of preparing population projections for subnational areas include variants of those used to prepare current population estimates for such areas. Here is a summary list:

1. Cohort-component method, which carries the age distribution of the population forward by birth cohorts separately for each component of change;
2. Multivariate linear regression, a non-component method that relates a number of independent variables in a regression equation to population or migration as an independent variable. Once the coefficients of the regression equation are determined from historical relationships and projections of the independent variables have been made, the regression equation can be solved for the dependent variable;
3. Econometric method, a combination of a component method and a statistical method whereby economic factors as independent variables are used to measure future migration in a regression equation, and the projections of migration are then combined with projections of births, deaths, and the latest population estimates to obtain projected population totals;
4. Time-series analysis (i.e., use of Box-Jenkins methods) of population totals or each component of change for an area and projection of the historical series; and
5. Mathematical extrapolation and proration.

Time-series analysis and regression are considered statistical methods, and can be applied as independent methods, or as supporting methods within the framework of demographic analysis.

Stochastic demographic forecasting is a special application of methods 1 and 4 that employs microsimulation and combines analytical and statistical approaches. Because stochastic forecasting derives its results by probabilistic techniques, it can provide confidence intervals for the resulting forecasts.

It is more practical for most users to avail themselves of the population estimates and projections marketed by private vendors, such as CACI, Demographic Data Center, Claritas, Polk, and Equifax, than to prepare the figures themselves. These companies provide estimates and projections not only of the total population of small areas but also for a wide variety of characteristics of the population. Such companies also prepare custom-made estimates and projections, on specification of the areas, demographic detail, and dates desired for the figures.

Indicators of the Local Health Situation

In dealing with local health issues, measures of the local health situation and indicators that aid in explaining the changes in the current health situation have to be prepared. These measures and indicators are then used to monitor progress in improving community health (IOM 1997; Stoto 2004). The measures of the local health situation usually express risks (e.g., diabetes rate), and the questions of interest may be:

How do these risks vary from one subarea to another (e.g., census tract to census tract)?
How do these risks change over time for a particular subarea?
How do these risks vary for the sexes, socioeconomic groups, or the various racial/ethnic groups in the subarea?
How are these risks related to community-level characteristics (e.g., number of health workers, number and location of highways, proportion of crowded housing) for a group of subareas?

We are accustomed to describing the characteristics of individuals in terms of their personal traits but individuals may also be described in terms of family or community characteristics. Individuals live in a neighborhood that has characteristics that may be applied to all its residents. For example, some neighborhoods may be described as stress-inducing because they have high crime rates, have large numbers of deteriorated structures, and are adjacent to highways or industrial plants (Curtis and Leitner 2006). Thus, neighborhoods may be classified according to degrees of positive or negative influence on the mental health of their residents.

Healthy People 2010, the federal program for health promotion and disease prevention, has developed a list of ten health indicators (really a mixture of health measures and explanatory variables) and set health goals on the basis of these indicators for the United States in 2010. The health indicators identified by Healthy People 2010 are:

1. Physical activity
2. Overweight and obesity
3. Tobacco use
4. Substance abuse
5. Responsible sexual behavior
6. Mental health
7. Injury and violence
8. Environmental quality
9. Immunization
10. Access to health care

Local governments can adopt these goals for their own programs and many have done so.

The Metropolitan Washington Public Health Assessment Center, which was established to provide a base for analysis of public health data for the Washington, D.C., area, developed 29 measures to implement the 10 indicators listed above for

9 jurisdictions located in 3 states, namely, the region as a whole, the District of Columbia, 3 Maryland counties, and 4 Virginia counties. The Center then compared the current measures for the Washington area with U.S. current figures and the U.S. targets for 2010. The specific measures are:

1. Physical activity
   - Percent of adults engaged in regular physical activity
   - Heart disease death rate
2. Overweight and obesity
   - Percent of adults obese
   - Percent of adults consuming fruits and vegetables
   - Diabetes-related death rate
3. Tobacco use
   - Percent of adults currently smoking
   - Lung cancer death rate
4. Substance abuse
   - Percent of adults binge drinking
   - Drug-induced death rate
5. Responsible sexual behavior
   - Teen (girls aged 15–17) birth rate
   - AIDS incidence
   - Incidence of chlamydia in girls and women aged 15–24
   - Incidence of gonorrhea
6. Mental health
   - Suicide rate
   - Percent of adults reporting 8+ days of "not good" mental health in past 30 days
7. Injury and violence
   - Motor vehicle crash death rate
   - Firearms-related death rate
   - Reported incidence among females of rape or attempted rape
8. Environmental quality
   - Days the national 1-hour ozone standard was exceeded
   - Reported incidence of Salmonella infection
9. Immunization and infectious diseases
   - Percent of 2-year olds immunized
   - Percent of adults aged 65+ receiving a flu shot in past year
   - Incidence of tuberculosis
10. Access to health care
    - Infant mortality rate
    - Percent of births under 2,500 grams
    - Cervical cancer death rate
    - Percent of adults receiving professional dental hygiene services in past year
    - Percent of adults aged 50+ having fecal occult blood test in past 2 years
    - Percent of women aged 40+ having a mammogram in past 2 years

Source: Website of Metropolitan Washington Public Health Assessment Center: www.mwphac.org. See also Metropolitan Washington Public Health Assessment Center (2001).

Nine of the 29 indicators come from the vital statistics system and another 10 indicators are based on special county-level tabulations of the Center for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System

(BRFSS). All three "states" (Maryland, Virginia, and the District of Columbia) conduct BRFSS surveys. Stoto (2004) noted that data were not available for many of these measures at the national level for tracking the Healthy People leading health indicators.

It is often difficult to compile relevant, accurate, and comparable health data for states and small areas for the selected indicators. National surveys are not well designed to provide local health data and the necessary data are not often collected by local agencies. The sample from a national sample survey is usually too small to provide reliable data of the type needed for states and small areas. The problem is often complicated by the fact that the areas to be delineated may encompass jurisdictions in more than one state (e.g., Washington, DC Metropolitan Area) or that the boundaries of the areas of interest (e.g., eastern half of a county) may not correspond to the boundaries of political jurisdictions for which data are being compiled. Yet subcounty data may be needed to show geographical health patterns since the data for whole counties may not reflect these patterns very well. Finally, the list of diseases that are reportable under the state surveillance systems varies from state to state and each of the states may have different procedures for reporting the diseases.

Estimation of Health Characteristics of Small Areas

Having obtained appropriate population estimates and projections for the areas of interest – total population, age-sex distribution, and possibly race and Hispanic composition and number of households – we want to derive various health measures and data on selected explanatory variables next. Estimates for a health characteristic can be derived directly from a sample survey, by estimation using independent data and demographic and statistical analysis, or by a combination of sample survey data and estimation methods. Specialists in small area estimation conceptualize these types of estimates in the following terms. An estimate for a "domain," such as a small geographic area or sociodemographic group in a small geographic area, is called direct or "design-based" if the estimate for the group is derived solely from the sample units within the domain (allowing for adjustment of the sample estimates for such populations to independent estimates for larger areas). If the sample domain is large enough, the standard design-based procedure may produce reliable domain estimates. Often, however, when a national sample is taken, the direct estimates from the sample for small geographic units within the country lack adequate precision because the domain and the sample for the domain are too small.

To derive more reliable estimates where the sample domain is small, independent estimates may be generated and combined by some weighting procedure or otherwise with the sample survey estimates to derive "model-based" estimates. The independent estimates employed in such model-based estimates may be derived wholly or partly from administrative records or by synthetic methods, regression analysis, or another type of demographic analysis. As an example of the combination of survey estimates and independent estimates, they may both be incorporated

as independent variables in a regression equation to derive alternative improved estimates for the characteristic of interest. Model-based methods are developed in an effort to improve the quality, that is, to reduce the bias and variance, of the estimates over the simple use of the direct survey figures, or to evaluate the quality of the direct survey estimates. Such model-based estimates have often been developed for small geographic areas. Yet they have their shortcomings. If the model is misspecified (i.e., does not employ appropriate variables or express the relation between the variables correctly), the resulting estimates can be seriously in error. In addition, the calculation of sampling errors may be complicated.

Design-Based Estimates: Small Area Estimates from Sample Surveys

The sample survey estimates for the United States, described in Chap. 2, are based on complex multistage stratified probability samples. The "raw" sample estimates have been adjusted ("controlled") to independent estimates of the U.S. population, disaggregated by age, sex, race, and Hispanic ethnicity. The national independent estimates are derived by demographic analysis, which combines data from the prior census and postcensal data on births, deaths, and net immigration. The purpose of the adjustment of the sample survey data is both to achieve consistency with the independent population estimates and to reduce the bias and variance of the raw sample estimates. Note that the estimates resulting from these adjustments are described as design-based or direct estimates even though the adjustment has modified the original raw sample estimates.

The sample size for the regions and the large states in some of the official national surveys is sufficiently large to provide estimates of acceptable accuracy (i.e., with relatively small sampling errors) for some health characteristics in these regions and states. The exact extent to which the bias and variance of the subnational estimates are reduced by the adjustment described above is unknown, but only a small part of the variance, if any, has been eliminated. The adjusted estimates from the survey are still subject to variance and their variances can be measured. The variance estimators – that is, the algorithms for arriving at variance estimates – take account of the sample design and the adjustment to the independent estimates.

Model-Based Estimates: Use of Indirect Data

Synthetic Methods

Perhaps the simplest model-based estimates are so-called synthetic estimates. The derivation of synthetic estimates of a health characteristic begins with design-based estimate of the population having the health characteristic (e.g., diabetes)

for a parent area (e.g., United States, region, or large state). An estimate of the population having the health characteristic (i.e., diabetes) for a subarea (e.g., small state or segment of a state) is obtained by applying proportions of the population having the health characteristic in one or more demographic categories (e.g., age, sex, race, etc.) at the higher geographic level to population figures for these demographic categories for the subarea for which the estimate is sought. In applying the proportions for the parent area to the subarea, the analyst assumes that these proportions are invariant for any demographic category used in the synthetic estimation among the subareas of the parent area.

Statistically, the resulting local estimate may be described as a weighted average of the proportions with a health condition for designated subgroups in the parent area, the weights being the proportional distribution of the demographic subgroups in the subarea.

$$p_a = {}_n\sum [(P_{an} \div P_{tn}) * p_{tn}] \tag{16.1a}$$

$$p_a/p_t = {}_n\sum [(P_{an} \div P_{tn}) * p_{tn}] \div p_t \tag{16.1b}$$

where p_a is the estimate of the population with the health characteristic a in a given subarea, P_{an} is the population with the characteristic a in a given demographic class (e.g., age, sex, race group) in the parent area, P_{tn} is the total population in the given demographic class of the parent area, p_{tn} is the total population in the demographic class in the subarea, p_t is the total population of the subarea, and finally, p_a/p_t is the proportion of the population with the health characteristic a in the subarea.

Typically in this method the proportion of the population with the health characteristic in specified demographic categories for the nation or region are applied to independent estimates of the population with these demographic characteristics for the subarea. For example, we may use the survey estimates from the National Health Interview Survey (NHIS) for a region to estimate the number of persons having the health condition in a constituent state. Here is a simplified hypothetical illustration of the derivation of a synthetic estimate of the number and proportion of persons in the state ever having had asthma. For this computation we use race and Hispanic origin as "symptomatic" variables or weights and regional proportions of persons ever having had asthma for each racial/ethnic group:

Race/Hispanic origin	Proportion asthmatic of race group: Region R (1)	Proportion of total population: State S (2)	Proportion asthmatic of race group: State S (3) = (1)*(2)
White non-Hispanic	.097	.605	.059
Black non-Hispanic	.118	.249	.037
Other races	.106	.061	.007
Hispanic	.079	.085	.007
All races	.089	1.000	.110[a]

[a]The sum of the figures in the column

Multiplying each proportion asthmatic for the race/ethnic groups in region R in column (1) by the race/ethnic distribution of the population in state S in column (2) yields the proportion asthmatic in state S in each race/ethnic group in column (3), and then summing the products yields the synthetic estimate of the proportion of asthmatics in state S. The synthetic estimate of the proportion of asthmatic persons in state S is 11.0%, compared to the observed value of 8.9% for region R.

A second illustration relates to the estimation of the teenage fetal loss rate for a county. Suppose an estimate of the teenage fetal loss rate in County C during year y is desired in connection with preparing an estimate of the teenage pregnancy rate in County C in year y. (Data on births and abortions need to be combined with data on fetal losses to derive the estimate of the total number of pregnancies.) Fetal loss data can be secured from the National Survey of Family Growth (NSFG) for the United States as a whole, but not for states or counties. For this purpose, the national fetal loss rates for teenagers from the NSFG can be "borrowed" as the county's figures. We can improve the estimate of fetal losses for ages 15–19 for the county by taking advantage of the available national data from the NSFG for separate age groups 15–17 and 18–19 for Hispanics, other whites, and other blacks. By applying the national fetal loss rates for these age-race-Hispanic origin groups to the corresponding county population estimates, we derive a synthetic estimate of the teenage fetal loss rate for the county. This rate can be expected to be a more accurate estimate for the county than the national rate. In this way if the county population is mainly black, the rate for blacks will be given major weight, as it should be. Table 16.1 illustrates the calculation of this type of synthetic estimate. The national teenage fetal loss rate for 1995 is 15.9, but the synthetic estimate for the county is 18.0.

Ratio-adjusted synthetic estimates. When sample survey estimates for a region can be obtained, synthetic estimates for the several states composing the region can be adjusted proportionately to be consistent with the regional sample figures. Assuming that the sample estimates for the region are unbiased, this procedure results in some reduction in bias in the state estimates (U.S. Office of Management and Budget/Malec 1993). The estimator is

$$ {}^{s}p_{va} = \left[\left({}^{r}P_{da} \div \sum {}^{s}p_{ea} \right) * {}^{s}p_{ea} \right] \tag{16.2} $$

where ${}^{s}p_{va}$ is the ratio-adjusted synthetic estimate in category a for the state, ${}^{r}P_{da}$ is the direct estimate in the category for the region, and ${}^{s}p_{ea}$ is the synthetic estimate in the category for the state. Such ratio-adjusted synthetic estimates for all states in the region sum to the direct estimate for the region.

Uses and limitations of synthetic estimates. Synthetic estimation has several advantages over other forms of local-area estimation: Simplicity of interpretation, ready availability of the necessary data, ease of calculation, and low cost. It has important limitations, however. The estimates are subject to error both as a result of the limitations of the basic assumptions and the sampling error of the input data. They tend to take insufficient account of factors causing local variations in the

Table 16.1 Calculation of the teenage fetal loss rate for County C by the synthetic method: 1995

Race/Hispanic origin	15–17 years	18–19 years	Total, 15–19 years
Female population, County C			
White, non-Hispanic	735	593	1,328
Black, non-Hispanic	5,647	4,634	10,281
Hispanic	3,041	2,375	5,416
Total	9,423	7,602	17,025
Fetal loss rate, United States[a]			
White, non-Hispanic	11.5	12.9	12.0
Black, non-Hispanic	19.1	18.3	18.8
Hispanic	12.0	26.0	17.5
Total	15.9	16.0	15.9
County C, total	16.2[b]	20.3[b]	18.0[c]

Source: Incorporates data from the U.S. National Center for Health Statistics. "Trends in pregnancies and pregnancy rates by outcomes: Estimates for the United States, 1976–1996," by S. J. Ventura et al., *Vital and Health Statistics* 21(56), 2000

[a]Spontaneous fetal losses from recognized pregnancies of all gestational periods as reported by women in the 1995 National Survey of Family Growth

[b]Derived by weighting the U.S. fetal loss rates shown in this table by the corresponding county populations shown above. Example: $[(.0115\times735)+(.0191\times5,647)+(.0120\times3,041)])\div9,423 = .0162$

[c]Derived by weighting the total rates for ages 15 to 17 and ages 18 to 19 shown for County C by the county populations shown: $[(.0162\times9,423)+(.0203\times7,602)])\div17,075 = .0180$. The same result is obtained by weighting the U.S. age-race-specific rates at once

prevalence or incidence of the characteristic being measured; that is, the relative frequency of the characteristic may vary locally and this local variation is not taken into account.

Moreover, the sampling error of the estimates may be large and difficult to measure. Much of the research on the synthetic method has dealt with the measurement of the sampling error of the estimate and ways of reducing it, although the method is subject to substantial bias as well.

In applying the synthetic method, it is desirable to select variables as weights whose distribution varies widely from area to area (e.g., race/Hispanic ethnicity) and for which the associated prevalence or incidence of the health characteristic of interest varies widely among the categories of the variable selected as weights (e.g., percent with work disability). Age and sex are usually not effective weighting variables for this purpose because populations usually do not vary greatly in age-sex composition, but race/Hispanic ethnicity is a highly variable characteristic, as suggested by the following figures. The national percentage of blacks in the United States was 13 in 2004, but in some states it was near or over one-third. In the nation as a whole 33% of the population was "other than non-Hispanic white," but the figure was 53% for California, 39% for New York, and 50% for Texas. Next, using work-disability as an example for the health variable, its prevalence among blacks greatly exceeds its prevalence among whites. The extent of work disability in 2004 was 13% for blacks, 8% for whites, and 7% for Hispanics.

Synthetic estimation provides a very crude estimate of the absolute number of persons or events in a demographic category (e.g., age-sex group) having the health characteristic (e.g., work disability) in a year in a subarea (e.g., small state or section of state). The sum of the numbers with the health characteristic over all categories (e.g., all ages, sexes, and races) in the subarea may be a useful rough estimate of the total number of persons with the characteristic in the given year in the subarea. The accuracy of the estimates of the overall number and proportion of persons with the characteristic in the subarea depends on the degree to which the group-specific proportions from the parent area apply to the subarea and on the accuracy of the estimates of the population weights for the subarea. If these conditions are favorably met, the synthetic estimates will be relatively unbiased and if, in addition, large enough samples were selected for the detailed health proportions for the parent area, the variances of the subarea estimates will be within acceptable limits. The direct sample survey estimate of a health characteristic for a local area is generally unbiased but has a relatively large variance, while the synthetic estimate is generally biased but has a relatively small variance. A weighted average of the two estimates will yield an estimate with bias and variance intermediate to those of the original estimates.

The National Center for Health Statistics (NCHS) has conducted considerable research on the synthetic method since the mid-1960s in an effort to make estimates of persons with various health conditions for states. The use of the synthetic method to derive local estimates of health characteristics for decennial census years has been very productive since the National Health Interview Survey (NHIS) and the other national health surveys produce data on a great variety of health characteristics for the nation as a whole and the decennial census provides population counts for subnational areas. The health characteristics that NCHS has measured by the synthetic method for states include, among others, hospital discharges, length of hospitalization, percent of persons limited in activity, and number of physician visits per year. With the advent of the American Community Survey, the rolling expanded U.S. sample survey (see Chap. 2), national, state, and local population data are becoming available annually or for groups of years, so that the synthetic method and its various extensions can be implemented in the way described in any recent year.

Regression Methods

There are numerous other ways of preparing model-based estimates than the synthetic method. In the regression method of preparing model-based estimates, a mathematical equation relates one or more independent variables (“covariates”) to data for the health variable under study (e.g., the direct survey estimate) or some simple function of it. The regression method offers a way of employing a wider variety of indicator or symptomatic data than the synthetic method in making estimates of health characteristics for subnational areas. (Note that the synthetic method may be viewed as a special case of the regression method.)

The symptomatic variables for use in the regression equation are selected because they are associated with the health variable being estimated. The independent variables may be a combination of a survey estimate of the health characteristic for the area, a population estimate based on demographic analysis, the components of change underlying the latter estimate (e.g., births, deaths), data from other administrative records (e.g., Medicare records), or a synthetic estimate. The independent variables may also be socioeconomic measures such as per capita income, percentage of the adult population completing high school, and the percentage foreign born, depending on the health variable to be estimated. The form of the dependent variable may be the number with the health characteristic, the percentage of the population with the health characteristic, or the percentage difference between the "true" number with the health characteristic and a synthetic estimate of it. The form of the multivariate regression equation may be linear, polynomial, or logistic.[3] The analyst should consider the alternatives possible to arrive at an optimum procedure for making regression estimates for particular health characteristics for a particular area. Statistical software packages commonly contain procedures for solving various types of regression equations (e.g., SAS, STATA, SPSS, SUDANN).

Regression-adjusted synthetic estimates. A special application of the regression method, called regression-adjusted synthetic estimation, may also be employed to generate estimates of health conditions for states or smaller areas. The regression-adjusted synthetic estimator relates the synthetic estimate in the form of its percentage error as the dependent variable to a set of independent variables in a regression equation (U.S. NCHS/Levy and French 1977; Levy 1979). The equation follows the conventional multivariate linear model, but Y, the dependent variable, has the form of a ratio:

$$Y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \ldots.. + \beta_{m-1} z_{m-1} + \beta_m z_m + e \qquad (16.3a)$$

$$Y = [(X_s - X_t) \div X_t] * 100 \qquad (16.3b)$$

where Y is the percent difference between the synthetic estimate (X_s) and the estimated true value (X_t) for the subarea, z represents the values of the independent variables z_1 to z_m for the subarea, the β's are regression coefficients to be estimated, and *e* represents random error. If the two equations are merged and solved for X_t, the following estimator is obtained:

$$X_t = X_s[1 + 0.01(\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \ldots.. + \beta_{m-1} z_{m-1} + \beta_m z_m)] \qquad (16.3c)$$

[3] Linear regression is not appropriate for estimating a percent as the dependent variable. For estimating a percent, logistic regression is a preferred choice. As may be recalled, it involves the conversion of the independent variable to logit form (the logarithm of the ratio of a percent to its complement). Solution of a polynomial regression may be simplified by taking the first or second differences of the data or converting them to moving averages.

The combination of the synthetic estimate and the regression estimator makes a further adjustment for the bias in the synthetic estimate.

As is evident, model-based estimation of a health characteristic from the NHIS or other health survey for states and its subdivisions at a postcensal date can be carried out in many ways. The regression-adjusted synthetic method is only one of the ways to combine multiple covariates with a synthetic estimate. Another device is simply to average the sample survey estimate and the regression estimate. First, secure the direct sample survey estimate for the area of interest for a current date and, then, develop a second estimate, a regression estimate incorporating several symptomatic variables. Next, average the two estimates with appropriate weights. The selection of weights can be a matter of judgment, but the guiding principal is to select weights inversely to the variance and bias (i.e. total mean square error) of the estimates. Finally, if feasible – and this step requires calculation of preliminary estimates for all states or other coordinate areas – adjust the set of preliminary estimates to the national total from the health survey.

The National Center for Health Statistics has published several model-based estimates of health characteristics for subnational areas using regression estimation. These estimates take advantage both of direct survey estimates of the health characteristics and administrative data symptomatic of change in the health characteristic.

Combinations of Methods

As indicated, the various methods may be combined. A direct survey estimate may serve as one variable in a regression equation. The synthetic estimate may be combined with a regression equation, either as the dependent variable (as in the regression-adjusted synthetic method) or as an independent variable (in a combined synthetic-regression method). Methods may be averaged, as in the composite synthetic method, where a direct sample survey estimate is averaged with the synthetic estimate, with appropriate weights. Direct estimates may be averaged with ratio-adjusted synthetic estimates, weighted together to form another type of composite synthetic estimate (Malec 1993), or a direct sample survey estimate may be averaged with a regression estimate incorporating several independent variables including a synthetic estimate.

The process of modifying the direct survey estimates in these ways is called “shrinkage” by some statisticians. In general, shrinkage refers to the imposition of restrictions on estimates with the goal of improving the results. These restrictions may be of several kinds and may be used alone or in combination. The most common are the weighting of survey estimates with synthetic estimates or with estimates derived from a regression equation, and the adjustment of survey estimates to “control” totals for larger geographic areas by some proportionate adjustment procedure.

Hierarchical/Empirical Bayes Modeling

Without employing the usual statistical name, I have in fact been describing empirical Bayes estimation. Empirical Bayes estimation is a general term encompassing a number of estimation techniques alluded to above. It involves the use of a *prior* estimate (e.g., the direct survey estimate) for an area, which is then used to derive a *posterior* estimate for the area in some way. The prior estimate may be combined (i.e., weighted) with a second estimate derived by the synthetic method, an estimate from a regression equation (in which the prior estimate may be a dependent variable as well as an independent variable), or some other method. In deriving the posterior estimate in empirical Bayes estimation, the prior estimate and the second estimate are combined with weights that are the inv erses of their variances or mean square errors.

$$EY_i = (1 - B_i)Y_i + B_i\hat{Y}_i \tag{16.4}$$

where EY_i = "true" estimate, Y_i = sample survey estimate, $\hat{Y}_i$ = regression estimate, and $B_i = V_i \div (V_i + W)$. V_i = mean square error of sample estimate and W = mean square error around the regression curve (Morris 1980). Note that the weights representing the errors in the two estimates are reversed when applied to the estimates. Posterior estimates may be further adjusted proportionately to the survey estimates for a broader geographic area, such as the region or the nation. This is to assure consistency with the latter estimates but also because the "control" estimates are assumed to have minimal variances and biases. The empirical Bayes estimation method is illustrated below where it is used to estimate the number of insured persons for counties in the United States.

The label "hierarchical" as applied to empirical Bayes estimation refers to the fact that survey data for a higher geographic level are used to produce estimates at a lower geographic level. Direct estimates at the lower geographic level may be unavailable (e.g., zero sample cases) or unreliable as a result of the small number of sample cases in the target areas. The sample size of the survey-based estimates can be increased by borrowing sample data from related areas (i.e., broadening the geographic area) or time periods (i.e., broadening the time period). This technique tends to reduce the variance but to increase the bias in the prior estimate. Ordinarily the larger the sample the closer the empirical Bayes estimate is to the direct estimate. Evaluation studies show that empirical Bayes estimates are a substantial improvement over the corresponding direct estimates (Farrell et al. 1997; MacGibben and Tomberlin 1989).

Health insurance estimates for counties. The following is an illustration of how to estimate the number of persons having health insurance for the counties of the United States by empirical Bayes modeling. The method uses sample survey data from the Current Population Survey (CPS). We begin with the premise that the bias and variance of the sample survey estimates for the counties are unacceptably large. The plan is to derive model-based estimates using a regression equation. First, 3-year averages of the proportions insured in each county from the survey

are calculated and converted to logarithmic form. Next, each logarithm is expressed as the linear combination of the product of several predictors and their regression coefficients, the intercept, and an error term. The proportion insured is the dependent variable and the predictors are the independent variables.[4] The estimator is,

$$\ln P_i = \ln \beta_0 + \ln \beta_1 z_1 + \ln \beta_2 z_2 + \cdots + \ln \beta_{m-1} z_{m-1} + \ln \beta_m z_m + \ln e \quad (16.5)$$

Most of the predictor variables are based on administrative records. Some of the predictor variables, expressed as logarithms, are: Proportion of persons under age 18 and proportion of persons 35–64 who are participants in the Medicaid program; proportion of persons 65 years and over estimated by demographic analysis; proportion of persons who are receiving food stamps; and proportion of persons with ratios of family income to 200-to-300%-of-the-poverty-level as estimated from tax returns.

Once the regression estimates for the approximately 1,200 counties in the CPS with acceptable sample data are obtained, the survey estimates are combined with the regression estimates, implicitly using Bayesian techniques. More exactly, the direct estimates are weighted with the regression estimates according to the precision of each of the estimates (i.e., inversely to their variances). The estimates for the remaining counties are obtained solely from the regression equation.

The proportions of insured persons are converted to absolute numbers insured by use of independent estimates of the household population of the counties derived by demographic analysis or obtained from a data-vending company. The resulting estimates of insured persons for the counties are then adjusted so that the numbers of insured persons over all counties sum to the CPS national survey estimate of insured persons. The state estimates are then derived as the sums of the county estimates in the state.[5]

A similar method can be used to derive county estimates for a great variety of health characteristics from the National Center for Health Statistics' National Health Interview Survey (NHIS). Data are available on more than a dozen health indicators. Data on persons having health insurance and other health characteristics for the nation and the larger states can be obtained directly from the NHIS, but the sample estimates for many states and nearly all counties have unacceptably large variances. Hence, the need to develop model-based estimates for them.

---

[4] All variables in the regression model are expressed in terms of logarithms, except for the region-indicator variables, which have the value of 1 for counties in the region and 0 otherwise. Approximately 1,200 counties with sample cases are used to derive the regression equation although insurance coverage is to be estimated for all 3,140 counties. As expected, CPS sampling variances are not constant over counties. Survey observations with larger variances may be given less weight in the regression than observations with smaller variances by weighting each observation by the inverse of its variance.

[5] The final estimates for counties where there are no sample cases are approximately the same as the regression estimates. The estimates for counties with large samples and high percentages insured (i.e., with low variances) tend to be close to the direct estimates and the estimates for counties with small samples and high variances tend to be close to the regression estimates.

The American Community Survey (ACS) provides current data on total population and population characteristics useful in selecting possible independent variables in a regression equation to estimate health characteristics for counties in the CPS and NHIS. In addition, disability data following the several definitions of disability described in Chap. 5 are available directly from the ACS for 2006 and later years for states, counties, and other geographic areas with populations of 65,000 or more.

Graphic Devices for Health Analysis

Data Portrayal and Analysis

As noted, many issues involving health data relate to small areas, such as census tracts, urban health districts, and so on. For the description and analysis of small area data, maps are particularly useful. They permit one to visualize variations in the incidence or prevalence of some health condition from one area to another at a glance. Moreover, they allow for the addition of variables that may explain geographic variations in the health condition. Maps confer the ability to extract meaning from spatial data and so are analytic tools as well as tools aiding visualization of the data. Maps can show to what extent a health condition is spatially autocorrelated (i.e., to what extent its distribution in space is not random) and to what extent the incidence of the health condition decays with distance from an epicenter. Maps can be used effectively to show the changes that take place in health phenomena over time. For example, a sequence of maps could show how AIDS has been spreading over time in a country.

Maps have been employed for a considerable time as a tool in the analysis of health data as well as of social and economic data. In fact, some of the earliest uses of maps involved the display of the distribution of cases of some disease over the parts of a city. Best known of these is a map prepared by the physician, John Snow, showing the distribution of cases of cholera in London in the middle of the nineteenth century (Fig. 16.1). This map enabled him to show that most of the cases of cholera in the city of London occurred among those who used one particular source of water, which was subsequently found to be contaminated from sewage. Once the handle to the water pump connected to the contaminated source was removed and people no longer had access to that supply, the incidence of cholera dropped rapidly.

As with the statistical data on health indicators, the utility of maps displaying such data can be greatly enhanced by relating them to the population exposed to risk. As we have seen, an increase in deaths from some cause may simply reflect the growth of population or a shift in age structure, but the calculation of age-specific or age-adjusted death rates can eliminate the independent effects of population size and age composition. The analyst may also wish to relate the health data to other series, such as land area or educational level.

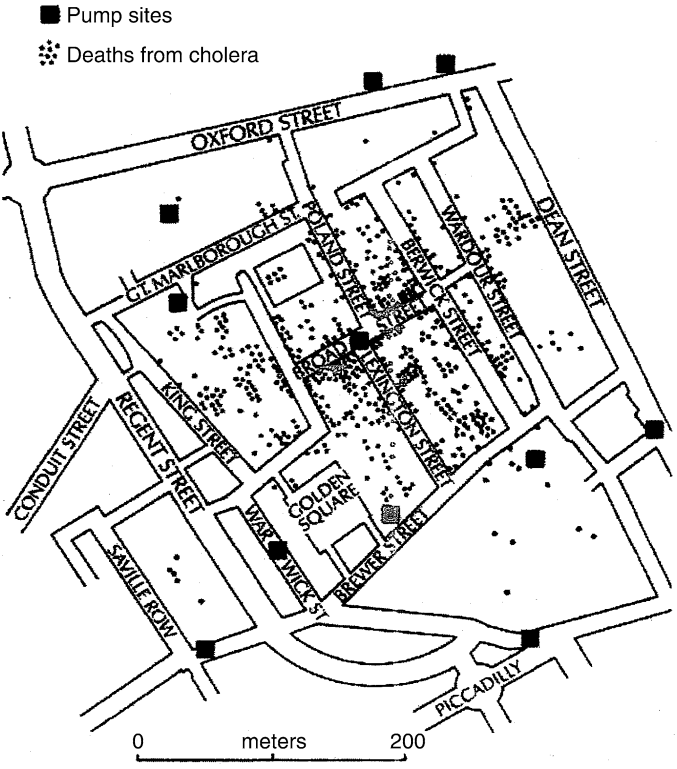


Fig. 16.1 Dr. John Snow's dot map depicting the distribution of cholera cases in London, England: 1849 (Source: Public domain; www/nationalgeographic.com/resources/ngo/education/ideas912/912choleraho3.html)

Types of Thematic Maps and Their Uses

Maps presenting subject material are called thematic maps. There are several types of thematic maps. Among these are choropleth maps, in which geographic units are differentiated by colors, shades, or patterns to show interarea variations with respect to some variable (s). Other types of thematic maps are dot density maps, proportional symbol maps, flow maps, and contour maps. A single map may combine more than one of these mapping techniques. In general, in the design of thematic maps, increasing levels of a variable are represented by increasingly darker (or lighter) shades, increasing density of dots, increasingly larger circles or spheres, greater thickness of flow lines, or increasing concentration of contour lines.

In deciding on the size of class intervals, that is, how many classes are desirable to assure that the map presents meaningful and informative differences in the data, the analyst has to choose from a number of possibilities, including equal intervals, natural clusters, quantiles of some sort, or some different alternative. Often the areas

with the largest and smallest levels or changes in the characteristic depicted fall into their own class intervals. A careful decision in selecting class intervals for the map is required to make it truly informative. To assure the comparability of a series of maps, whether depicting different areas or the same area at different dates, the same scale, format, symbols, and classification should be used from map to map, to the extent possible.

Choropleth Maps

Choropleth maps typically display the variation in some variable by using four to six shading schemes to distinguish the geographic units in the map. The shading scheme may be the density of cross-hatching, color tones from white to black, tones of a single color, or several colors. Figure 16.2 is a type of choropleth map with six class intervals of a single variable distinguished by blends of "color" from white to black. (Figure 16.2 is shown also in color in U.S. Census Bureau, *Census Atlas of the United States*, Census 2000 Special Reports: CENSR-29: 207 and 255). Typically only one variable is displayed on a map, but two variables can be combined on the same map. One way to do this is to use different schemes for the two variables. A shading scheme for one variable may be combined with a set of symbols (e.g., circles) for the second variable. Figure 16.3 employs this design. Alternatively, color coding of the cross-classified data may be employed. For example, a map may show the death rate from cardiovascular diseases in relation to educational attainment for counties or the crossclassification of median family income and percent of high school graduates for counties. (The latter relation–that between income and education for the counties of the United States in 2000–is shown in U.S. Census Bureau, *Census Atlas of the United States*, page 217). The Census Bureau map has two intervals for the percent of high school graduates, each represented by a shade of red and green and two intervals for median family income, each represented by a different shade of green and red. Each country has been assigned a mixture of a shade of red and green (totaling four colors), varying according to the level of median family income and the percent of high school graduates for the country. (For further instructions on designing such a map, see Barabba 1975.)

Dot Density Maps and Proportional Symbol Maps

Dot density maps display a dot for a designated fixed number of persons, say a dot for each 1,000 persons. The dots are assigned to the coordinates of geographic or political units and distributed approximately according to the actual location of the population being described. Figure 16.1 is a type of dot density map.

Proportional symbol maps use a symbol in each geographic subarea to show the magnitude of the variable (Fig. 16.3). Usually a circle is employed; then, the radius of the circle is made proportional to the square root of the population. If the population of one subarea is twice that of another, the radius of the circle representing the former area increases by $\sqrt{2}$, or 41%, over that of the latter subarea.

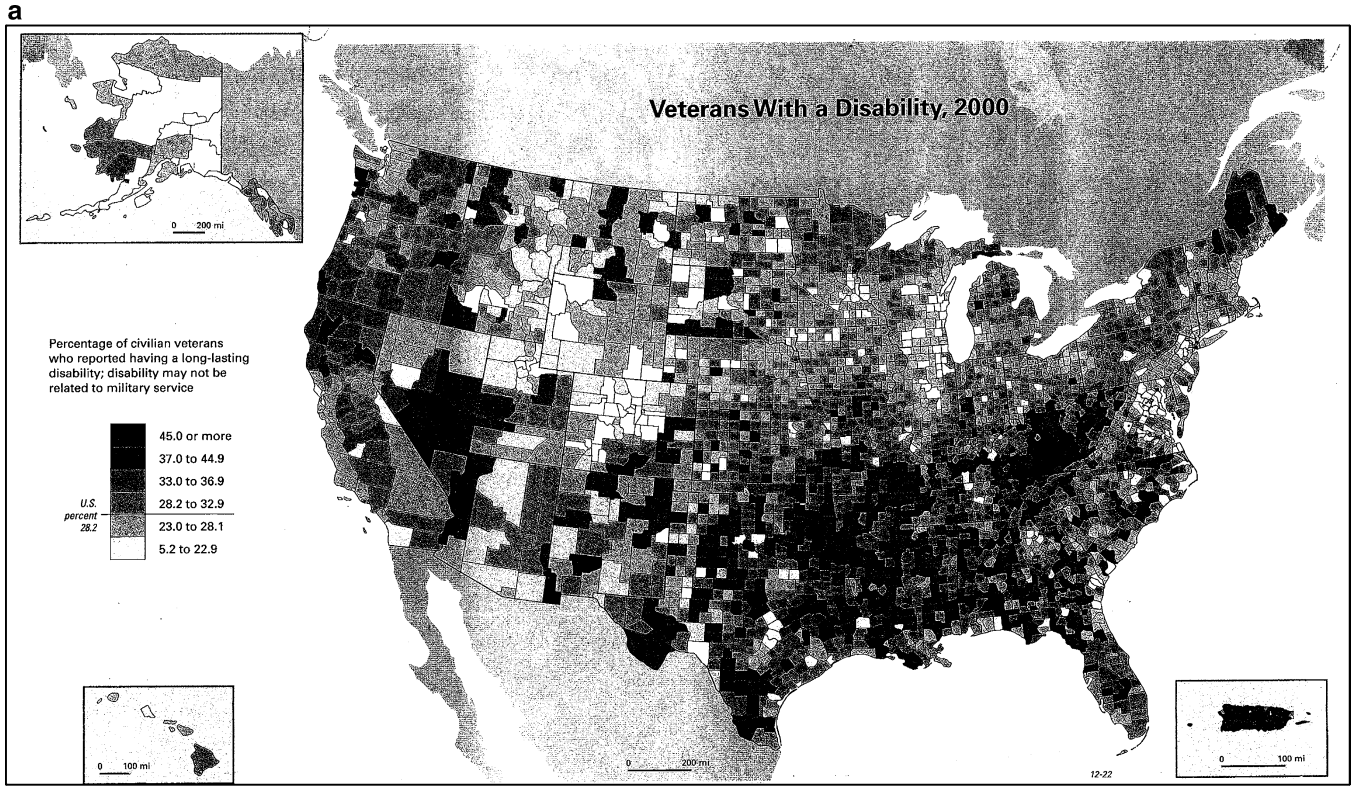


Fig. 16.2 **(a)** Choropleth map showing disability rates for veterans, for counties of the United States: 2000. **(b)** Choropleth map showing crowded housing, for states of the United States: 1940, 1970, and 2000 Source: U.S. Census Bureau, *Census Atlas of the United States*, Census 2000 Special Reports: CENSR-29: 207 and 255

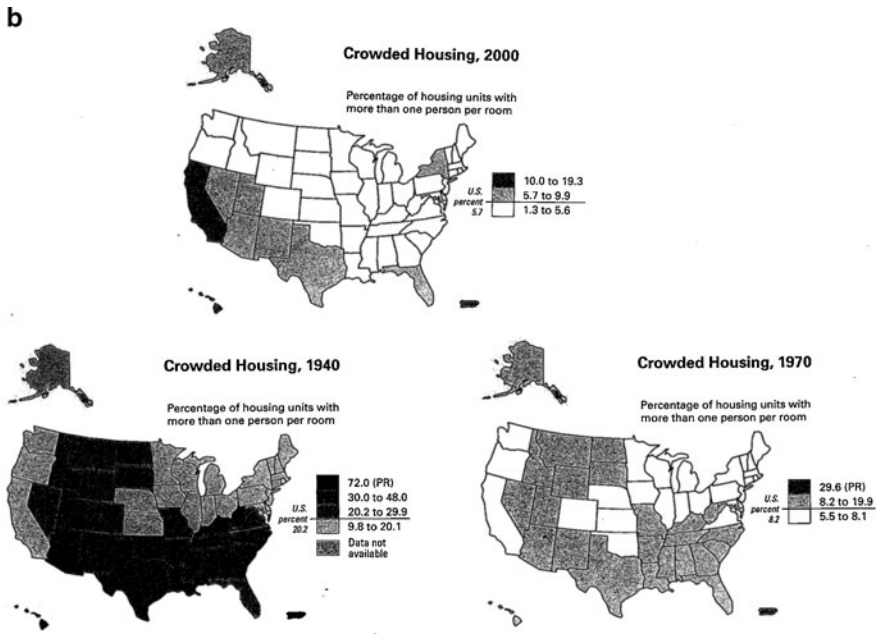


Fig. 16.2 (continued)

Sometimes, a sphere is employed; in this case the radius of the sphere is made proportional to the cube root of the population. If the population doubles, the radius of the sphere increases by $\sqrt[3]{2}$, or 26%.

Isoplethic/Contour Maps and Flow Maps

In an isoplethic/contour map, lines connecting equal values are drawn between equal points of data. Darker shades of a color may be used to represent higher values of the variable. One type of contour map incorporates distance and direction simultaneously in measuring the geographic distribution of characteristics from a designated point. Lines on isoplethic/contour maps are independent of conventional political and statistical boundaries and connect lines of "common intensity" of the variable.

In flow maps a straight or curved line with an arrowhead indicates the direction, including the origin and destination, of population movements or the spread of a disease. The width of the lines shows the size of the movement and the lines' colors or patterns can indicate qualitative differences or another quantitative variable (Fig. 16.4). This type of map can help users visualize actual or potential routes of the spread of a disease or noxious agent or the movement of migrants. For example, the lines could indicate the direction and volume of the spread of HIV/AIDS through a country.

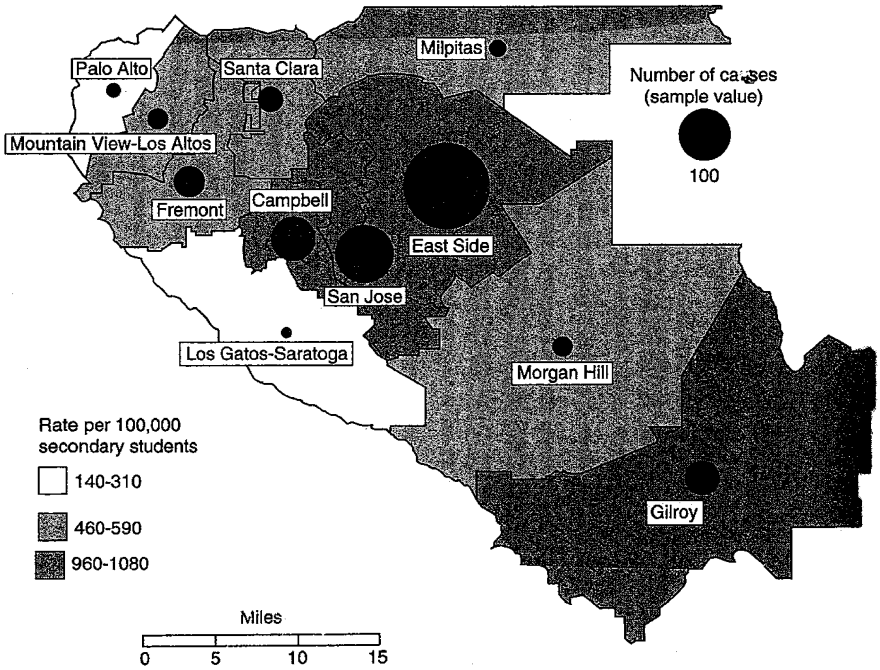


Fig. 16.3 Map showing number and rates of sexually transmitted diseases among adolescents in the secondary school districts of Santa Clara County, California: 1993–1994 (Source: With kind permission from Springer Science + Business Media: Gobalet and Thomas (1996). Copyright © 1996 Springer Science + Business Media)

Geographic Information Systems[6]

With the advent of computers, the automatic production of maps with customized specifications became possible. I consider this new development, called geographic information systems, in some detail next. Prior to the advent of the computer, maps with data for geographic areas were usually prepared by hand, that is, draftspersons made the appropriate entries on map outlines according to instructions from the statistical analyst. Today such maps are prepared almost entirely by computer.

Characteristics

Geographic information systems (GIS) represent the marriage of mapping procedures, statistical data, and computer technology. GIS is a computer-automated

[6] This section is drawn in part from: Bryan and George (2004); and Siegel (2002), Chap. 5.

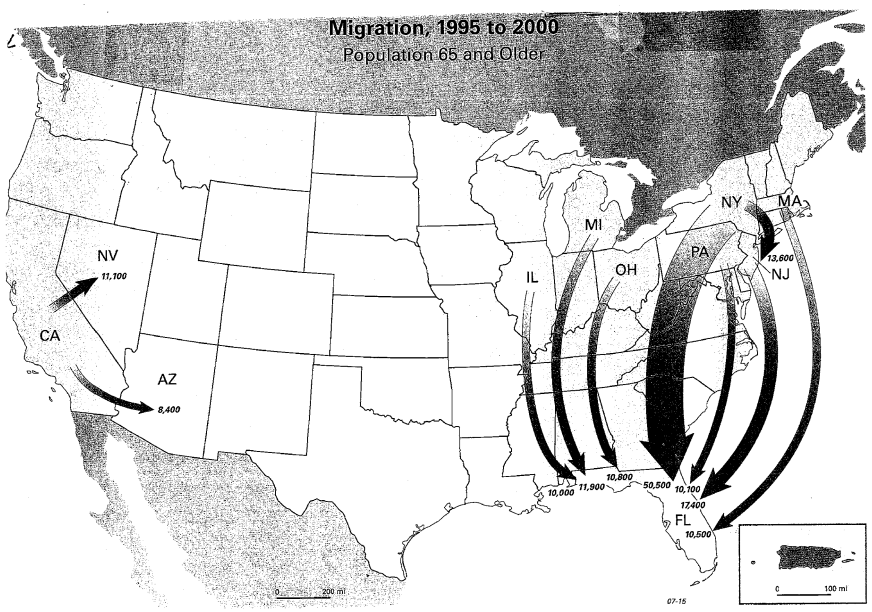


Fig. 16.4 Flow chart depicting the 10 leading net migration flows of persons 65 years and over for states in the United States,1995–2000 (Note: *Arrows* represent direction from origin to destination; *width of lines* represent relative volume of net flow between the states;Source: U.S. Census Bureau, *Census Atlas of the United States.* Census 2000 Special Reports: CENSR-29. Page 115)

system of compiling and mapping statistical data identified by geographic coordinates, for specified sets of geographic areas identified by geographic coordinates. A geographic information system requires an automated geographic database to support it. To accomplish this, the cartographic coordinate points must be digitized, i.e., coded in binary language that a computer can read. The resulting product is a geobased file system for manipulating data. The U.S. Census Bureau, in collaboration with the U.S. Geological Survey, developed such a database, identified by the acronym TIGER (Topographically Integrated Geographic Encoding and Referencing) for use in taking the 1990 census. The system was largely completed between 1984 and 1987.[7]

The TIGER/Line Files are a digital cartographic database that contains all the information normally found on a Census Bureau map in a form that a computer can manipulate. They are used to assign geographic codes to addresses for collecting

[7] Precision in mapping addresses is provided by the Global Positioning System (GPS). This technology, developed by the U.S. Department of Defense, involves 24 satellites that orbit the earth at an elevation of 22,000 miles and continuously beam their locations and temporary positions toward the earth's surface. With a GPS receiver, it is possible to record the precise coordinates of any point on earth. The coordinates can then be entered into a geographic information system (GIS).

data, and to provide the geographic structure – that is, the relation of one geographic area to other geographic areas – that permits assigning an address to the correct census block, census tract, or other geographic unit. The TIGER system makes it possible for users to generate maps by computer at different scales for any geographic area in the country.

To apply the TIGER files, user applications software is required. The applications software is available only commercially. Environmental Systems Research Institute (ESRI) markets ArcView GIS, one of the most widely used desktop GIS and mapping software products.[8] With these tools, computer users can generate complex maps on their personal computers, including thematic maps in color for very small geographic areas. Desktop analysis systems have the ability to integrate data sets from a variety of sources, both internal and external, the capability of importing data sets in a variety of formats, and the ability to incorporate these data sets into existing applications.

Supported by the appropriate software, TIGER/Line files enable the user to combine geographic and cartographic data with data from censuses, vital registrations, and administrative records in order to present the statistical data graphically and analyze them spatially. The statistical data in the summary tape files from the 2000 census, for example, can be linked with the TIGER/Line files, which contain the census-2000 geographic codes for these data. GIS with TIGER is useful whenever demographic, socioeconomic, and health phenomena present, or are expected to present, distinct patterns that vary according to different geographic or administrative units. It provides a basis for visualizing and analyzing information whenever variables such as location, distance, area, and geographic shape are involved.

GIS on Personal Computers

Several factors had to converge to make possible preparation of statistical maps displaying census and other data by personal computer. One is the availability of these data in digital form on disk, CD-ROM, or the internet at the Census Bureau's and other agencies' web sites. Another was the development of personal computers that were powerful enough and had a sufficient memory to be able to store and manipulate large files. Finally, there was the development of computer mapping, by which each data item could be coded and matched electronically to a point or area on a map. The geo-referencing of data to places on a GIS-generated map allows for the combination of many different types of data for the same place. Multiple geographic data can be stored in a common coordinate system, these map layers can be viewed simultaneously, and a composite map can be created from the several

[8] Other software packages for implementing GIS, in addition to ArcView, are produced by MapInfo, Atlas GIS, Arc Info, GeoMedia, Infomark, Maptitude, Surfer for Windows, Geo Da, and Sat Scan.

overlays.[9] The new developments added to the user's ability to visualize the changes in health and related variables that are taking place over time and the differences in these variables that have developed over geographic space. Thus, a composite map with information on population exposed to risk can be combined with information on health conditions to investigate their possible association.

Applications

There are numerous applications of GIS/TIGER to health matters. GIS allows the analyst to display health phenomena such as the incidence of disease and the location of treatment facilities geographically. Maps can be prepared for the purpose of analyzing geographic patterns and correlates of various health conditions. Addresses of patients with specific diseases can be combined with population census data to determine the incidence rates of the diseases for specific geographic zones in a city. The geographic display of the combination of health data with associated demographic and socioeconomic data aids in the analysis of their relation and in planning preventive and treatment programs. Hospital, patient, and physician data can be linked. GIS can aid in identifying the characteristics and locations of the principal victims of a particular disease, whether salmonella poisoning, Legionnaire's disease, or other toxic exposure. By mapping demographic and socioeconomic factors such as age, income of householder, and housing type in relation to the distribution of the disease, target populations for the disease can be determined. Thus, one can compare the health situation of elderly persons in different neighborhoods and measure the differences in the incidence and prevalence of a particular disease among census tracts, racial or ethnic groups, or income classes.

Other applications of GIS/TIGER include locating the address of a 911 emergency call, monitoring the ongoing occupancy status of the beds on a hospital floor (Lang 2000), exploring consumer markets for hospital or other health services. and analyzing patients' access to care. Demographic profiles of a target population such as the elderly disabled population can be generated for designated geographic reference areas such as counties, cities, and census tracts. Still other possibilities

[9] The data layer is a basic organizing principle of GIS. Groups of similar features are combined in one of these data layers. A GIS database may include a layer of physical features such as roads, rivers, and structures, and another of geologic and topographic data. GIS is designed to integrate these different data sets into a common reference framework – the geographic coordinate system. There are two common methods of representing spatial information in a GIS database. They are the vector data model and the raster data model. The vector model is generally used to symbolize discrete features of the environment such as houses and roads, and the raster model is used to symbolize continuously varying phenomena such as climate or elevation. The vector system follows a geometric method of designation using points, lines, and polygons, while the raster system divides space into a regular array of rows and columns and the value of data at a particular location is stored in the corresponding cell of the raster.

include merging files on health conditions and social characteristics, drawing new boundaries for health districts, planning new community health services, and determining how far patients have to travel to their physicians' offices. GIS can be used as adjuncts to health care facilities surveys. Such surveys seek to determine how many of a specific type of facility currently exist in an area and assist in ascertaining the best locations for new ones.

Illustrative studies of spatial distribution of diseases. Gobalet and Thomas (1996) describe several cases in which GIS was used to display and analyze the spatial distribution of public health data. One involves sexually transmitted diseases (STDs) among adolescents in the high schools of Santa Clara County, California. A map depicts the number of cases of STDs by proportionally sized circles, and the rates of STDs by various patterns of shading, for each secondary school district in the county in 1993–1994 (Fig. 16.4). Patient-address data were collected from physicians and health facilities by the public health department of the county, geocoded, and aggregated for secondary school districts. The geocoding of the STD cases and their efficient aggregation for school districts were made possible by GIS software.

Lead overexposure among children has been a serious public health issue in the parts of cities with older housing. Wartenberg (1992) analyzed screening programs for lead overexposure by identifying areas with common risk markers such as older housing and poverty. GIS can make screening programs more effective and less costly by mapping cases of overexposure and identifying neighborhoods with high incidence rates and at high risk. The first step is to enter into a computer database the addresses of homes in a city (or other geographic area) with lead pipe and lead paint. The next step is to link these addresses to a map of streets of the city stored in the GIS. Then a map can be displayed showing the locations in the city of the homes with lead pipe and lead paint. Maps can also be prepared showing the distribution of homes according to average age, householders according to educational level, and householders according to median household income, to distinguish neighborhoods that were merely poor and neighborhoods with lead-pipe homes.

Given the expectation that the learning ability of pre-teenage children and their growth could be adversely affected by drinking water containing lead, it is of interest to determine whether there was a relationship between a lead-pipe water supply and student test scores. The mapping of elementary schools, with records of fifth-grade standardized achievement test scores, in relation to the neighborhoods with lead-pipe water service, could reveal a possible relationship between these factors. The aim is to ascertain whether the students in schools located in areas with lead-pipe water service, and thus at higher risk, have poorer test scores than schools located in other water service areas.

Such a study could be informative, but it could also be misleading because of the problem of ecological correlation. The grouping of data into neighborhoods for the analysis could mask the true extent of the relationship between achievement scores and type of water service for individuals. Less biased interpretations could be made by matching individual test scores with individual homes. GIS technology

can be applied to individual or microdata as well as to area data. GIS can match individuals residing at specific addresses or working in specific geographic areas to the GIS geographic database. In this way the concern about misinterpretation of the relation between the series of data because of ecological correlation is eliminated and more valid inferences can be drawn from them. Even with individual data there is a risk of biased interpretations if possible confounding factors are omitted from the analysis. Note that the research design with microdata is more difficult and costly to implement and carry out.

A study of Glass et al. (1995) carried out an analysis at the individual level, using individual level data for disease outcomes. The study investigated residential and environmental risk factors for incurring Lyme disease, a disease caused by a bacterial agent transmitted to humans by ticks. The researchers used GIS to overlay six different land databases that included land use/land cover, forest distribution, soils, elevation, geology, and watersheds and that incorporated 53 environmental variables. Case control methods and GIS were combined to identify residents of Baltimore County, MD at high risk of incurring Lyme disease. The GIS database included the addresses of 48 cases of Lyme disease that occurred between 1989 and 1990 and 495 randomly selected control addresses. This information was combined with 11 selected environmental variables in a logistic regression to determine the importance of various risk factors for Lyme disease. This study depended on knowledge of the epidemiology of Lyme disease to identify potential environmental factors involved in its spread, the methods of GIS, and existing computerized databases to develop the models of population exposed to risk.

Temporal spread of diseases. During the 1980s, when the HIV virus and AIDS were spreading rapidly, GIS was used to track the geographic spread of both the virus and the AIDS epidemic. With this knowledge, targeting measures could be developed and treatment and prevention programs could be implemented most efficiently. GIS can contribute to developing estimates of how a smallpox epidemic would spread in the event of a terrorist attack. It can contribute to estimating the number of persons who would be afflicted by such infectious diseases as cholera and dysentery following an environmental disaster and the number who may later die from these diseases as a result of the disaster.

Kitron et al. (1994) demonstrated the use of GIS in malaria surveillance in Israel. In Israel there is a risk of localized outbreaks of malaria as a result of infection of local mosquitos by people who come into the country already carrying the disease. A national computerized surveillance system of breeding sites and imported malaria cases was established in 1992 using GIS. Distances between population centers and breeding sites were calculated, and maps linking epidemiological and other data were generated. If a localized outbreak does occur, the surveillance system can quickly associate it with a likely breeding site, a specific vector (i.e., the flight pattern of each species), and a probable human source, so that control measures can be instituted promptly and effectively.

International mapping of diseases. GIS is being used increasingly in the less developed countries (LDC) by both native and foreign analysts for the study of LDC health issues. GIS has been brought to bear in the international mapping of a variety of diseases, including dengue fever and malaria. The GIS surveillance system is especially useful for countries with indigenous malaria transmission. It has been used also in studying the geographic variations in immunizations and in humanitarian emergencies, including risk assessments, conduct of surveys, and investigations of disease distributions (Kaiser et al. 2003). Major difficulties arise in using GIS in the LDC because of the costs of equipment and software, the problems of training personnel, and the difficulties of maintaining the equipment and the skill of the analysts (National Research Council 1996). For these reasons there is still a considerable gap between the MDC and the LDC in the application of GIS to health problems.[10]

Matching the supply of health services to demand is especially important for countries with limited resources and limited access to health facilities. In the LDC there is often insufficient information on how many persons and what share of the population are being served, and what services the populace seeks and requires. The number of persons served by existing hospitals, the number served by other means, and the number not receiving services need to be determined. Such information may be known for political districts, but districts have different boundaries than health "catchment areas," so that special estimates for catchment areas have to be made. Estimates of the number of people living in each catchment area are also needed to determine the share of the population of each area that is served by the health facilities. GIS can assign each household to a health catchment area, as well as determine distance and travel time to a health facility.

Problems with Geo-Referenced Computerized Databases

Existing computerized databases are subject to several limitations. They may not be sufficiently precise, complete, and current to serve a particular purpose. Vine et al. (1997) list some conditions for computerized databases to be useful in spatial epidemiological research: (1) they must contain spatial coordinates and (2) they must contain temporal and quantitative information regarding the measured factor (e.g., dates and level of exposure). Databases often lack appropriate spatial detail to be useful for the analysis planned. Vine et al. note other problems with the various data sources contributing to each map layer. They may not be comparable with respect to the geographic unit to which the data apply, the map scale for which

[10] The United Nations Statistical Office and UN member agencies have recognized the importance of GIS in the portrayal and analysis of spatially distributed data. Their publications, international conferences, and demonstration projects assist the LDC in setting up GIS as a tool for analyzing spatially distributed statistics for health surveillance and humanitarian assistance. The publications provide examples of the use of GIS techniques in demographic studies in various LDC. See, for example, United Nations, ESCAP (2004).

the data were collected, the time frame to which the data apply, the accuracy and completeness of the data, and the format in which the data were computerized. The combination of non-comparable data layers could result in erroneous information on exposure to risk within the various geographic units.

Address geocoding may be necessary when the geocodes contained in the datasets are not specific enough or the data have not previously been geocoded. The U.S. Census Bureau's TIGER/Line files are most commonly used for address geocoding and they nominally cover all areas in the United States. The files contain geographic coordinates for all census geographic units down to street segments (i.e., parts of a street if intersected by other streets). The street segments are coded by the street name and the address range (e.g., 1500–1537). Specific addresses are geocoded with the street name and the address range, and then by interpolation. However, the geocoding for rural areas is often incomplete (Vine et al. 1997). For rural areas more extensive geocoding may be necessary using available lists of addresses (e.g., water billing lists, telephone books) and maps. Post office box numbers and rural route numbers are not useful for geocoding.

With GIS, analysts can display the exact address of the cases studied. Hence, there is an issue of data confidentiality. The analyst must protect the confidentiality of the individual even while preserving the geographic precision of the data for study purposes. The Census Bureau has done considerable research on this issue and employs a device whereby data liable to violate respondent confidentiality are switched between households with similar characteristics. Some researchers using GIS have developed their own solutions, such as displaying only grouped data or reassigning points geographically on a random basis to distort the actual location of individual cases.

Projections of the Health Situation and Resources

Projections of the Health Situation

Methods of projecting mortality and health conditions at the national level were discussed in Chap. 14. Here a note relating to the application of these methods to subnational areas is added. The several methods described earlier, with modifications, can be applied, alone or in combination, to the preparation of subnational projections of the health characteristics of the population. A list of the methods is as follows:

General analytic methods:

Prevalence-ratio method
Event-exposure method
Cohort-component method
Stochastic demographic forecasting

Statistical methods

Time-series methods/Box-Jenkins methods
Regression methods/econometric methods

The most common and most simple analytical methods of making such projections are the prevalence-ratio method and the event-exposure method. These methods work with aggregate data and produce aggregate results. In these methods the frequency of a health characteristic or event in a local area during a year or other period is divided by the population at risk, the resulting ratio or rate is projected to a future date, and the projected figure is applied to the population projection for that future date. The cohort-component method is another macrosimulation method but it is more complex and requires more detailed data for its implementation, such as recent data on change in the health characteristic or event for birth cohorts at the local level. Another method, stochastic demographic forecasting, is a microsimulation method and so it deals with microdata. We leave it to the reader to follow up on the details of these methods of making health and mortality projections inasmuch as this work is a specialized field of applied demography. Some references for getting started were given earlier.

Projections of Health Resources

For purposes of public planning, population projections and projections of health conditions must ultimately be converted into measures of the future need for health facilities (e.g., hospitals), health personnel (e.g., physicians), or health services (e.g., medical imaging). Nations, states, and smaller jurisdictions have to anticipate their needs for workers in various health fields, to assure that a sufficient number will be trained and available for service. In planning the requirements for health personnel, projections for a variety of specialties, i.e., physicians, dentists, nurses, and auxiliary health personnel, are needed at least for the nation and broader geographic areas. Governments also have to anticipate their needs for health facilities to assure that a sufficient number of hospitals, rehabilitation centers, hospices, and nursing homes will be constructed, adequate bed capacity will be installed, and other required health facilities will be available. Such projections are needed not only for a nation's major political divisions but for local jurisdictions such a metropolitan areas. This section offers some general guidelines for preparing projections of health personnel and facilities.[11]

[11] Some projections of health resources are prepared by U.S. government agencies. In addition to the basic projections of national population prepared by the Census Bureau and the Social Security Administration, and the projections of national labor force prepared by the Bureau of Labor Statistics and the Social Security Administration, national projections and occasionally subnational projections for several types of health resources are prepared by two Federal agencies. The Bureau of Health Professions prepares projections of the number of workers required in certain

Projections of applied health projections may be derived directly from population projections, or they may be calculated on the basis of intermediate projections of health conditions. Projections of population are critical when forecasts of the need for health facilities, health personal, or health services, are derived by a method involving per capita need or demand. The calculations may be done in one or two stages. When projections of health conditions are derived first, two ratios are calculated and applied in tandem. The first ratio is designed to project the share of the population that will have a health condition or conditions and the second is designed to project the need for the health resource in relation to the numbers with a health condition. The most common analytic tools for making such projections are the prevalence-ratio method and the event-exposure method. These are simple devices that require little data.

The following is an illustration of the general design for making projections of visits to neurologists, and ultimately of the need for neurologists, by the prevalence-ratio method in two stages. The first ratio derives a projection of the numbers of disabled persons, and the second ratio converts the number of disabled persons into the number of visits to neurologists.

$$\text{Numbers disabled} = \frac{\langle \text{Numbers disabled} \rangle}{\langle \text{Population} \rangle} * \text{Projected population} \quad (16.6)$$

$$\text{Numbers of neurologists' visit} = \frac{\langle \text{Numbers of 'neuros' visits} \rangle}{\langle \text{Numbers disabled} \rangle} * \text{Numbers disabled} \quad (16.7)$$

All calculations should be carried out for age-sex groups if possible. Projections of hospitals and hospital beds can be tied in with prior projections of health conditions and average bed use per condition. Note that the design of these projections calls for prior projections of population disaggregated by age and sex, and then the use of the resulting projections of the disabled population to secure projections of neurologists' visits.

The Equilibrium Issue

The projections we have described so far are projections of *demand* for health personnel, facilities, or services. They are needed to determine probable requirements of health personal or facilities. Alternative projections can be made of the *supply* of health personnel or facilities. They are needed to determine the probable

---

health professions, specifically physicians, dentists, and nurses, and the U.S. Health Resources and Services Administration prepares projections of hospital construction. The U.S. government is directly involved in the planning for hospital construction since it subsidizes new construction with state grants.

supply of personnel or facilities. Both demand and supply projections call for the application of demographic methods to demographic data, but they typically employ different data and techniques. Demand projections are usually based on population projections and projections of per capita demand for health personnel or services. Supply projections of health personnel, for example, are based on the current supply of health providers, the numbers of newly trained health providers graduating training centers, and projections of demographic changes (i.e., deaths, migration, and retirements) among them. We can expect the two series of projections to disagree. It is recommended that the analyst prepare both types of projections and then try to reconcile them in order to arrive at projections of the "actual" numbers for use in public planning. In the reconciliation of this "equilibrium model," special steps may have to be taken to increase supply to meet the greater demand, or to reduce the supply to meet the smaller demand.

For example, suppose we want to make projections of the number of dentists that will be available for practice in a specific metropolitan area in future years. It may be useful first to prepare projections of the number of dentists in the United States. For supply projections the general method would involve adjusting the current count of practicing dentists to take account of future mortality, immigration, emigration, retirement, and the addition of new recruits from dental schools who intend to set up dental practices in the United States. Next, a share of the national projected total can be taken for the specific metropolitan area of interest on the basis of past trends in this percentage. Alternatively, one can make direct projections for the metropolitan area by adjusting the current count of dentists practicing in the area for their future mortality, retirements, and additions through graduation from local dental schools. Then it is necessary to allow for movement of dentists into and out of the local area for the projection period.

Any discrepancy between the projected demand and the projected supply of dentists for the area suggests the prospect of either a shortage or surplus. The more likely scenario is that there will be an excess of demand for dentists over the supply of dentists. To avoid this problematic eventuality, special action must be taken to try to increase their supply.

Shortages of Health Providers

In the United States nationally there is a shortage of nurses and the shortage has been increasing, as indicated by a comparison of both demand and supply estimates. Thirty states report a shortage of nurses and this count is expected to grow to 40 states by 2020 (*AARP Bulletin*, Nov/Dec. 2003: 54). In these states the demand for nurses exceeds the supply both currently and prospectively. Hence the supply could be determinative as to the actual number employed unless aggressive efforts are made to increase the supply. This labor situation underlies the special dispensation for foreign-born nurses in our immigration legislation.

There are current and prospective shortages of other types of health providers in the United States. Especially acute is the shortage of caregivers for frail and disabled

persons. The situation is expected to become acute with the advent of the baby boom cohorts to the older ages. Many of their parents now require the attention of caregivers and ultimately most of them will also. By 2030, when all the baby-boom cohorts will have passed age 65 and many will have passed age 75, the United States will need between 5.6 and 6.6 million caregivers (R.N. Butler, *AARP Bulletin*, Jan. 2007:61). Because of the shortage of paid caregivers, much of the burden, where it is filled, is provided by unpaid family caregivers. This arrangement may sometimes have a positive consequence for the patient but can have negative consequences for the family, the family caregiver, and the economy.

Planning for Public Health Emergency Services

Measuring Vulnerable Populations and Losses in a Disaster

Government agencies can use GIS with TIGER to measure and describe the population vulnerable to, and affected by, floods and other natural disasters, and plan public action to respond to the emergency. Public transportation planners can plan emergency-vehicle routing, partly on the basis of data on residential patterns and data on housing characteristics and partly on the basis of the geographic distribution of past emergencies and the location of the stations used for the emergency vehicles. Applied health demographers may be called on to estimate population losses in a disaster such as a flood, a hurricane, or a heat wave, or even more destructive, a tsunami, a major hurricane, or civil war. Censuses or sample surveys of survivors can be used to obtain retrospective reports of deaths or missing persons, and the current counts or estimates can be compared with counts or estimates made just before the disaster occurred to ascertain the extent of the excess losses.

Illustration of Use of GIS in Response to a Health Emergency

The following narrative illustrates the use of GIS in response to an actual health emergency. It recounts the case of the derailment in 1996 near Alberton, Montana, of a train carrying 122,000 pounds of chlorine, a lethal gas, and the escape from the train of plumes of chlorine gas (Lang 2000). Since 1991, the CDC's Agency for Toxic Substances and Disease Registry (ATSDR) has used GIS software to create computer models showing the spread of chemicals through soil, air, or water and the effects on people who are exposed to hazardous substances. When the train derailment occurred, ATSDR gathered information about the chlorine spillage, such as the total supply of chlorine on the train, the extent of the leak, the nature of the terrain, and the area in which the gas fumes leaked and would likely spread. Maps were prepared showing, among other things, the location of the escaped gas in relation to population, especially the vulnerable population. The latter were considered to be young children, the very old, people with lung conditions, and

pregnant women. For this, the analysts used, among other sources, census data on the age-sex composition of the population in the various geographic units in the vicinity of the spill. The number of women of childbearing age was taken as a proxy for pregnant women. The population maps helped to decide whom to evacuate and when to evacuate them.

The Environmental Protection Agency (EPA) staff estimated how the chlorine would move through the area, depending on rain and wind. As a result of this analysis, about a thousand people in Alberton and the surrounding area were evacuated from their homes and a 49-mile portion of the highway was closed. Federal, state, and local staff interviewed people evacuated from their homes to assess whether they were experiencing any symptoms from the gas spill, and entered these data into the GIS. ATSDR analysts used GIS to create maps that the interview teams could use to track the progress of the fumes and estimate the size of the potentially exposed population and the length of time it would take for the fumes to dissipate through the river basin. By recording each person's symptoms alongside their address, the analysts were able to estimate how much chlorine was in the air at a particular location. Evaluation of the emergency response was made by comparing the predicted events with actual events regarding the number of people affected, the path of the chlorine fumes, and other phases of the disaster.

Measuring Excess Deaths and the Excess Death Rate

The demographic analyst or epidemiologist may be asked to determine the number of excess deaths resulting from a human-caused or natural disaster, and the extent to which the death rate "spiked" as a result of the disaster. A variety of emergency situations may produce an abnormally high death rate and there have been many such situations in recent memory. In addition to national wars (e.g., Iraq-Iran war), civil wars (e.g., Lebanese civil war), and the nuclear plant explosion in Chernobyl, Ukraine, of earlier decades, we can enumerate the attack on the Twin Towers of the World Trade Center, New York City, in September 2001; the heat wave in France in 2003; the Sumatra-Andaman Earthquake and Tsunami in December 2004; hurricanes Katrina and Wilma on the southeast coast of the United States in September/October, 2005; the earthquake in the Pakistani-controlled area of Kashmir in October, 2005; the petrochemical-plant accident spilling chemical toxins into the nearby river and polluting the water supply in Harbin, China, in November 2005; the famine in Niger in 2005 and the prior period; and the Iraqi occupation and civil war initiated in 2003.

Estimates of the number and percent of excess deaths due to a disaster can be derived by the following general calculations, assuming that the reference period and area have been clearly specified and that migration to and from the area is minimal:

(1) Ascertain the total number of deaths that occurred during the period of the disaster.

(2) Estimate the number of deaths that would have occurred in the normal course of events during the period, assuming that the disaster had not occurred.
(3) Subtract (2) from (1), to secure an estimate of excess deaths in the affected area. The result would include not only the deaths directly due to the disaster but also any secondary deaths that resulted indirectly from it because of the turmoil (e.g., disruption of health services and lack of safe water and sanitation facilities).
(4) Take the ratio of (3) to (2), to derive the percent of excess deaths during the period.

The concept of the number of excess deaths as the number of deaths above the number that would normally have occurred or would be expected to occur during the period had the disaster not occurred is relatively straightforward, but the measurement of it is complicated by several factors. First, disasters are often accompanied by large population movements, especially movements away from the disaster area. Hence, given the rapid changes in the size of the population, it is not evident for which population the estimate of excess deaths is to be calculated and serious concerns can be raised about the representativeness of the sample if a survey is taken. The greater the turmoil and the migration, the more problematic it becomes to estimate deaths. This is the case in trying to evaluate the mortality situation in the counties of Louisiana and Mississippi during hurricane Katrina and the flood of 2005. Second, during many disasters, government operations are disrupted, official record-keeping is irregular, and the required demographic data are not available. This is the case in trying to evaluate the mortality situation in Iraq during the occupation and civil war that occurred there in 2003 and after.

Most excess deaths in a disaster occurring currently in MDC are due to a few exogenous causes, typically serious injuries and other violent causes, but victims may also die from a variety of other causes, particularly if the disaster is a war. It is necessary to include in the estimate of excess deaths the nonviolent secondary deaths resulting from the disaster, such as deaths from the destruction of the sanitary system, disruption of the sewage system, reduced energy supply, forced migration of refugees to temporary abodes, departure of health personnel to secure areas, and other disruptions of health services. In many conflicts in the LDC, disease and malnutrition and similar indirect factors accompany the conflict and account for more deaths than violence.

Not all deaths caused by a disaster occur immediately after the disaster strikes, however. The disaster may be followed by a trail of infectious diseases, starvation, or delayed deaths from radiation, so that the period of reference may have to be extended beyond the years of the original event. This would apply to the nuclear plant explosion at Chernobyl, Ukraine, in 1986. The effects of radiation are still being experienced by current and former residents of the area and so it is difficult to delineate the area and period of health impact. While the geographic area of reference should be kept as limited as possible, it may have to be expanded if the venue of the disaster shifts or spreads because of, say, pollution of rivers, scattering of plumes of noxious gas by wind, or dispersal of nuclear radiation.

In some cases the situation is stable enough so that the actual number of deaths and death rates are not difficult to ascertain, and the problem is to impute the "normal" or expected number of deaths and death rate during the crisis (e.g., French heat wave in 2003). In other cases there is considerable turmoil and the population situation during the crisis is quite difficult to evaluate (e.g., Iraqi occupation and civil war, 2003 to date; Kashmir earthquake, 2005).

If an effective death registration system has existed, it should be able to provide item (1) above unless the disaster has disabled the system. Item (2) must be estimated. It can readily be derived by interpolating between the death rate for the years just before the disaster occurred and the years just after the disaster occurred. Simple mathematical interpolation should serve this purpose adequately if the disaster period lasted only a few years and the disaster has not decimated the population greatly. Otherwise, this determination becomes problematic.

A parallel calculation provides an estimate of the excess of the observed death rate over the death rate expected in the absence of the disaster. This calculation may require an estimate of the population of the affected area during the year(s) of the disaster. The death rate will tend to rise as a result of both the increase in deaths and the reduction in population caused by the disaster. A rough estimate of the relative excess of the observed death rate over the expected death rate can be obtained by calculating the relative excess of the number of observed deaths over expected deaths.

Excess Deaths from the Conflict in Iraq, 2002–2006

As an illustration of a methodology for estimating excess deaths under circumstances of civil disorder, let us consider the leading study designed to measure the number of excess deaths of Iraqis due to the U.S. occupation of Iraq in 2003 and the ensuing sectarian violence. The study was conducted by epidemiologists at the Johns Hopkins University School of Public Health and Iraqi colleagues, and designed to produce estimates of excess deaths of Iraqis between January 2002 and March 2003, and between March 2003 and July 2006 (Burnham et al. 2006). Measuring the excess deaths resulting from this conflict is of special importance because the findings could affect public opinion in the United States and the decision of the U.S. government to continue to pursue the war. At the same time it is an extremely difficult task because of the wartime conditions prevailing both during the reference period and the period during which the data were collected.

Various methods can be used to count deaths under such circumstances: A count of deaths in hospitals; mortuary tallies; media or military forces reports; household interviews, alone or supplemented by other sources; and linkage of two or more collection systems by multiple-systems analysis. Multiple-systems analysis is the case-by-case matching of the records of two or more collection systems for the purpose of evaluating each of the sets of records against the other and deriving a single estimate. Experience in other countries shows that surveys inquiring about deaths tend to give much higher estimates of deaths than passive surveillance reports.

In no recent conflict situation except Bosnia were more than 20% of the total deaths recorded by passive surveillance as compared with population-based methods (Burnham et al. 2006). Nevertheless, surveillance estimates are important in monitoring trends over time, particularly in years of little violence.

As stated earlier, in times of war the death toll is augmented not only by "war deaths" but also by a variety of other factors, such as the deterioration in health services and an increase in infectious diseases. All these peripheral factors were at play in Iraq, contributing to the risk of death. The data on deaths during the period 2002–2006 were collected in a national sample survey of households taken between May and July 2006. A multistage cluster design was used: 47 clusters, each consisting of 40 households, were randomly selected from 16 Governorates. Selection of these initial survey sites was proportional to population size. Then streets and blocks in the survey sites were selected by random number. One household was first selected randomly in each cluster, and then the 39 adjacent households were added.

Various adjustments in the design of the survey were made to accommodate to war circumstances. For example, global positioning units (GPS) were not used since interviewers felt that being seen with a GPS unit could put their lives at risk. For the same reason callbacks were not made. Vacant households (i.e., households in which all members had died or had departed) and households that refused to be interviewed (totalling 1.7% of the households) were simply passed over, and other households were added to reach the quota of 40 interviewed households. The sample finally consisted of 1,849 households and 12,801 individuals. The households were asked about the members who live in the unit at present, members who had lived in the unit on Jan. 1, 2002, and births, deaths, in-migrants, and out-migrants since that time. Where deaths were reported, interviewers asked to see a copy of the death certificate. Death certificates were produced for 80% of all deaths reported. Fifteen percent of the households reported in- or out-migration.

The interviewed households reported 629 deaths (87% in the post-invasion period). The absolute number of deaths corresponded to a death rate of 13.3 deaths per 1,000 population per year (95% CI, 10.9–16.1) in the 40 months following the invasion, compared to a pre-invasion death rate of 5.5 deaths per 1000 population per year (95% CI, 4.3–7.1). The study produced an estimate of 655,000 excess Iraqi deaths as a consequence of the war (95% CI, 393,000–943,000). Of these deaths, 601,000 (95% CI, 426,000–794,000) were due to violence (mainly gunfire) and the remainder to other causes. In contrast, the several official and unofficial estimates, such as the Iraq Body Count, were far lower.

According to the report of the research, death rates were calculated on the basis of the mid-interval population and with log-linear regression models. To estimate the relative risk of mortality, a baseline rate, and then the relative rates for three post-invasion intervals, were determined. The number of excess deaths was estimated by subtraction of the values for the pre-war mortality rates from the wartime mortality rates in the three post-invasion periods.

In addition to the possibilities for bias characteristic of all surveys, the extreme insecurity prevailing during this survey could have introduced additional biases,

given the small size of the field staff, the short length of time that could be spent in each interview, the omission of callbacks, the omission of households all of whose members were killed or had departed, and the quality of the population estimates. These limitations would generally suggest underreporting of deaths. On the other hand, missing persons could have been reported as deaths by households and on death certificates, and the results could have been biased as a result of the migration of several million families during the reference period of the survey.

In sum, serious questions could be raised about the representiveness of the survey sample, given the small number and geographic distribution of clusters, the great variability in mortality levels from cluster to cluster, the considerable volume of migration, and the arbitrary adjustments that had to be made in sample selection. The results do not pass the tests of reasonableness. The total estimate of deaths implies an incredibly high number of deaths in Iraq each day since the invasion (560), and the (95%) confidence interval ranges from about 40% above the estimate to 40% below the estimate. In keeping with a common practice in demographic estimation under such circumstances, it might have been wise for the authors to have averaged their estimate with the much lower estimates from the other leading sources, such as the Iraq Body Count, weighting the estimates inversely in proportion to their estimated or presumed standard errors.

The collection of data on casualties and the estimation of casualities in times of war and other mass violence generally presents extremely difficult circumstances, especially in less developed countries. Sampling designs may have to be modified, compromising on representativeness, to deal with the dangerous conditions. Respondents may be afraid to respond honestly or to respond at all. They may have difficulties communicating with the inteviewer because of language or cultural/socioeconomic differences. Adequate administrative records may not be available for the LDC for the period before the violence started or after the violence has ended as well as for the period of violence, so the measures are even more tenuous. Several estimation methods should be applied and merged to derive the most reasonable estimate of events.

Mortality Analysis for Hurricane Katrina and Flood, 2005

Different problems of measuring excess mortality are presented by hurricane Katrina and the flood that followed it in the U.S. Gulf states in 2005. If we want to determine how many excess deaths resulted from this disaster, we have to decide first on the delineation of the area affected and the population to be analyzed. A substantial share of the original population of the Louisiana/Mississippi delta fled from the area during the flood. Should we measure the excess deaths only in the delta region or follow the migrants to their new residences in other parts of the Southwest and elsewhere? On reflection, it appears impractical to carry out the measurements for any area other than the area directly affected by the disaster. Yet there may be serious difficulties in securing accurate counts of deaths and developing accurate population estimates for this area during the time of the crisis.

We can consider a few ways of handling the problem. We have reasonably accurate population estimates for the affected area, some 117 counties, before and after the disaster from the 2000 census, the Census Bureau's current population estimates program, and the population estimates from the American Community Survey. We can derive estimates of the "normal" or "expected" death rate during the disaster period for the affected area, defined in terms of whole counties, by accepting the death rates just before the disaster or interpolating between the death rates before and after the disaster. We can derive estimates of the number of "normal" deaths on the basis of the available population estimates and our imputed death rates. Counts of actual deaths during the disaster can be reconstructed from the official records and other sources, such as records of undertakers. The U.S. Census Bureau has developed special population estimates for January 1, 2005, July 1, 2005, and Jan. 1, 2006, and the components of change for the 117 counties during this period. With estimates of the actual numbers of deaths and of the "normal" number of deaths, the excess number of deaths resulting from the disaster can be determined.[12]

Another approach is to apply estimated survival rates for Louisiana and Mississippi to the initial population of the affected area, to derive estimates of expected deaths in this area (16.8a). To derive the excess number of deaths from the expected deaths, we can use the actual number of deaths given in the official records or generate alternative estimates from equations 16.8b,c, and d. The survival rates can be based on life tables, adjusted by extending the life tables for these states for 1989–1991, or for 1999–2001 when these state life tables become available. The following formulas may be used:

(1) To estimate expected death on the assumption of no change in population,

$$D_e = P_0 P_o S \tag{16.8a}$$

(2) To estimate net migration approx., apply

$$M = (P_1 - P_o S) \div \sqrt{S} \tag{16.8b}$$

(3) Use the basic population equation,

$$P_1 = P_0 + B - D + M \tag{16.8c}$$

(4) To solve for D, insert M from (2)

$$D = P_0 - P_1 + B + M \tag{16.8d}$$

where P_1 and P_0 refer to the actual population of the affected area before and after the disaster, S refers to life table survival rates, B refers to recorded births, and

[12] There appears to have been about 3,000 excess deaths in Louisiana between July 1, 2005 and July 1, 2006. In this period there was an estimated 241,000 net outmigrants as compared with 12,000 in the previous year.

M refers to net out-migration. The calculation in (1) should be carried out for age groups. The population figures could be secured from the ACS.

Social Impact of Disaster

Civilians bear a heavy burden in war and other conflicts that are waged in the areas they inhabit. In the Vietnam war 3 million civilians died, and an estimated 200,000 of the 800,000 total population died in the conflict in East Timor (Burnham et al. 2006). Two hundred thousand people have died in Darfur, Sudan. In most cases of disaster, the burden of death and disease is borne unequally by the different age, sex, and socioeconomic classes in the population. Among the beleaguered populations, old persons, children, and poor persons tend to be the most vulnerable. In the Kohistan District of Afghanistan an estimated death rate of 2.6 per 100,000 per day occurred in a 4 1/2-month period from late 2000 and early 2001 during the civil war and drought. Diarrhea, respiratory tract infections, measles, and scurvy caused most of the deaths, battle deaths accounted for only a minority of the deaths, and children were among the leading casualties (Assefa et al. 2001).

In the heat wave that occurred in France in 2003, residents of retirement homes, persons over 75 years of age, women, and unmarried persons were disproportionately affected. The poor are typical leading victims. The poor are less likely to have home air conditioning to confront a heat wave; they are less able to afford the extra heating fuel needed in a severe winter; they are less likely to have the means to evacuate quickly and safely or to have storm-proof homes in a hurricane or flood; and they are more likely to live in the low-lying coastal areas where the floods and hurricanes cause the most severe destruction. The events following hurricane Katrina and the accompanying flood in New Orleans and vicinity in the fall of 2005 took its greatest toll on the poor.

References and Suggested Readings

General

Pol, L. G., & Thomas, R. K. (2000). *The demography of health and health care* (2nd ed.). Norwell, MA: Kluwer Academic Publishers.

Estimating and Projecting Population and Health Characteristics of SmallAreas

Farrell, L. P., MacGibbon, B., & Tomberlin, T. J. (1997). Empirical Bayes small area estimation using logistic regression models and summary statistics. *Journal of Business and Economic Statistics, 15*, 101–108.

Gonzalez, J. F., Placek, P. J., & Scott, C. (1993). Synthetic estimation in follow-back surveys at the National Center for Health Statistics. In U.S. Office of Management and Budget (Ed.), *Indirect estimators in federal programs* (Statistical Policy Working Paper 21) (pp. 2–1 to 2–11).

Institute of Medicine, National Research Council. (1997). *Improving health in the community: A role for performance monitoring*. J. S. Durch, L. A. Bailey, & M. A. Stoto (Eds.). Washington, DC: National Academy Press.

Levy, P. S. (1979). Small area estimation – synthetic and other procedures, 1968–78. In U.S. National Institute on Drug Abuse, J. Steinberg (Ed.), *Synthetic estimates for small areas: Statistical workshop papers and discussion* (NIDA Research Monograph 24) (pp. 4–19). Washington DC: U.S. Government Printing Office.

MacGibben, B., &. Tomberlin, T. J. (1989). Small area estimates of proportions via empirical Bayes techniques. *Survey Methodology*, *15*, 237–252.

Malec, D. (1993). Model-based state estimates from the National Health Interview Survey. In U.S. Office of Management and Budget (Ed.), *Indirect indicators in federal programs* (Statistical Policy Working Paper 21) (pp. 8–1 to 8–21). Washington DC: U.S. Government Printing Office.

Malec, D., Sedransk, J., & Tompkins, L. (1993). Bayesian predictive inference for small areas for binary variables in the National Health Interview Survey. In C. Gatsonis, J. S. Hodges, R. E. Kass, & N. D. Singpurwalla (Eds.), *Case studies in bayesian statistics* (pp. 377–389). New York: Springer.

Malec, D., Sedransk, J., Moriarty, C. L., & LeClere, F. B. (1997). Small area inference for binary variables in the National Health Interview Survey. *Journal of the American Statistical Association, 92*, 815–826.

Metropolitan Washington Public Health Assessment Center. (2001). *Community health indicators for the washington metropolitan region*. Washington, DC: MWPHAC.

Namekata, T., Levy, P. S., & O'Rourke, T. W. (1975). Synthetic estimates of work-loss disability for each state and the District of Columbia. *Public Health Reports, 90*, 532–538.

Siegel, J.S. (2002). *Applied demography: Applications to business, government, law, and public policy* (Chaps. 9 and 10). San Diego, CA: Academic.

Siegel, J.S., & Swanson, D.A. (Eds.). (2004). *The methods and materials of demography* (2nd ed.) (Chaps. 20 and 21). San Diego, CA: Elsevier/Academic.

Smith, S., Tayman J, & Swanson, D.A. (2001). *State and local population projections: Methodology and analysis*. New York: Kluwer Academic/Plenum Publishers.

Stoto, M.A. (2004, November 6). *Statistical methods for state and local public health data. Community health assessment*. Washington, DC: APHA/ASA Continuing Education Institute.

U.S. National Center for Health Statistics. (1977a). *State estimates of disability and utilization of medical services: United States, 1969–71* (DHEW Publication No. (HRA) 77–1241. Health Resources Administration). Hyattsville, MD: U.S. National Center for Health Statistics.

U.S. National Center for Health Statistics. (1977b). Synthetic estimates of state health characteristics based on the Health Interview Survey. By P. S. Levy & D. K. French. *Vital and Health Statistics: Data Evaluation and Methods Research, 2*(75). (DHEW Publication No. (PHS) 80-1356). Hyattsville, MD: U.S. National Center for Health Statistics.

Graphic Methods and Geographic Information Systems

Albert, D. P., Gesler, W. M., & Levergood, B. (2000). *Spatial analysis, GIS, and remote sensing applications in the health sciences*. Detroit, MI: Ann Arbor Press,

Barabba, V. (1975). The graphic presentation of statistics. In *A numerator and denominator for measuring change*, Technical Paper 37 (pp. 89–103). Washington, DC: U.S. Census Bureau.

Bryan, K.N., & George, R. (2004). Appendix D, Geographic information systems, In J. S. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed.). San Diego, CA: Elsevier/Academic.

Carbaugh, L. W., & Marx, R. W. (1990). The TIGER system: A Census Bureau innovation serving data analysts. *Government Information Quarterly, 7*, 285–306.
Curtis, A., & Leitner, M. (2006). *Geographic information systems and public health: Eliminating perinatal disparity*. Hershey, PA: IRM Press. Esp. Chapters II and III.
Gesler, W. (1986). The uses of spatial analysis in medical geography: A review. *Social Science and Medicine, 23*, 963–973.
Glass, G. E., Schwartz, B. S., Morgan, J. M., III, & Johnson, D. T. et al. (1995). Environmental risk factors for lyme disease identified with geographic informtion systems. *American Journal of Public Health*, *85*(7), 944–948.
Gobalet, J., & Thomas, R. (1996). Demographic data and geographic information systems for decision making: The case for public health. *Population Research and Policy Review, 15*, 537–548.
Gould, P. (1993). *The slow plague: A geography of the AIDS pandemic*. Oxford, UK: Blackwell.
Kaiser, R., Spiegel, P.B., Henderson, A.K., & Gerber, M.L. (2003). The application of geographic information systems and global positioning systems in humanitarian emergencies: Lessons learned, programme implications and future research. *Disasters, 27*(2), 127–140.
Khan, O. A., & Skinner, R. (Eds.). (2003). *Geographic information systems and health applications*. Hershey, PA: Idea Group Publishing.
Kitron, U., Pener, H., Coston, C., Orshan, L., et al. (1994). Geographic Information System in malaria surveillance: Mosquito breeding and imported cases in Israel, 1992. *American Journal of Tropical Medicine and Hygiene, 50*, 550–556.
Lam, N. S. (1986). Geographical patterns of cancer mortality in China. *Social Science and Medicine, 23*, 241–247.
Lang, L. 2000. *GIS for health organizations*. Redlands, CA: Environmental Systems Research Institute.
Mitchell, A. (1997–1998). *Zeroing in: Geographic information systems at work in the community*. Redlands, CA: Environmental Systems Research Institute, See esp., Chap. 10.
National Research Council. (1996). *Bridge builders: African experience with information and communication technology*. Washington, DC: National Academy Press.
Rushton, G., Krishnamurti, D., Krishnamurthy, R., & Song, H. (1995). A geographic information analysis of urban infant mortality rates. *Geographical Information Systems, 5*, 52–56.
Siegel, J.S. (2002). Geodemography: Geographic area concepts and information systems for demographic applications, In *Applied demography: Applications to business, government, law, and public policy* (Chap. 5). San Diego, CA: Academic.
United Nations, ESCAP. (2004). *Guidelines on the application of GPS in modern mapping and GIS technologies to population data*. Accessed at www.unescap.org/stat/pop-it
Vine, M.F., Degnan, D., & Hanchette, C. (1997). Geographic information systems: Their use in environmental epidemiologic research. *Environmental Health Perspectives, 105*(6), 598–605.
Wartenberg, D. (1992). Screening for lead exposure using a geographic information system. *Environmental Research, 59*(2), 310–317.

Measuring Vulnerable Populations and Losses in a Disaster

Assefa, F., Jabarkhil, M. Z., Salama, P., & Spiegel, P. (2001). Malnutrition and mortality in Kohistan district, Afghanistan. *Journal of the American Medical Association, 286*(21), 2723–2728.
Burnham, G., Lafta, R., Doocy, S., & Roberts, L. (2006). Mortality after the 2003 invasion of Iraq: A cross-sectional cluster sample survey. *Lancet, 368*, 1421–1428.
Population Council. (2006). Population assessment: Council study clarifies state of vulnerable populations after Pakistan earthquake. *Population Briefs, 12*(3).
World Health Organization. (2005). Rapid assessment. In M. A. Connolly (Ed.), *Communicable disease control in emergencies: A field manual* (Chap. 1). Geneva, Switzerland: World Health Organization.

Chapter 17
Some Ethical and Legal Aspects of Health Demography

Health demography presents numerous issues having ethical and legal implications, as we shall see. This chapter describes only some of them and hence is only illustrative, not comprehensive. The discussion is divided into two main parts, issues relating to fertility and issues relating to mortality, with health issues being the umbrella theme. Ethical and legal aspects of health demography encompass issues pertaining to sex, age, race/ethnicity, and socioeconomic categories, access to and quality of health care as affected by membership in one or other of these classes, and the definitions, and hence the determination, of life and death. In recent decades many bioethical issues, some old and some new, have been vigorously debated publicly. These issues include such matters as: The commercialization of childbirth and health; the control over the characteristics of newborn children; the timing of the beginning and end of human life; the right to bear children, terminate a pregnancy, and die; the right to medical care and the rationing of health care; and the proper subjects and objects of medical research. The resolution of these issues involves the application of ethical principles, moral guidelines, and value judgments, following various philosophical, religious, and legal traditions. Research in this area defines the field of medical bioethics. This field has grown as the variety and complexity of issues involving value judgments in health and medicine have grown. Technological developments relating to the beginning and end of life, new diagnostic and therapeutic equipment, and advances in medical knowledge have all contributed in no small part to the emergence and salience of these issues.

Fertility Issues

Genetic Testing and Gene Therapy

Since the last quarter of the last century there have been great advances in genetic testing and fertility technology that have proved important in addressing various problems in human fertility. Among the new developments is the ability to determine

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4_17, © Springer Science+Business Media B.V. 2012

the sex of the fetus, its general health condition *in utero*, and whether it is a carrier of certain genetic diseases and disorders. It is now possible to treat some health conditions of the fetus *in utero*. Other developments include the ability to freeze and store both egg and sperm for use at a later date, possibly a distant date, including use after the death of the donors.

Experiments are being carried out by geneticists with fertility genes that are shared by humans and other species. Experiments now under way to replace genes that contribute to infertility in mice may eventually be used to address the problem of human infertility. The more difficult problems we face today are how to address the ethical implications of genetic modification, how to deal with the social stigma related to the newly acquired knowledge about the harmful genes we possess as individuals, and how to avoid doing physical harm in the ongoing efforts to manipulate the human genome in treating either human infertility or other disorders.

Sex Determination of Births

Parents throughout the world have a mild to strong preference for boys. The preference or boys is especially strong in the Far East, but even in the Western industrialized countries boys are mildly preferred. It is now possible to determine with considerable accuracy the sex of the fetus after just 3 or 4 months of uterogestation. In countries where the preference for boys is strong and sex-selective abortion is allowed or disregarded, in time the balance of the sexes in the population will become grossly distorted and the consequences can be socially destructive. Accordingly, many Western countries, among them Canada, Germany, and Great Britain, allow genetic testing and amniocentesis (i.e., examination of amniotic fluid) in order to determine the sex of the fetus and genetic abnormalities in the fetus, but prohibit their use for sex-selective abortion.

As in these countries, so in other countries where abortion is permitted, it may be desirable to limit sex-selective abortion or prohibit it entirely. This is the view expressed in the report of the President's Council on Bioethics (2003). There is a general principle guiding public policy here. As a general rule we cannot assume that the aggregate outcome of individual decisions, particularly reproductive decisions, however positive for the individual, is consistent with the public interest. In the public interest it may be necessary for society to prohibit individuals from implementing some decisions, such as sex-selective abortion, because of their destructive effect on society.

There are no data on the sex of aborted fetuses in the United States, so that we cannot determine from this source if there is a preference for boys. The Guttmacher Institute, a research institute that compiles data on abortions, reports that virtually none of the women seeking abortions in the United States mention sex preference as the reason for requesting the procedure (L. Darabi, June 10, 2004, personal communication from the Guttmacher Institute, to the author). There is no evidence from other sources to suggest any statistically significant sex selection in the United States.

Assisted Reproductive Technology

Clinical procedures designed to aid prospective parents to have children who are unable to conceive or carry children on their own are encompassed by the term assisted reproductive technology. They include cases where either the woman or man, or both, are infecund, and possibly cases where the woman does not want to bear her own child. The forms of assisted reproduction include intrauterine insemination, donor insemination, *in vitro* fertilization (IVF), and intracytoplasmic sperm injection. Some persons and organizations oppose all forms of assisted reproductive technology on moral and religious grounds, as does the Catholic Church. Hence, the practice is controversial, and the peripheral aspects of IVF, such as the storage and disposal of unused sperm, eggs, and embryos, raise additional ethical and moral issues.

Infecundity

As the reader may recall, infecundity refers to the physiological inability to parent a living child – a condition characterizing several percent of American couples. Women are naturally infecund after menopause, and men become infecund at somewhat later, less determinate, ages. The chances of infecundity or subfecundity rise with increasing age of the woman and man, especially for women who are over age 35. The percentages of infecund and subfecund women and men have been rising in recent decades because of the trend of late marriage and postponed childbearing. Gene abnormalities and immune system disorders may play a role in female infecundity between menarche and menopause. Men is subfecundity is due to the production of too few sperm or low-motility sperm (National Collaborating Center for Women's and Children's Health 2004). Genetic factors are also associated with male infecundity. These include chromosomal abnormalities (i.e., a Y chromosome deletion) and impaired gonadotropin secretions.

Dealing with infecundity. In the case of the inadequacy or insufficiency of the sperm of the husband, a sperm sample may be obtained from an outside donor and injected into the uterus of the prospective mother, where it may become implanted. In this case, donor insemination, the donor is not the spouse of the mother but is the biological father of the prospective child. This process has raised a number of bioethical issues, among them the confidentiality of the identity of the biological father and the disposition of the unused frozen sperm obtained from donors.

To resolve the problem of excluding the husband, intrauterine insemination or a form of *in vitro* fertilization named ICSI (intracytoplasmic sperm injection) may be employed. In the latter procedure a single sperm from the husband is injected directly into an egg cell taken from the prospective mother and, then, if an embryo is formed, it is implanted in her uterus. According to the Centers for Disease Control and Prevention, more than 90% of U.S. fertility laboratories offered the latter procedure in 1999 (*Scientific American*, April 2002:38). Of course, if intrauterine

insemination with the husband's sperm succeeds alone or in combination with donor insemination, the other forms of assisted reproductive technology are not necessary though they still may be preferred.

Now consider the issue of the identity of the biological mother where the prospective mother is infecund. There are two types of situations to be noted, one where an "outside" woman contributes eggs for *in vitro* fertilization and the second where the "outside" woman carries the baby to term as a substitute or surrogate for the prospective mother, using the latter's egg. In the former case, that is, where another woman contributes the egg, the identity issue is the same as where an "outside" male is the biological father. In the latter case there is little issue with the identity of the biological mother since the surrogate mother is often designated by the biological mother. The surrogate mother who carries the fetus may be a sister or friend of the biological mother.

Ethical/legal issues also arise in connection with the selection and screening of the sperm or ovum donor. The surrogate mother may change her mind and want to keep the child, the "outside" participant may identify herself or himself contrary to the wishes of the "adoptive" parents, or the ownership and disposal of stored embryos may be disputed by divorced couples. IVF presents a serious risk of creating a child with serious chromosomal abnormalities, particularly where either donor is of advanced reproductive age.

Human Cloning

The most extreme possibility for assisted reproduction is called cloning (or genetic reprogramming by its ardent supporters). First, a distinction must be made between reproductive cloning and therapeutic cloning, the former being for human reproduction and the latter for treatment of specific human disorders. The administration of President Bush (2nd) opposed both therapeutic and reproductive cloning. In 2002 a panel of the National Academy of Sciences recommended support for therapeutic cloning and a ban on human reproductive cloning. The Academy's position on human reproductive cloning was like that of the President's Council on Bioethics, which recommended that reproductive cloning be strictly banned along with any other techniques for human procreation except the fusion of a human egg and sperm. The Council also urged that experiments on human embryos would be acceptable only if the embryos are not maintained past an early stage of development (e.g., 14 days or less). These guidelines would allow investigators to collect the needed stem cells from the embryos for therapeutic cloning. In an executive order issued March 3, 2009, President Obama lifted the existing ban on the use of Federal funds to support embryonic stem cell research and expressed his strong opposition to human reproductive cloning. He called for research with embryonic stem cells to be directed toward therapeutic cloning only, not reproductive cloning.

Most persons oppose reproductive cloning. Nevertheless, the well-known Italian gynecologist Severino Antinori has launched a broad effort dedicated to the

production of human clones. Arguing in favor of reproductive cloning, Antinori maintains that cloning does not produce an exact copy of the individual donating the DNA sample, as is commonly believed. "Even if most of the clone's DNA comes from the donated nucleus," he argues, "the oocyte still contributes a small percentage of genes from the mitochondria," and this means that cloning to produce two identical individuals is impossible (*Scientific American*, April 2002:39). An oocyte is a cell from which a egg or ovum develops by division.

Use of assisted reproductive technology for human cloning presents several ethical and legal issues. Many millions of men in the United States do not produce any sperm and therefore cannot have a child of their own by mating or in-vitro fertilization (IVF). How much priority should be given to aiding such infecund prospective parents to produce children who are genetically similar to themselves? Some parents want to replicate a child who has died in an accident. An infecund woman may seek a genetic link to her recently deceased husband by desiring a clone from a sample of his tissue. Even more extreme, a husband may want a clone of his wife so that he can raise a "copy" of his wife as a child. How meritorious is the goal of siring a genetic relative rather just adopting a child? Is there a right to genetic similarity? Those who believe that there is such a right want to have infecundity classified as a disease so that more public funding and research can be directed to it. Bioethicist Glenn McGee (2003) believes that society should direct its limited research funds toward other goals than producing "vanity" children.

In addition to the moral and religious questions regarding reproductive cloning, there is the concern that some deformity will occur as a result of the procedure. Cloning may present a far greater risk of creating a child with serious chromosomal abnormalities than IVF. Most animal clones die before they are delivered or they suffer from severe birth defects, and the experts working on animal cloning warn of similar possibilities for humans. Some animal clones that look quite healthy at birth develop many diseases after birth, and there is currently no way to test for these prospective defects. Bioethicist Arthur Kaplan (2004) believes that the possibility of producing defective human clones is the most cogent argument against this procedure. Gynecologist Antinori rejects these arguments, noting that accurate prenatal screening could eliminate most of the defects. Furthermore, IVF for humans results in far fewer defects for humans than for animals, and this could carry over to cloning of humans.

Fertility at Unconventional Ages

The public and public policy endorse a given range of ages of women as appropriate for motherhood and biology sets its own limits. Questions are raised about motherhood at ages outside this "acceptable" range – that is, at very "young" ages and at very "old" ages of the reproductive period. The pertinent socioeconomic and health issues have been well documented but motherhood at extreme ages also raises bioethical issues as well.

As Antinori has shown, even post-menopausal women may be assisted in bearing a child with the aid of a donated egg and hormones. In 1989 he enabled a 47-year old woman to be the first post-menopausal woman to give birth, and in 1994 he enabled a 63-year old woman to give birth, the oldest woman ever to do so to that date. (In 2007, a woman 66 years of age was reported to have born a child.) The bioethical issues concern the acceptability of facilitating motherhood for "old" women, who are very likely to bear children with birth defects, are presumed to lack the energy required to rear an infant through childhood, and who are less likely to survive to the child's adulthood. Another issue concerns the propriety of teenage parenthood, particularly parenthood below 18 years of age. We know that very young parents are more likely to raise children destined to a life of low income, low education, a menial and hazardous occupation, and relatively poor health.

Issues in Female Reproductive Rights

Social, political, and legal trends in the United States have moved in the direction of enabling women to achieve gender equality in reproduction, including the right to terminate a pregnancy before birth occurs. The right of a woman to terminate her pregnancy remains controversial in the United States, however, even though the supreme law of the land supports this right. Given the sharp division in American public opinion, some restrictions on a woman's right to a medically safe abortion have been sought by opponents of legalized abortion. There have been indirect assaults on abortion rights in the form of legal efforts to restrict the conditions under which abortions may be secured and to have fetuses declared persons under the law. These efforts have been successful and a number of states have enacted such laws.

If abortion is recognized as a woman's right, the range of acceptable reasons for an abortion, particularly in the first two trimesters, becomes almost unlimited. This has been described as abortion on demand, and includes such reasons as preventing the birth of a child who is likely to have severe health problems (e.g., Down's syndrome), when the parents cannot afford to take care of an additional child, when the mother has a health condition making childrearing very onerous, when close spacing of children would be unhealthful for mother and child, and when the mother is still a child herself.

Mother's Right to Abort a Fetus

Let us note first that opponents of abortion view a mother's "right" to terminate a pregnancy, or a mother's "right" not to give birth, as a nonexistent right and see it rather as a "duty" of a pregnant woman to give birth. We may also phrase the issue as the "right" of a fetus to be born or not to be born under the circumstances confronting it with respect to its future health and socioeconomic status. In this view the issue can be conceptualized as the "right" of a fetus to be born well and

wanted, naming the mother as the surrogate for making the decision for the child. If we accept this position, it is the "duty" of a mother to bear a child only if it is likely to be born well and wanted. Much overlooked have been the rights and duties of fathers in this context. An unwanted pregnancy is not simply a question of a woman's mothering an unwanted child. If the father has legal obligations with respect to the unborn fetus and the prospective child, then he should have rights, in collaboration with the mother, in the decision to abort or not to abort the fetus.

My rephrasing of the issue is for heuristic purposes and should carry no implications as to whether a fetus does or does not have the rights of a human being. As I argue below, however, the operative question in resolving the abortion controversy should be, not when does life begin, but when does human personhood begin. (See section on "Stem Cell Research" below also.)

To understand the philosophical as well as legal basis of the decisions of the U.S. Supreme Court on reproductive health cases, it is useful to recognize the two very different principles of jurisprudence that guide U.S. Supreme Court Justices in such matters. Some Justices are strict constructionists or "originalists," and others are activists. The former feel that judges should adhere to the precise words of the Constitution and that the meaning of the words in the Constitution was locked into place at the time they were written. This philosophy usually leads to conservative positions on social issues like the right to terminate a pregnancy and the right to die. The alternative view is that the Constitution does not have a static meaning and that its principles must be adapted to deal with current conditions. This view recognizes that standards of decency change, appeals to a basic sense of fairness and justice while overlooking legal niceties, and flouts judicial restraint to pursue an egalitarian agenda. According to its opponents, this approach leads to the discovery of imaginary new rights, such as the right of privacy.

The right of privacy is the foundation of the decision in *Roe v. Wade* (1973), which established a woman's right to an abortion. In *Griswold v. Connecticut* (1965), a case involving the right to transmit contraceptives in interstate commerce, the Court recognized the right to privacy for the first time. In *Roe v. Wade* the Court ruled that a fetus was not a person with constitutional rights and that the right to privacy was inherent in the 14th Amendment's guarantee of personal liberty. This right to privacy protected a woman's right to an abortion. At first, the Court maintained that the decision to have an abortion should be left entirely to a woman and her physician during the first trimester of pregnancy, but that some regulation of abortion procedures was allowed in the second trimester and that further restrictions on abortions could be imposed in the third trimester. The Court later rejected the trimester framework, replacing it with a woman's right to an abortion up to viability, defined as being "potentially able to live outside the mother;s womb...," and adding that viability "is usually placed at about seven months..." This matter was previously left to state legislatures.

Strict constructionists believe that the right of privacy, and hence the right to an abortion, cannot be found in the Constitution and that the 14th Amendment's protection of liberty does not imply the right to privacy. When the Constitution's text is silent or ambiguous on a subject, the strict constructionists try to determine

whether a long-standing American tradition has supported the practice. Because abortion was outlawed in the United States for more than a century, they argue that the American tradition does not uphold a right to it.

The Court has wavered in its readiness to put restrictions on abortion while never rejecting *Roe v. Wade*'s essential ruling. Its decisions have been made against the background of continuing efforts by the states to limit the role of *Roe v. Wade*. Most states have parental notification rules requiring parents to be notified if a minor seeks an abortion. In 1983 the Supreme Court invalidated a series of city and state laws that had been designed to make it more difficult to obtain abortions; and in 1986 the Court invalidated a Pennsylvania law that imposed restrictions on access to abortion. On the other hand, in 1989 the Court upheld several prohibitions of Missouri state laws on abortion; and in *Planned Parenthood of Southeastern Pennsylvania v. Casey* (1992), the Court allowed the state of Pennsylvania to place certain obstacles in the way of obtaining an abortion. Through all these decisions, perhaps for lack of the necessary fifth dissenting vote, the Court upheld the essential principle of *Roe v. Wade*, that is, that abortion is a constitutionally protected right. In part, this position relied on *stare decisis*, that is, the principle of respecting precedent or letting a previous decision stand. None of the several Supreme Court decisions in the last several decades relating to the right of a mother to abort a fetus (or the "right" of a fetus not to be born) imputes personhood to the fetus.

Recent controversy has centered around so-called partial-birth abortions.[1] In *Stenberg v. Carhart* (2000) the Supreme Court struck down a Nebraska law that banned partial-birth abortions. It argued that the state law could be interpreted as banning other abortion procedures and that the state law should have made an exception for reasons of the health of the mother. However, in 2003, as if to support Nebraska's position, Congress passed the Partial-Birth Abortion Ban Act, a law banning "partial-birth abortions." In the consolidated cases of *Gonzales v. Carhart* and *Gonzales v. Planned Parenthood* (2007), the Supreme Court upheld the federal law. The court decision did not make an exception for the health of the woman following the federal law, but allowed the woman to seek a (federal) court's exception if she could show that her health was compromised by not employing the procedure identified as partial-birth abortion.

This latest court decision has little impact because abortions after the first trimester are rare. According to the CDC, only 4% of abortions occurs between the 15th and 20th week of uterogestation (early in the second trimester) and a negligible 1.4% occur after that (late second and third trimesters). The pro-abortion group saw the decision as "chopping away" at the fundamental right of women to secure an abortion, and the anti-abortion group saw it as abolishing a cruel and inhumane

[1]The exact definition of what constitutes "partial-birth abortion" is disputed. It is a procedure performed, usually in the second trimester of pregnancy, in which the life of the fetus is terminated after partial delivery of the fetus through the birth canal. Its medical purpose is to protect the health of the mother. Partial-birth abortion is the name given to the procedure by its opponents, but it is medically called "intact dilation and extraction."

procedure and endorsing a law of Congress that already had wide public approval. Proponents of the procedure decried the decision, in addition, on the grounds that it interfered with the right of a physician to make the best medical decision in an emergency, jeopardized a woman's health, and was paternalistic in seeking to protect the health of the woman on her own behalf. The opponents of the procedure find it objectionable in part because it resembles the killing of newborn infants, but their fundamental objection appears to be that they do not believe that the Constitution confers the right to an abortion and they oppose the practice of abortion in general.

The court decision still allows that abortions early in pregnancy must be protected, validating *Roe v. Wade*, but is consistent with public opinion in generally opposing abortions after the fetus is viable, that is, able to live independently of the mother. Some fetuses in the second trimester are viable so that an intact-dilation-and-extraction/partial-birth-abortion could be done on a viable fetus. Most fetuses are not viable at this early period, however. Public opinion strongly opposes abortions in the third trimester, and according to its opponents all third-trimester abortions fall in the class of partial-birth abortions. By this time nearly all fetuses are viable and the more salient issue is the risk to the life of the mother if the fetus is nearly brought to term and then aborted.

Some women base their decision to abort a pregnancy on the basis of the fact that the fetus has a genetic disorder. The existence of a genetic disorder may be determined by either of two procedures, apart from probabilities based on the health condition of the parents. They are amniocentesis and chorionic villus sampling (CVS). Prenatal diagnosis must be done in time for parents to make a safe decision to abort the fetus, treat the fetus during pregnancy, or treat the newborn child after delivery. CVS is usually performed late in the first trimester and amniocentesis is usually performed early in the second trimester. The former test incurs a slightly higher risk but the results are more readily available. The mother has to decide between the two tests with their different advantages for the health of the fetus.

Right to Parenthood

Historically, it was accepted that children resulted naturally from marriage and numbers were never an issue. Even the eugenics movement of the early part of the twentieth century concerned itself only with the negative quality of some births. Now that, in the western developed countries, most couples plan the number of their children, the total number of children born to women in their lifetime falls at or below replacement level – about 2.1, the number of children needed to replace each woman over a generation. In the event of sub-replacement fertility with no net immigration, a national population would soon decline in numbers. If national population policy considers population decline unacceptable, then governments must also try to remove obstacles to parenthood and encourage larger families as a

public duty. Some countries have family allowance programs and others deal with the issue through tax policy. Some are depending on immigration as a solution to the problem of prospective population decline.

A general ethical question is whether, in the context of the state's encouragement of responsible parenthood, every male and female of reproductive age should have the right to become a parent without restriction. The idea of restricting parenthood for some categories of persons has been debated for a long time. For some characteristics of the mothers or fathers the issue has been adjudicated in the courts. Prior to the last third of the last century, law and public opinion frowned on women who were not married having children, for example. Many were legally forced to give their children up to the state for adoption or residence in foster homes. More often today it is a question merely of negative public opinion. The following additional questions can be raised: Should a woman be permitted to have as many children as she wishes regardless of her ability to support them? What of unmarried women who are very young or very old (for parenthood)? Should women who have a serious illness that may put themselves or the fetus at risk, particularly those women who have a transmissible illness such as HIV or a potentially lethal illness such as ovarian cancer or aortic heart valve stenosis? Finally, what about men and women who are developmentally retarded or suffer from serious mental illness?

These questions arise today not primarily because of eugenic concerns, but rather because the "rights" of the fetus may be violated by the almost certain prospect of being born inadequately cared for, unwanted, extremely ill, or burdened with the expectation of serious illness at a later date. There is evidence that certain mental diseases or a predisposition to them are transmitted from one parent or other to the child. Or the concern may be that society will be forced to pay the bill for the birth and/or the care of the child, under the circumstances described, through insurance or public funds. Society can be caught between its general reverence for the life of a newborn child and the newborn child's right to be born "well and wanted."

Marital Status and Age

Consider the decision of a single woman who is 44 years of age, infecund and premenopausal, to have a child through insemination with donor sperm that will be paid for by an insurance company. The critics would say, she should not have waited so long. Supporters would say, she has a right to have a child when she feels it to be right for her even if she might have chosen to use the procedure when she was younger. But then it probably would have cost the public far less. Is not adoption a more socially desirable option? In spite of the availability of various types of assisted reproductive technology and changing concepts as to who can be a good mother and parent, the issue becomes, should an unmarried, infecund woman, particularly at the end of childbearing, be supported with community funds in having a child?

What about an "old" postmenopausal woman in her 50s or even 60s? As stated earlier, motherhood is now possible for such women. The public seems to view motherhood by a woman over 50, where the child is born through assisted

reproductive technology, as undesirable. The argument is raised that parents of this age would not have enough energy to rear a child properly, that when the child is of high-school age, the parents will be old, and that the child has a high risk of genetic disorders. In addition to age and marital status of the parents, other characteristics may be viewed as inappropriate conditions for further parenthood, for example, poverty status and/or receipt of public welfare, having sired several previous children, and illegal alien status. The specter of national or world overpopulation has been raised in such cases also. In fact, in countries with a democratic tradition, few or no restrictions are placed on parenthood, except for a subdued negative public opinion.

Physical and Mental Health

A separate group of questions relates to the mental and physical health of the prospective mother and father. The health condition may be severe and could impede their ability to care for a child. It may be a transmissable disease and the child could be burdened with the condition. HIV/AIDS mothers often bear HIV/AIDS children. One or both members of a couple may be mentally retarded. The right of mentally retarded parents to bear children was once a hotly debated issue that was considered frequently in the courts. The sanction in this case was sterilization of the woman. In addition, the child could be removed from the home and made a ward of the state. The assumption was that a mentally retarded woman was likely to give birth to a mentally retarded child and that she would not be able to rear the child properly, adding to the state's burden of care.

In 2007 the state of North Carolina threatened to remove a child from his mother's care on the grounds that the child's health was compromised by remaining at home. The case raises the general question as to the extent to which a mother can be held responsible for the health of her minor children and whether, by failing to comply with some general or specific health standard, she should lose custody of her child. The facts of the case are that the 8-year-old boy was morbidly obese and that a hormonal or genetic basis for the condition could not be found. As a result, the state claimed that the mother was physically abusing the child by allowing him to eat inordinate amounts of food. This assumes that there are established standards for "parentally allowed" obesity? What if the child is at risk of incurring a health condition because of lifestyle factors, such as sleep deprivation? Suppose the child does not exercise in addition to not eating properly, and so develops a heart condition? What if a child develops periodontal disease and the parent neglects the problem because of insufficient income? In these cases it may be argued that the woman is not in a position to provide a proper home for the child initially or in later years, particularly without major public support. This state of affairs would apply to multitudes of families in one degree or another.

Concluding Note

The issues relating to the right to parenthood represent a task of balancing the interests of the unborn fetus, the interests of the parents, and the interests of society. If the focus is on the welfare of the child and the public good, not on the satisfaction of the parents, all of these types of cases merit close public scrutiny. On the other hand, in general, in western societies the size of families is not subject to direct formal controls for any reason. For the most part, no real legal restrictions are in effect prohibiting parenthood for any of the conditions cited above. Couples are free to decide whether they want children or not, regardless of a public opinion that may look with disfavor on parenthood under certain conditions. Changes in public morals are increasingly making parenthood acceptable, even where the parents are unmarried, extremely poor and have had several children, are in ill health, or are otherwise viewed as unsuitable by some for rearing children. Moreover, giving up the children for adoption or foster home care is required only in cases of extreme physical or mental abuse. All this implies mutual consent of the parents and the absence of an abusive situation in the home.

I have been essentially discussing the situation in the More Developed Countries (MDC). There is even less involvement of society in parental fertility decisions in the Less Developed Countries (LDC). People in the MDC often object to parenthood in the LDC on some of the very grounds that have been raised above. In many of the latter countries, fertility rates are high and population growth is out of line with familial and societal economic resources. This objection, where valid, would apply to many families in their very own town.

Mortality/Health Issues

Implications of Extension of Human Longevity

In the United States, among those who accept the probability of a considerable extension of the human life span, there is a sharp ethical divide between those who believe that the extension of human life will have essentially positive outcomes and those who believe the opposite. Some see it as a continuation of medicine's quest to cure disease and as the promise of an increase in the quality of human life (Overall 2003). They believe that an extension of human life is both possible and desirable.

Others believe that an extension of human life is possible but undesirable. They are concerned about the personal and social changes that would accompany a considerable extension of human life, anticipating widespread social disruption and public fear of "indefinite" life. They see it as a diversion of precious health resources to old persons, who have already had an opportunity to accomplish in a long life what they needed to accomplish. Death in old age avoids burdening society with

many sick, disabled, and frail people. Still others regard a substantial extension of human life as both impossible and undesirable. This position merges a belief in what *is* with a belief in what *ought to be*. In the view of all those who consider the extension of human life undesirable, whether it is possible or impossible, old persons should accept the inevitability of death and the finiteness of life. This view joins the arguments given above with respect to the problems created by the overly long life.

Daniel Callahan, an eminent philosopher and ethicist, has written several books elaborating the latter position (1995a, b, 1998). He believes in rationing health care, in caring for, rather than curing, the very old and frail, in distributing health care so as to benefit the largest number of persons, and in particular, in rationing specialized care, such as transplant surgery, on a cost-benefit basis. He maintains that biomedical research has been pursuing the wrong goals and neglecting other important interests, being too concerned with life extension and too responsive to commercial interests (Callahan 2003). R.L. Barry (1991) has written a strong rebuttal of the Callahan position. The reader should note that health care rationing is not consistent with existing federal Civil Rights laws and with the United Nations' Universal Declaration of Human Rights.

The Right to Die vs. the Obligation to Live

Euthenasia and Physician-Assisted Suicide

The decision to assist in or be assisted in the process of death often arises on the occasion of the severe physical or mental incapacitation of the patient. While the patient is deemed to be of sound mind, he or she may request assistance in terminating his or her life if a severe and debilitating condition has made further existence intolerable. An advance medical directive (or living will, health power of attorney) can be also prepared by anyone and used in the event that the person cannot speak for themselves. The U.S. Supreme Court ruled in *Cruzan v. Director, Missouri Department of Health* (1990), that individuals have the right to refuse medical treatment and may specify their wishes about treatment in advance of a terminal illness or a life-threatening injury (Glick et al. 1996). Subsequently, Congress enacted the Patient Self Determination Act (1990), which requires many medical organizations to notify their patients of their right regarding advance medical directives.

There is widespread support in the United States and several other Western industrial countries for assisted death under well-defined safeguards in the event of irreversible or terminal illness. Belgium, Denmark, the Netherlands, and Switzerland have right-to-die laws. In cases of extremely serious illness terminated by assisted suicide in the United States, there is a tendency to "look the other way." Numerous cases of assisted suicide have been reported where the patient was

wasting from AIDS or other debilitating illness. Only one physician has been found guilty of homicide and no one has been charged with attempted suicide.

Our current national policy, however, does not support this practice and the Bush II administration strongly opposed it. The states regard it as their right to legislate on this matter. In accordance with the states' rights principle, in *Washington v. Glucksberg* (1997) and *Vacco v. Quill* (1997), the Court ruled that states can ban doctor-assisted suicide. In defiance of the Bush II administration's position on the question, the state of Oregon permits assisted suicide.

Apart from the moral and religious views opposing physician-assisted suicide, there is concern about abuse of the practice by patients and physicians. The experience in Oregon and Denmark, however, demonstrates that the latter concern is not justified; relatively few persons have opted for use of assisted suicide and these have been carefully screened.

In the United States there is little public support for a general unrestricted "right to die," or a right to suicide, with or without assistance. We can say then that, instead of a "right to die," the predominant public position in the United States is an "obligation to live." Public opinion, public policy, and prevailing moral standards endorse the principle of the "obligation to live." This is true even where the death of the individual would be a blessing to the ill person and his/her significant others, and would not negatively affect anyone. Many theologians and ethicists as well as other lay persons believe that there is often an unnecessary prolongation of dying as "personhood" is being lost.

The development of human life is a process, involving progressive stages from conception to implantation to fetal development, and to full personhood when the fetal brain has developed self-consciousness and self-awareness at birth. In this process the potential human life becomes actualized in full personhood. These stages of potential human life have different values relative to one another and to other animal life. While all these forms of life have value, they have different values. These different values guide us as to what and whom we should save first in the event of a disaster such as a fire or flood.

Mental Incapacitation

Similarly death is a process, involving a loss of personhood and self-consciousness that culminates in brain death with irreversible non-consciousness. Unlike the fetus, which is progressing into personhood with each stage of fetal development, dying persons have already experienced personhood throughout their lives since birth and are losing personhood with each stage of brain death. (Various stages of brain death are described below.)

With increasing longevity more and more persons are incurring conditions where there is impairment of the consciousness levels of the brain, with partial or total cerebral death. This impairment results when there is a malfunction of the nerve fibers connecting the brain and the sensory organs, or the blood flow and hence the oxygen supply to the brain are disrupted. For older persons the underlying cause

is often cardiac arrest and for younger persons the underlying cause is often a head injury. Cerebral impairment may also result from infections of the brain, toxic effect of drugs, hypoglycemia, a cerebral aneurysm, and similar causes. Medically-assisted procedures such as insertion of a feeding tube are required to keep the person alive, but more drastic procedures such as use of a respirator may be required.

The following states of unconsciousness may result from cerebral impairment: stuporous/minimally conscious, persistent vegetative, comatose, and brain-dead. These four states correspond approximately to progressively lesser degrees of awareness, or ability to respond to external stimuli or internal needs. They are sometimes not clearly distinguishable, even by neurologists, who may disagree on the exact type of consciousness-impairment of the patient. These distinctions are presented because they help to clarify the arguments made in the various right-to-die cases being disputed, but touch also on the criteria of death.

A *stuporous/minimally conscious state* is an unresponsive, sleeplike state from which a person can be aroused after numerous attempts, but only briefly and with strong stimuli. The patient is aware of the environment but makes few responses.

A *persistent vegetative state*[2] is characterized by a complete lack of cognitive function or awareness although there may be random responses to some sensory stimuli. The person may have normal sleeping and waking patterns and may make occasional noises. The possibility of recovery depends on the duration of the condition and the cause. When the cause is a head injury and the condition has persisted less than a few months, the individual has a fair chance of recovering. When the cause is cardiac arrest and the condition has persisted for more than a few months, recovery is very unlikely.

A *coma* is an even less responsive state from which a person cannot be aroused, however aggressive the means used. The individual cannot make the most basic responses although the reflexes may work.

Brain death[3] is the most severe form of unconsciousness and is evidenced by the fact that the brain can no longer perform any vital functions: There is an absence of muscle activity, cessation of breathing and other vital reflexes, and inability to respond to stimuli. There is no movement even in response to painful stimuli, no reaction of the eyes to light, and an inability to breathe without breathing-support equipment. Brain death has been considered as an alternative basis of defining death that would replace the current WHO definition.

Many persons in the United States at any particular time are in these states of limited consciousness and are maintained on life support equipment, removal of which would end their lives. In these cases physicians and hospitals are unwilling to "pull the plug" without explicit instructions from the patient to do so in the form of a Living Will or an Advance Directive requesting them not to use "heroic" methods of sustaining life. The Living Will should preferably indicate the specific kinds of procedures not to be applied in case of the incapacity of the patient, for example,

[2] A persistent vegetative state results when the cerebellum, the part of the brain controlling thought and behavior, is destroyed but the brain stem and the thalamus, which control sleep cycles, breathing, and heart rate, are intact.

[3] Brain death results from irreversible brain damage. The loss of brain function is shown clinically by a flat reading on an encephalogram. In addition, there is no flow of blood to the brain as shown by any of several medical procedures (e.g., Doppler ultrasonography).

use of feeding tubes or respirators. In such cases the physician may comply and the courts will tend to support that decision. In *Cruzan v. Missouri* (1990), as noted, the Court ruled that a person had the right to refuse life-sustaining medical treatment.

Before treatment could be withheld from a comatose patient, however, a state could require "clear and convincing evidence" that the patient would not have wanted to live. The difficult cases are those where the will of the patient is not made clear or specific in writing, and members of the immediate family disagree on the matter. In general, according to state laws in the United States, in the event of the lack of a Living Will and disagreement of members of the patient's family, a spouse's testimony as to the expression of the patient's wishes, and the spouse's own wish, take precedence over those of other members of the family. Lacking this guidance, the decision will probably be made on the side of keeping the patient alive, however miniscule the chance of resuscitation.

In the celebrated Terri Schiavo case of Florida, the patient was in a persistent vegetative state and was kept alive on life-support equipment (i.e., a feeding tube for hydration and nutrition) for over 15 years (1990–2005). For the last 7 years of this period the husband and the parents had been battling over their difference as to whether the patient should be kept on life-support equipment any longer. A Living Will would have resolved the issue long before, but none was available and the courts had to depend on the testimony of members of the family. The husband argued that the feeding tube should be removed since his wife had made known to him her wish not to be maintained under such conditions and neurologists testified that there was no chance of resuscitation from her persistent vegetative state. The parents argued that their daughter would have wanted to live, that they would be willing to care for her, and that her constitutional right not to be subjected to cruel and unusual punishment (i.e., slow starvation until death) would be violated if her life-support was removed.

The state courts of Florida and the lower Federal courts (the U.S. Supreme Court having refused to hear the case) supported the husband's position even though the President, the U.S. Congress, the Governor of Florida, and the legislature of Florida all intervened, without legal precedent, to set aside the various courts' decisions and to void the husband's wishes. In this and many similar cases the medical, legal, and ethical issues have long been resolved.[4] From a broad bioethical view, the "conservatives" are committed to the prolongation of life because they believe that life has intrinsic merit, while the "liberals" place importance on the quality of life. This case would have had far reaching financial, legal, and bioethical implications if the conservative view had triumphed, but the court decisions and the husband finally prevailed. Terri Schiavo was allowed to die in March 2005.

[4]The resort to nonjudicial devices was widely condemned by legal and bioethical scholars as well as by the public as a political tactic by the party in power to win the support of religious conservatives.

Stem Cell Research and Bone Marrow and Organ Transplantation

Stem Cell Research

For most bioethicists, the issues in therapeutic cloning, including stem cell research, are far more complex than the issues in human reproductive cloning. In their view, as in the view of the public and political leaders, reproductive cloning is simply wrong.

Before discussing the issues pertaining to therapeuting cloning, I want to give some background information on stem cells. Stem cells are pluripotent, that is, they have the distinguishing feature that they are able to differentiate into a variety of types of body cells. Stem cells can be obtained from embryos and umbilical cord blood. These are the most efficient, dependable, and safe sources of stem cells. Some adult cells and fetal cells can also be used as stem cells, but they have a lesser ability to differentiate into target cells and they tend to migrate to the bodily area from which they were taken.

Researchers have been trying to "train" adult stem cells to work like embryonic stem cells and are finding new sources of adult stem cells, including fat cells, skin cells, and periodontal ligaments (i.e., the fibrous tendons that hold teeth in their sockets). Such adult stem cells are so-called induced pluripotent stem (IPS) cells. In 2006 S. Yamanaka of Japan reversed the process by which mouse skin cells were developed, creating the first IPS cells, i.e., stem cells developed without the use of embryos. Then in 2007 Yamanaka and J. Thomson of United States independently created the first human IPS cells. Such cells could be used to create any one of the body's 200 tissue types, and the techniques could eventually be used to replace tissues for patients using their own cells. The procedure would avoid the need for donors and the risk of rejection. Stem-cell transplantation has some serious risks, especially if the donors are mismatched; hence, it is preferable to secure stem cells from the individual being treated.

Therapeutic cloning in general and stem cell research in particular would make possible the replacement of defective or diseased genes, regeneration of organs, the growth of replacement organs, and ultimately a successful attack on many of the chronic degenerative diseases. As the reader may recall, this field has been labeled regenerative medicine. Already experiments have been conducted in which human fetal cells have been injected into rats that have experienced a stroke; the fetal cells migrated to the brains of the rats and reversed the effects of the stroke. Apparently the area affected puts out a message that draws the fetal cells to it. Research is under way on the use of fetal cells from the pancreas, heart, and brain with the goal of treating, respectively, diabetes, heart disease, and Alzheimer's disease and Parkinson's disease. Stem cells from human retinas may one day be used to cure eye diseases such as macular degeneration and retinitis pigmentosa, and even to restore vision in the blind.

Such developments could reduce the extent and degree of chronic illness among many older persons and could extend life expectancy. However, they would not

necessarily involve any substantial slowing of the rate of aging or an increase in maximum human life span. Moreover, stem cell technology is no panacea, it could take years to be a general tool in treating human disease, and its promise may not be realized fully. What works in mice may not work as well in men or women, or worse, induce disease. Because of the difficulty of procuring embryonic stem cells directly it could from individuals for their own therapy, it may be much easier to use adult stem cells from a patient's own bodies for their treatment.

Scientific, philosophic, and religious views. There is moral and religious opposition to embryonic stem cell research. Extracting stem cells from human embryos is interpreted by its opponents as resulting in the destruction of human life because the embryos are destroyed in the process of harvesting the stem cells. The basic premise of this position is that human life begins at conception because life with human cells begins at conception, and therefore that embryonic stem cell research involves the systematic and intentional termination of human life.

In evaluating this viewpoint and the opposing one supporting embryonic stem cell research, I find that the former view has a purely religious foundation linked with a questionable philosophical foundation. To explicate the opposing philosophical and scientific argument, we need to draw a distinction between life, viability, and personhood. An ovum or sperm is a living organism, but few would argue that such organisms are living human beings. An ovum or sperm cannot become a human being unless they are united as a fertilized egg. A fertilized egg is a living organism, but it cannot become a human being unless it is implanted in the wall of the uterus. Millions of pre-implantation embryos and millions of implanted embryos naturally die every year in women's wombs. Many implanted fertilized eggs die naturally as miscarriages, and a large share of these miscarriages are unknown to the would-be mothers.

Many fetuses in the second and third trimester are lost as stillbirths, or intermediate or late fetal deaths. Although these fetal losses are usually considered natural or unintentional, with appropriate prenatal care of the mother many of these fetuses could be saved for full-term uterogestation and birth. However, the fact that many of these losses are preventable does not change the more basic fact that there is tremendous natural loss of life in the progression toward birth, including fetal losses. It should be apparent that life may best be viewed as a process in which living organisms increasingly move toward a potential state of (human) personhood with consciousness and there are many natural failures, losses, and "dropouts" along the way.

It is not until almost the end of the second trimester that some of the fetuses can be judged viable, that is, having the capacity to exist independently of the mother, possibly with modern medical intervention. Even at viability, however, the fetus has not become a person, that is, a human being that can live independently of its mother and has integrated functions permitting its own survival. It may be said to become a person and attain personhood at the point of birth. Scholars have commonly reached a consensus that personhood requires such characteristics as consciousness, self-awareness, and the ability to direct one's attention and act purposively.

Most of the available embryos are stored as surplus in fertility clinics, awaiting disposition by their owners. The owners have the right to dispose of the embryos as they wish, even request that they be discarded. Are they "murderers," as some would dub them? Since the embryos are no longer required for the fertility needs of the clinic, they will eventually be discarded anyway. Hence, most stem cell researchers consider baseless the concern that the destruction of stored embryos would be tantamount to destroying human life.

The issue can also be structured in terms of the relative value of living things at different stages of development. Surely a living child is more valuable than a tray of 20 frozen embryos in a fertility clinic if a choice has to be made between them (for example, in case of fire or hurricane). Surely the mother of three living children carrying a fetus, the birth of which could kill her or seriously compromise her health, is more valuable than the fetus. Consistency in the logic of the anti-embryonic-stem-cell- believers would favor the tray of 20 frozen embryos over the child and the unborn fetus over the mother, or at least would equate them. We should place a greater value on the prospect of saving untold numbers of individuals with chronic degenerative illnesses as compared with saving frozen embryos waiting to be discarded in fertility clinics.[5]

In an executive order in 2001 President Bush banned the creation of new lines of embryonic stem cells with federal funds. Federal funding for stem cell research would be limited to the 60 embryonic stem cell lines or colonies then in existence. In a more recent evaluation, only 15 of these lines still exist and these 15 lines have become contaminated. There is no restriction on the use of adult cells or fetal cells, but experts maintain that embryonic stem cells hold greater promise for medical advances than adult and fetal stem cells. While it would seem that the development of induced pluripotent adult stem cells has "uncomplicated" the debate regarding the morality of using embryonic stem cells, they are viewed as a less natural means of regenerating defective and diseased body tissues and subject to greater risk of failure or disease. Researchers maintain that all scientific methods should be available to them for exploration without limiting them to one course or another.

Public opinion in the United States supports embryonic stem cell research. In 2004 60% of California voters supported a ballot initiative that would establish and finance a public institute to conduct embryonic stem cell research and the state appropriated several billion dollars for this purpose. The state of Maryland has moved in the same direction. Within days after his inauguration in 2009 President Obama issued an executive order lifting the ban on the use of federal funds for research on embryonic stem cells.

[5] A logical corollary of the anti-embryonic-stem-cell view is that the estimate of the U.S. official population should include every pregnancy and every stored embryo, and the growth of the population should be measured in terms of all conceptions as well as every fetal loss and death. Another practical aspect of this issue arises when a pregnant woman is killed by accident or homicide: Is there liability for the death of one person, two persons, or possibly more?

Bone Marrow Transplantation

Bone marrow is the source of red and white blood cells and bone marrow transplantation provides red and white blood cells for therapeutic uses. Matches are based on inherited tissue markers called human leucocyte antigens (HLA), some of which are very rare. Normally a sibling would be the best candidate to provide the healthy marrow that may be needed in cases of such malignancies as acute leukemia. If a sibling is not available or none proves suitable, the next best hope is a close relative and then a member of the same racial/ethnic group. However, in the United States the nonwhite races, Hispanics, and some other ethnic groups, e.g., Jews, are underrepresented in bone-marrow registries.

There is both a U.S. National Marrow Donor Program, the main U.S.-based registry that is federally funded, with about eight million donors, and an international collaborative registry of about six million donors called Bone Marrow Donors Worldwide. According to the U.S. National Marrow Donor Program, Caucasians in the year 2000 had an 88% chance of finding at least one unrelated donor. For racial and ethnic minorities, the numbers were much lower, ranging from 58% for blacks up to 83% for American Indians. Although Jews are classed with the Caucasians, their chances are believed to be lower than those mentioned. As a result of the Holocaust, which wiped out the families of millions of Ashkenazi (European) Jews, a match for an Ashkenazi Jew is extremely difficult to secure. In trying to locate donors, the federal government distinguishes in its recruitment program between the ethnic/racial affiliation of the prospective donors, but it does not subsidize efforts to recruit and test Jewish donors as it does other ethnic groups. This may be because Jewish donors are "lumped" with the Caucasians in the donor pool (*Pennsylvania Gazette*, Nov./Dec. 2003:48–52).

Another group for which it is extremely difficult to locate a blood marrow donor is persons of mixed race. Inasmuch as they constitute only a small proportion of the total, few eligible donors of mixed race appear in the registry; yet the number of these in the population is increasing rapidly. There are only about 250,000 mixed-race donors listed in the National Marrow Donor Program – only 3% of the total persons registered. Their parents, children, and siblings are typically unable to be their donors as well. As a result, the patients usually must resort to old conventional treatments. To an important extent, this applies to blacks as well, who are largely of mixed race.

Organ Transplantation

Organ transplantation presents numerous ethical issues concerning the donor, the recipient, the relation of donor and recipient, and the procurement of organs and tissues. Much of the problem arises from the limited supply of organs. First, I want to provide some background information regarding organ and tissue transplantation.

The number of organs available for transplantation falls far short of the demand. Further, the number of persons receiving vital organs and surviving are too few

to make a serious dent in the death rate. With the prospect of rapid aging of the population, the situation may improve or deteriorate, depending on the degree to which the supply of donor organs increases in relation to the demand. Kidney transplants, cornea grafts, bone grafts, and skin grafts are considered routine for certain conditions while heart, lung, liver, and pancreas (pancreatic islets) transplants are becoming fairly common. Kidney transplants are the most common whole organ transplant. The supply of organs for transplant and the way the organs are distributed affect not only the number of people who survive but also the characteristics of the survivors (U.S. GAO 2008).

Sources and supply of donors. The shortage in the supply of organs for transplant is illustrated with the following pair of numbers. In 2008 an estimated 79,900 persons were waiting for kidney transplants but only 16,517 kidney transplants were performed. Even the number of organs available from cadavers falls far short of the number expected in that it is less than the number of persons who claim in polls that they have donated organs. For this reason, transplant physicians have had to resort to organ transplants from living donors. For some organs, adult live transplants are still quite rare. This is so for liver transplants, for which only 2% of donated livers come from healthy relatives and friends of patients. Halving livers from cadavers, which could double the supply, may be necessary since the number of persons requiring liver transplants is expected to increase by several hundred percent in the next several years.[6]

Risks. Transplants of different types carry different risks. In 2007 95% of the patients receiving kidney transplants survived at least a year but only 88% of liver transplant patients did so. A liver transplant from a living donor is far more risky than a transplant from a cadaver. A liver transplant from a living donor is far more risky than a kidney transplant from a living donor. Transplantation has a relatively limited "warranty" period, during which time a transplanted organ such as a liver can become compromised. However, the majority of transplant recipients live 5 or more years. Cadaver transplants of kidneys last 5–7 years on average and living transplants last 15–20 years (Boyd 2007). Rejection of the transplant is the principal cause of failure of a transplant since it is difficult to find a total match of tissues. The closer the genetic relation of the donor and recipient the more likely the transplant will last. Hence. a living identical twin offers the best opportunity for success. For most transplants the recipient must take immunosuppressive drugs for the remainder of the life of the transplant.

With careful medical evaluations in selecting donors, kidney donors have very few complications. Some donors have complications and a small percent die. However, a death rate of one percent of donors is small compared to the certain death of a family member.

---

[6]Two other developments add in small degree to the availability of organs for transplant. One is the recent growth in the number of donors who are not related to, or known by, the patient. These altruistic donors donate an organ such as a kidney to others because it is "the right thing to do." In addition, there is a growing number of internet websites that act as brokers between patient and donor for a fee.

Cost and other factors. In addition to the scarcity of organs for transplant and the health risks associated with donating an organ or receiving one, the opportunity to obtain a necessary organ is limited by the great cost of the operation as well as the urgency of the transplant, the recipient's age, and the recipient's general health. For many insured patients who receive kidney transplants, the insurance covers the cost of the evaluation, donor surgery, patient surgery, and follow-up care. On the other hand, insurance may cover only a fraction of the cost for some patients or no part for uninsured patients. The great exception relates to Medicaid-eligible persons, who receive virtually complete care, including 100% of hospital and out-patient care and medications, with minimal co-payments. The most seriously disadvantaged are those who earn too much money to be Medicaid-eligible and either cannot afford to purchase good coverage or have very limited plans from their employers (i.e., high deductibles, large co-payments for daily hospital stays, etc.). For example, if a patient has limited prescription-coverage for the required immunosuppressive medications, which are extremely expensive, the hospital may not carry out the transplant, but a Medicaid patient will be accommodated.[7]

Ethical issues. The ethical issues relating to organ transplantation revolve around such questions as the following: Should the body and its parts be treated as a commodity that can be bought and sold, or should the sale of body organs continue to be banned? Since the supply of organs is limited, removing the prohibition would make more organs available. How can the limited supply be fairly distributed? How can coercion of donors be prevented? Are there situations where coercion of donor or recipient is right? Should prospective recipients who have abused their bodies through excessive alcohol, smoking, and poor eating habits receive transplants in their turn or passed over in favor of those whose need is based on less problematic causes? Given the cost, time, and effort involved in transplantation, should one distinguish need to receive a transplant on the basis of age and health of the recipient or probable years the transplant will be viable, or simply on whether or not the organ is needed for survival? Should a person whose transplant has failed be given a second organ? Since the costs of transplantation is high, in order to prevent discrimination on the basis of income and socioeconomic status, should the cost of all transplants be covered by public funds?

The general principles for dealing with these questions might be summarized in the following terms: Informed consent, generosity, autonomy, respect for the dignity and equality of human beings and their body parts, and avoidance of inflicting harm. Informed consent is a basic principle of organ donation; it calls for informing the

[7]Living-donor transplantation is more expensive than cadaver-transplantation. In the case of kidney transplantation, if there is a living donor at hand, the surgery can be expedited and years of dialysis can be avoided. The availability of a living donor obviates the wait for a cadaver and the greater risk of medical complications. After a 3-years break-even point, transplantation becomes a less expensive option than dialysis. Moreover, transplant recipients have fewer complications and fewer hospitalizations than dialysis patients. Transplantation is a far less expensive option than lifelong dialysis (Boyd 2007; U.S. GAO 2007).

donor and recipient of the risks of the surgery and the likelihood of success or failure. Since incompetent persons cannot give informed consent, they should not be used as donors. They may be coerced into being recipients if appropriate authority rules that it will preserve their lives. Being a donor is an act of benevolence that should not be coerced, even by family members.

Violating the integrity or totality of one's body is frowned on by Judeo-Christian doctrine but donating a part of the body like a kidney for benevolent reasons is not viewed as compromising this principle. Donating one's heart or maybe even an eye would do so. The principle of distributive justice, recognizing the dignity and equality of human beings, would seem to rule out making determinations of transplant recipients on the basis of socioeconomic status, health habits, or previous receipt of a transplant, but age and state of health or likely period of viability of the transplant may be appropriate for ranking the candidates. Trading in body parts is an objectionable practice and can become a tool for exploiting the poor. Some argue, however, that it would increase the supply of transplant organs and therefore have a positive effect on a transplant program. The U.S. National Organ Transplant Act of 1994 makes the sale of organs illegal and it is prohibited by the U.N. Declaration of Human Rights, but compensated donation is common in some parts of the world, whether legal or not.

Religious principles frown on having another child merely to provide a donor organ for a sick brother or sister, or on becoming pregnant and deliberately aborting the fetus merely to harvest an organ or tissue part for a sick member of the family. There should be no indication that an organ is harvested from a dying donor before the individual is completely dead as indicated by an encephalograph, and the transplant team should be completely independent of the dying patient's own team of health care providers.

Provision of Health Care

Mortality/Health Policy

It may reasonably be assumed that every government has an explicit policy of reducing death rates and improving the health of its population. If this is so, can we assume that governments and societies view health care as a right? Not really. Universal health of a population may be seen only as a theoretical goal, and in general governments have not implemented programs to assure that everyone receives the health care they need. Although universal health care is often promised by governments or nationally mandated, this promise cannot be fulfilled for practical reasons of limited resources and unlimited demand, and health care is not guaranteed as a right. Even if health care is promulgated as a fundamental human right, this is more a slogan than a reality. Rarely do governments implement programs to fulfill such expectations. Given the limitations of resources and human nature, actual policies regarding health programs typically make this "right" a "pipe dream."

Although health care is generally available for persons 65 years of age and over under the Medicare program in the United States – the system of public health insurance for the elderly – the amount of health care offered under this and other health insurance programs is closely monitored and quite limited. Most people with health insurance are underinsured and lack sufficient health care. For example, insurance coverage of mental health, dental health, and ocular health is insignificant in the face of the need, and insurance for medications is too costly for many. In addition, about 47 million persons in the United States are uninsured although many of these secure health care in emergency clinics. Many defeat the intentions of health policies and programs by failing to pursue good health practices or to take full advantage of such health care as they are offered. Limited income and education are major obstacles to realizing good health.

Relative Role of Government, Private Industry, and the Family

With limited governmental resources, the limited power of government over individual lives, and the limited knowledge of diagnosis, treatment, and cure of disease, health cannot be guaranteed as a right. Unlimited health care cannot be guaranteed as a right either. We have to ask whether government can and should be responsible for everyone's health even if universal health care is available. It is more reasonable to maintain that the government, the family, the individual, and the health establishment are jointly responsible for the health of each person.

Present knowledge of how good health is maintained imposes a major burden on individuals to take care of their own health, whether by eating nutritious meals and eating moderately, exercising regularly, avoiding smoking and other self-destructive practices, getting regular physical examinations, and so on. Hence, even where the government offers universal health care, taking care of one's health is a partnership in which the individual must actively participate.

Related questions concern the definition of a sickness or health condition, the types of health conditions covered under insurance plans, and underlying these definitions, the applications of the concepts of medical necessity and preexisting conditions. All government and private insurance plans set limits on the care that is allowed if the health system is to remain financially solvent. In the United States the federal government determines the health conditions covered by Medicare and the list of conditions is changed from time to time. In 2004 the federal government declared that obesity is a sickness. It may be a risk factor for sickness but is it a sickness? The ruling raises the question of where the line should be drawn between conditions that predispose individuals to serious diseases – so-called risk factors – but are not themselves diseases, and the illnesses caused by these conditions. Should the government and taxpayers pay for treatment of a condition which is essentially a risk factor and for which the individual is largely responsible? Use of tobacco products, poor diet, and lack of exercise represent lifestyle choices and risk factors that are even more difficult to monitor than obesity, but they too contribute to many serious illnesses.

Insurance programs employ the concept of medical necessity to exclude claims for cosmetic treatments and other reasons not viewed as requiring medical care. This principle is often invoked in claims for mental health conditions, particularly those that are arguably the less serious ones, such as headaches, anxiety states, and some phobias. They are also costly ones to cover because treatment may stretch out over long periods and they affect large numbers of persons at one time or another. In another thrust to control costs, insurance plans are anxious to exclude claims for "preexisting conditions." (Exclusion for preexisting conditions appears to be prohibited in the proposed new health reform legislation.) This problem is encountered, for example, by patients who have certain musculoskeletal and neurological conditions that occur intermittently, such as backaches and headaches, and a number of conditions that can become symptomatic some time after they are found clinically (e.g., atherosclerosis). In addition to the rejection of claims for these conditions, this practice inhibits patients from feeling free to change plans for fear of learning that some condition will be considered preexisting by a new plan.

Illustrations of Ethical Dilemmas and Opposing Interests

There are many cases where legal and technical considerations clash with ethical considerations. Suppose, for example, that a medical intervention does not prove to be cost-effective under analysis compared with cheaper interventions, but a patient demands that the intervention be carried out because it may possibly save his or her life. A difficult ethical question is raised in denying this treatment to persons who have a lethal condition in its advanced stages (i.e., with a low cure rate), for example, metastatic breast cancer. The argument that such individuals might have prevented this condition, had they regularly taken a mammogram will not "fly" because it cannot be proven that the mammogram would have detected it and the public empathy for the patients with such a condition is too strong. Another difficult ethical question is raised for the physician whose patient will die without an immediate transplant, although the patient has admitted that he or she is a "substance abuser." This practice would normally make him or her ineligible for a transplant.

An issue of rights vs. risks (plus obligations), or patient decision-making based on informed consent vs. government coercion, arises where an HIV-infected mother wants to breastfeed her infant child even though there is a risk that the breast milk is a carrier of the HIV virus and the child will also become infected. Should such the mother have the right to choose how to feed her child or should the state have the right to prevent her from increasing her child's risk to become ill? The decision in an Oregon court in 1999 ruled that an Oregon mother who was diagnosed as HIV-positive was not permitted to breastfeed her child.

The private sector provides most health care in the United States and the agents of this care have different incentives in their provision of care. Hospitals want their procedures to be cost-effective. These costs are, of course, offset by revenues and, as businesses, hospitals are motivated to make a large "profit." They are not motivated to employ simple procedures that preclude the need for other more costly procedures

now or in the future. Nor are the physicians or surgeons so motivated. On the other hand, insurers want to keep costs down and avoid costly procedures, but if the bulk of the cost of the more drastic interventions is to be paid by Medicare, they have little motivation to use cost-saving procedures. As a consequence, they are little interested in prevention. Government agencies are reluctant to try new procedures unless they can be offset by immediate savings, and this is not usually the case. The interests of patients, measured by access to the most health care of the best quality, their time, and their out-of-pocket expenses are often forgotten in this process.

In the general private marketplace, competition plays a major role in controlling prices and assuring quality products, but in the provision of health care, market forces do not function this way. In the "market" of health care, prices are out of control and the best care is not provided. It is evident that the provision of health care is unlike other services, and close oversight is essential if the appropriate amount and quality of health care are to be made available.

The internet has introduced a new vehicle for patients to express their views regarding the quality of care they are receiving. This right of free expression is protected by law but is obviously subject to abuse. Physicians are concerned that some of these evaluations may be malicious, baseless, or ill-advised. They argue that patients are not always in a good position to judge the quality of care they are receiving for such reasons as the following. Patients cannot properly judge the skill, knowledge, or judgment required to practice a medical specialty. Patients vary in their risks and hence should expect different outcomes. Health treatments take varying periods of time to work for different patients. Treatment is usually the result of a system effort and, since the system is now fragmented, patients may not know who has played what role in their treatment. Finally, patients themselves are active participants in this system and hence in their own care, and they may fail to perform their own role well. Physician reaction to internet evaluations may lead to their asking patients to sign away in advance of treatment their right to use the internet to voice their opinion of their physicians.

Race, Age, and Other Issues in Provision of Health Care

Persons of different nonwhite races and ethnic groups, women, children, and old persons in the United States have been the subject of selective treatment in the administration of health research and health care. As a result, they have figured in a variety of bioethical issues. The most egregious examples are the cases of abuse of blacks in medical research in past years. Some groups, particularly blacks and very old persons, have been affected by various discriminatory practices through "adverse impact," that is, indirectly because of other characteristics they tend to have.

Race issues. With the completion of the mapping of the human genome, there has been considerable interest in determining the percent of the human genome that distinguishes one race group from another. The availability of supercomputers in

combination with the human genome map makes it possible to visualize the entire genomic profile of an individual (with all three million DNA markers) and note the differences between the profiles of one individual and another. It has been established that the DNA profiles of any two individuals differ by only 0.1%. The technology has also been applied to look for patterns of DNA that differentiate groups of individuals, including the races. Criminologists in particular have been trying to see whether a tissue sample left at a crime scene could be used to identify the race of the offender. Efforts to identify segments of DNA that distinguish the races have not usually been successful. This has been interpreted to mean that all races are genetically the same. Quite clearly, however, the races do differ biologically since specific physical traits that distinguish the racial groups are inherited and are seen "loud and clear" when there is little or no racial interbreeding. Moreover, some diseases that have strong genetic origins affect the races in different degrees and in different ways. Consider the following example.

The results of some clinical trials testing the drug BiDil as a treatment for heart disease indicated that blacks were aided by the drug after it had been shown that whites were not aided by it. Hence, the FDA in June 2005 approved the use of this drug for blacks only. This became the first drug approved on a racial basis, although previously a treatment for sickle-cell anemia, a hereditary condition that almost exclusively affects blacks, had been approved for that condition. Are drug treatments that specifically target blacks a positive or negative development in treatment programs? Is this good medicine or simply a "reinscription" of race as a category in clinical medicine (Duster 2005). The answers to these questions may be debated. I would argue that failure to offer the drug to the most vulnerable group (e.g., blacks) because that group happens to be a racial group is medically unethical, as the FDA decision implied. This view is tenable even though, admittedly, categorization by race is considered by some as a controversial and "messy" classificatory tool.

Given the knowledge that the human genome is essentially the same for all racial groups, why are some groups deliberately overrepresented in clinical trials and treatment protocols for various diseases? The fact that some genetic diseases disproportionately affect certain racial and ethnic groups has led to the deliberate decision to include, even overrepresent, such groups in the trials. A mutation in a single gene, common in a racial group, could account for a particular genetic disease. If convincing information were not available that race and ethnic affiliation specifically "caused" some diseases, there would be little reason to insist that these groups be handled distinctively in the trials. The merits of selective representation in clinical trials have been demonstrated for women, not only for specifically female diseases but also for "sex-neutral" diseases as well.

Racial exploitation in medical research. The history of medical research in the United States is replete with instances of testing of blacks without informed consent, research on black subjects resulting in untreated illnesses, and neglect of research on conditions that affect mainly blacks (Washington 2007). More recently, amidst new charges of the emergence of "neoracial medicine," the research establishment has been criticized for insufficient inclusion of blacks in clinical trials, on the one

hand, and overrepresentation of blacks in clinical trials, on the other. The first charge is one of neglect of a group with distinctive health problems and the second is one of reifying a category that has no distinctive genetic identification.

It is desirable to seek adequate representation of blacks in clinical trials whether or not the incidence of a health condition is higher for blacks than for other races. The treatment may prove to be effective in different degrees for the races. Moreover, if blacks were omitted, it would be uncertain whether the treatment results were applicable to them. There are diseases that, whether for genetic or environmental reasons, occur almost entirely among one racial group, such as sickle-cell anemia and pellagra among blacks, and there are diseases that disproportionately affect blacks, such as hypertension and hepatitis B. It is ill-advised to deem adequate representation in clinical trials a reflection of the emergence of a neoracial medicine, especially if informed consent is scrupulously observed. The situation in the armed forces with respect to informed consent presents a special problem since enlistment in the military service subjects the members to a special body of law where informed consent is limited. In view of their past history, blacks have special reason to be concerned about the erosion of informed consent in military or civilian life.

Medical experimentation with human subjects has been imposed on vulnerable groups such as children, the poor, blacks, and institutionalized persons. Blacks especially have been subjected to a number of risky medical procedures, partly because they were readily accessible as victims, but also because experimentation on them would protect whites from often painful and terrible sicknesses. I cite a few of such experiments as illustrations. Perhaps the most abusive experiment on blacks was the Tuskegee Syphilis Study, in which 400 poor blacks in Macon County, Georgia, afflicted with syphilis, were observed over a period of 40 years beginning in 1932, to determine how the disease progressed and to measure its final effects in postmortem autopsies. In another study blacks in Florida were exposed to mosquitoes carrying yellow fever and other diseases in experiments conducted by the U.S. Army and the CIA in the early 1950s. In the 1970s black youths in New York were injected with Fenfluramine, an ingredient in the discontinued drug FenPhen, by medical researchers at Columbia University, who were investigating the genetic origins of violence.

A consequence of this experience with medical research has been to make many blacks regard the medical establishment with fear. Blacks have a pervasive distrust of the health care system, believing that it is generally inequitable. This is shown by an unwillingness to be examined by physicians, a reluctance to participate in clinical trials, and a low rate of organ donation. Blacks are more distrustful than whites of the equity of the organ donation system, are more likely to believe that the rich and well-known are favored in the receipt of donated organs and transplants, and are more likely than whites to support the provision of tangible benefits (e.g., money and funeral expenses) to donor families (Siminoff et al. 2006).

Health Conditions, Economic Status, and Age in Provision of Health Care

Health care is implicitly or explicitly rationed on the basis of the economic status of patients, their health condition, and their age. Generally patients who cannot afford to pay the out-of-pocket costs of a prescribed treatment or cannot buy the required drugs must do without them in a private health care system. Health insurance programs tend to rule against treatments for many conditions that are considered merely preventive, self-indulgent, not life-threatening, and not "medically necessary." They demand early discharge of patients from hospitals to reduce costs, usually forcing patients to take care of themselves at home or in nursing homes at their own expense. Even with health insurance, the economic status of patients has an impact on their access to good health care and the impact is adverse for the large segment of the population of modest means. This is particularly true in the case of mental and dental health. In sum, health care is rationed implicitly to a fair extent on the basis of economic status.

Age. Older people experience health-care discrimination, both explicit and implicit, in a variety of ways. It is a common view that the very old do not merit the same degree of health care as younger persons. Special limited treatment of very old persons is based partly on the premise that the life of an old person is less valuable because of his or her age and degree of illness and that the very old have already enjoyed a long life (Overall 2003). Other reasons given are that treating their conditions adds greatly to the public cost of health care while adding only a few short years, if not months, to their lives. Moreover, the cost of their treatment deprives others, especially children and youth, of the resources needed for the latter's health care. This view is justified further on the grounds that the individuals are often extremely ill, would not recover even if massive medical resources were applied, and are very near in age to the limits of human life. With the rapidly increasing number and share of persons at the older ages, particularly at the very advanced ages, more and more persons have multiple serious chronic conditions and only a few years to live.

These circumstances raise a difficult ethical question for society with regard to the treatment of such persons, particularly because human and financial resources and the supply of donor organs for transplants are limited. This situation poses an ethical dilemma for families as well. The patient and his or her family value the patient's life as much as others value the life of their younger family members. The general question is, should society allocate the considerable resources needed to prolong the life of persons of advanced age, possibly by a mere few months or years? For example, should society pay for a kidney transplant for a 93-year old person when hemodialysis may be less expensive for the first few years? Or for fitting a person of very advanced age with a prosthetic limb when he or she will have only brief use of this expensive prosthesis? Or for repairing a large abdominal aortic aneurysm of a nonagenarian when the chances of death on the operating table are only slightly less than no treatment at all?

As suggested, the charge of discrimination on neglecting these conditions would appear to be mitigated by the fact that the services have a limited value in terms of human life years and that the likely benefits for younger persons are greater. Where should the "age" line or "health-conditions" line be drawn when a zero-sum game is created by limited resources and limited benefits? Is it even appropriate to draw any lines at all? As we recall from the discussion earlier in this chapter, some do favor the restriction of health care at the advanced ages as a device for maximizing the output of the health care system (Callahan 1998), a health philosophy I call geriatric nihilism and one strongly opposed by the very old and their families.

The forms of discrimination against the elderly in health contexts are many and varied. For example, older persons are often treated less aggressively than younger persons, their conditions being considered less worthy of intensive or prolonged treatment (Overall 2003). This is partly because physicians generally have little training in geriatrics and partly because the multiple conditions of the elderly are difficult and time-consuming to treat and commonly incurable. Physicians may also view the condition as normal for advanced old age, even if unhealthy. When demand for health care is high and health resources are strained, patients who are extremely ill and are likely to die soon may be deemed justfiably unworthy of intensive treatment. Implicitly a form of triage is practiced. Cases of serious mental illness among the old are viewed in this way by many psychiatrists as are cases of cognitive impairment by many neurologists. The treatment needed by many old persons is long-term care and hospice care, and these are either not reimbursed or poorly covered by Medicare.

Older persons are not adequately represented in clinical trials designed to test the efficacy of procedures or medications. Yet it is well known that older persons respond differently to treatment protocols and medications than younger adults and often suffer from the very condition being researched. Most, recent announcements made by or on behalf of, NIH for participants in clinical trials set age limits excluding elderly persons. Where an age limit is not set, the participants must be "medically healthy" or "in good health," or not have "any sleep disorders," conditions not likely to apply to elderly persons. The reasons given for their exclusion from clinical trials are that it is too complex to design studies that include a sufficient number of elderly persons, that the outcome of these studies would be compromised by the elders' existing health conditions, and that they are too vulnerable to the treatment protocol.[8] It is true that most older persons have multiple disorders and are taking a variety of medications that could complicate

[8]In one solicitation for volunteer participants in a study of bone-marrow transfer, older persons were excluded on the ground that there is reduced bone marrow regeneration at higher ages and older persons are vulnerable to osteoporosis and other bone conditions. In another NIH research study on the antidepressive effects of sleep deprivation, the reason given for exclusion of the elderly is that the medication employed (yohimbene) causes a rapid increase in blood pressure and older persons do not tolerate this well because of the reduced elasticity of their vascular systems. There is age "discrimination" in the U.S. National Bone Marrow Program also; donor information remains on file only until the donor reaches age 60.

the research design if they were included or studied separately. However, given the importance of having direct evidence about elderly patients, the modification of the design of research protocols to include them needs to be explored further. Some new studies at NIA seem to be moving in this direction.[9]

Right Not to Receive Treatment

Various groups exercise a real or assumed right to reject medical care under certain circumstances. I enumerate five such categories of "refusers." First, some religious groups argue that their religion forbids the use of artificial means of treatment and that their condition will be healed naturally by divine intervention. A similar problem is posed by noninstitutionalized adults with mental illnesses who refuse treatment for physical ailments that may endanger their life, some out of simple fear of the procedure and others out of a delusionary concern that the procedure is being used as a weapon against them. Third are those who, defying ample research evidence, refuse some treatment or preventive procedure for themselves or other member of their family because of a misguided interpretation of the research findings. For example, a current public health problem is being created by persons who attribute autism in their children or the children of others to early childhood vaccinations and who, as a result, oppose such vaccinations, at the risk of having their children incur a variety of childhood infectious diseases and spread them to others.

A fourth category of refusers of appropriate health care is becoming increasingly common as the population ages. It consists of people who are mentally competent and choose to terminate a life-saving procedure such as kidney dialysis, because they no longer wish to live on a machine. The final category of refusers includes groups of persons who, having certain health conditions, have developed protective and supportive communities, have adapted to their special health state, and have chosen not to endure further accommodation to the larger community. An example is those deaf-mute persons who resist cochlear implants and turn against associates who have received them. In this age of "designer babies," some deaf parents may so embrace their deaf condition that they may select embryos likely to produce a child unable to hear. Another example is those persons with Asperger's Syndrome – a developmental disorder closely related to autism but less severe – who redefine their condition as normal and do not wish to be rehabilitated.

Should society have the right to intervene and force treatment in these cases? Some might pose the issue differently: Do individuals have the right to risk their

[9]The Baltimore Longitudinal Study of Aging (BLSA) announced early in 2009 that it is initiating a new research project intended to probe the genetic, environmental, social, and behavioral factors that contribute to the health and longevity of persons over age 80. It is called Insights into the Determinants of Exceptional Aging and Longevity (IDEAL). Persons over age 80 in superior health are being recruited to probe the reasons for their lengthy and healthy life. They will be compared with the regular participants in the BLSA.

lives or even commit suicide by refusing treatment for serious illness? These cases raise legal issues of "standing" (which ordinarily requires evidence of damage to oneself already sustained and precludes preemptive treatment to prevent damage), and of the state's power to make an adult a ward of the state when the individual is severely injuring himself or herself or is at risk of doing so. The courts have drawn a distinction between treatment that is required to save the person's life and other treatment that is less drastic; and treatment of mentally competent adults who are on life-support devices and others who are on such equipment but are not competent because of their mental condition.

Summary Note

Even if we assume that health care is a right, it cannot be a right without limits. Society's resources of money, manpower, and time, as well as the supply of transplant organs, are limited. The problem is exacerbated in the United States because this country does not have a system of universal health care, and the present system of widely dispersed private control assures an inefficient use of the health dollar and of an inadequate response to the health needs of the public. In the United States there is silent public acceptance of discrimination in the allocation of health care on the basis of income, and hence on the basis of education, but little public tolerance for discrimination on the basis of sex, race, and age per se as shown by the legal proscriptions against it in the workplace. Discrimination on the basis of income and education, however, has a disparate impact, resulting in discrimination on the basis of race and recent immigrant status. Moreover, discrimination on the basis of age and health condition is still widely practiced in the diagnosis and treatment of people of advanced ages and with various health conditions.

Society is ambivalent in its views regarding the rights and obligations of persons in health matters. It generally favors freedom of choice for adults in health matters, especially pertaining to themselves, whereas children's rights and parent's rights over their children are limited. Society draws the line at active attempts at suicide, but endorses the right of the individual to choose passive suicide by refusing life-support equipment.

Genetic Testing and Gene Therapy

New developments in genetics and its relation to health, particularly developments in genetic testing and gene therapy, suggest several areas of ethical concern. The list includes such issues as privacy in the handling of genetic records, discrimination in insurance and employment on the basis of the results of genetic tests, inequality in the availability of genetic testing on the basis of income, adherence to the requirement of informed consent, and minimization of the risks to patients undergoing gene therapy.

Genetic Testing

Hundreds of genetic tests are now available. They tend to be very expensive and, like many physical tests, can produce false positives and false negatives. Choices have to be made by individuals, physicians, and health insurance plans as to the ones to be employed by health providers and reimbursed by insurance plans. Some tests are more justified than others. If a woman has previously had a child with a genetic disorder (e.g., Down's syndrome) and is pregnant again, testing the condition of the fetus, either by amniocentesis or chromosome testing, is clearly justified. However, testing just for general screening is not justified and should be confined to cases where serious illnesses are suspected. Public opinion polls in the United States wholeheartedly support prenatal genetic testing for serious illnesses, but only mildly support testing for a personal preference such as sex selection.

The sequencing of the human genome opens up considerable possibilities for aiding individuals in learning about their genetic health and in tailoring treatments and medications to individual health histories. To sequence an individual's entire genome, however, costs a prohibitive amount today, perhaps a half million to a million dollars. Those who can afford genetic testing will be able to secure the special designer drugs needed to treat a problem; others will have to use the conventional drugs or do without medication entirely. A scanning of the genome for an individual may show, for example, that the person carries the gene for macular degeneration. This information may be used to treat the person with gene therapy at an early date. If, however, the tests are costly and are limited to those who can afford them, there are the same risks of discrimination on the basis of income as in other health matters, with the result that there will be an intensification of the existing differences in health between various segments in our population.

Resistance to genetic testing. Individuals may fear genetic testing, even though it may improve their prospects for longevity, because of fear of discrimination by employers and insurance companies. The genomic code can now be used to aid persons in changing their lifestyle and thereby improving their health, but persons may not want to take advantage of this opportunity if penalties could follow. If insurance companies can exclude them from coverage or limit their coverage, or employers can refuse to hire, retain, or promote them because of the results of genetic tests, they may decide not to take the tests. Privacy of medical records is seen as insufficient to protect the patients who secure genetic testing. As a result, legislation prohibiting use of this information in insurance underwriting or employment has been necessary.

The Heath Insurance Portability and Accountability Act (HIPAA) passed under the second Bush administration prohibits group health insurers from excluding presymptomatic persons from coverage based on results of genetic tests, and over 40 states have enacted laws relating to genetic discrimination. HIPAA does not apply to long-term care insurance or life insurance and does not deal with genetic discrimination in the workplace. Accordingly, in 2008 Congress passed the Genetic Information Nondiscrimination Act, making it illegal to discriminate in matters of

health insurance and employment on the basis of genetic information. The law prohibits group health insurance plans from basing determinations about premiums or contributions on an individual's genetic information. These companies may not request, require, or buy the results of genetic tests and they are prohibited from disclosing genetic information. Employers are bound by these same restrictions and cannot use genetic information in decisions regarding hiring and firing, or in setting conditions of employment. The law was to be implemented with a delay of 1 year to a year and a half.

Gene Therapy

Gene therapy for various serious illnesses in humans is now being tested in various laboratories. Experiments testing the safety and efficacy of gene therapy for use in osteogenesis, amniogenesis, and other forms of regenerative medicine are under way.[10] Since the procedure is still experimental, there are concerns about unexpected and undesirable side effects, such as the triggering of the patient's immune response, the mutation and resulting virulence of the viral vectors (i.e., the vehicle for delivering therapeutic genes to the patients' cells), and the transfer of altered genes to succeeding generations. Some genes increase the risk of a disease even if they do not "cause" it. There are also concerns as to how the manufacture of the viral vectors – once the safety and efficacy of this type of procedure have been demonstrated – can be expedited and the demand for large quantities of the product can be filled, given that the continuation of existing medications taken daily brings in far more money to the pharmaceutical firms. In addition, there are the usual concerns about informed consent of the patients, care in the development program to assure the validity of the results, and care in the design and interpretation of the clinical trials.

Gene therapy is designed to correct genetic disorders by injecting a normal copy of a gene into cells that contain the harmful form. Canine models are mainly being used to evaluate the safety and efficacy of gene therapy as a potential treatment for humans that have similar diseases. Researchers try to locate homologous regions in the human genome that can cause diseases in humans similar to those found in dogs, for which gene therapy has succeeded. Specifically, the goal is to locate the specific mutation involved in a specific disease of dogs and to link it to a corresponding chromosomal region in persons with the human disease equivalent. Then, the goal is to find the biochemical basis by which specific mutations produce disease. Gene therapy on dogs is now being tested for a number of diseases, particularly eye diseases, at a number of medical centers. NHLBI (NIH) has a research protocol to test the efficacy of gene therapy for angiogenesis in peripheral vascular disease. The selected patient population consists of persons who have advanced peripheral vascular disease with intermittent claudication and whose mobility has

[10]Osteogenesis is the development of new bone structure and amniogenesis is the development of new blood vessels.

not been sufficiently improved by revascularization (i.e., catheterization with laser and/or balloon angioplasty) of their femoral and associated arteries, or for whom revascularization is not indicated.[11]

Risks of gene therapy. Before the geneticist can proceed to the use of gene therapy on persons with a disease, he or she must conduct sufficient animal studies to make sure that the gene therapy protocol is safe. At this time gene therapy is in its infancy and so it is still a tentative and risky procedure and, even if the risk is limited, its efficacy and safety in humans have yet to be demonstrated. It is with respect to such a procedure that informed consent of the patients is essential. For example, to evaluate the efficacy of the procedures in the NIH protocol mentioned above, the subjects will be asked to give their informed consent to engage in a variety of diagnostic tests and therapeutic procedures, such as an MRA (i.e., magnetic resonance angiogram) of the peripheral vascular system (to determine the exact location and extent of the stenosis or blockage), a treadmill stress test (for risk classification purposes), and balloon angioplasty of the blocked peripheral arteries (to eliminate from the protocol those treatable by this and other more conventional procedures). If the balloon angioplasty fails or is not feasible, the patient may then opt to proceed with gene therapy for the treatment of his or her peripheral vascular disease.

Even an exercise stress test on a treadmill and balloon angioplasty carry risks, and informed consent requires the approval of the patient to take these risks. Informed consent at each step in the whole process is essential to the integrity of the research. It was partly on the grounds that the consent form and process "did not disclose all anticipated toxicities" to Jesse Gelsinger that the Gelsinger family successfully sued the University of Pennsylvania's Institute for Human Gene Therapy. Gelsinger died in 1999 while participating in a gene therapy study at this institute for treatment of a hereditary liver disorder. His death came a few days after being injected with a modified cold virus designed to carry corrective genes to his liver.

Preimplantation Genetic Diagnosis

Preimplantation genetic diagnosis (PGD) is offered as an alternative to prenatal diagnosis and induced termination of pregnancies that are affected by adverse health conditions. There are two main applications of this procedure but other applications have been considered. In the first application HLA (human leukocyte antigen) typing is used to match a future child as a compatible donor of hematopoietic stem cells to a sick sibling who is in need of a stem cell transplant (Devolder 2005). In this

[11]Peripheral vascular disease refers to atherosclerosis in the arteries of the arms and legs, particularly the legs, and intermittent claudication refers to pain in the calf on walking that results usually from blockage of the superficial femoral artery in the thigh. Revascularization encompasses the interventional procedures by which a catheter is used to enter and examine the arteries of the legs and arms and reduce the blockage by applying a laser and/or balloon angioplasty of the arteries.

case a couple has a sick child and chooses PGD to select embryos in the hope of conceiving an HLA identical donor sibling. Unlike the usual medical model, PGD here is not implemented for the health of the particular potential child, but for the health of a second party within a family, a sibling.

Ethical arguments have been made for and against application of PGD for selecting embryos in the hope of conceiving an HLA identical donor. One general ethical issue is whether PGD is more acceptable than prenatal diagnosis and termination of pregnancies for genetic and congenital disorders. The argument is made that the donor child is merely created as an instrument to be exploited rather than for its own sake. However, conceiving a child to save a child may be considered a morally defensible practice if the operation that the future child will undergo would be acceptably performed on any sibling (Pennings et al. 2002). Other questions may be asked. Does the procedure pose risks for the prospective child? Are there limits that should be placed on what may be done to the donor child? Is a stem cell transplant in a PGD procedure acceptable also to save the life of other members of a family, not just a sibling? Would it be acceptable to create such a human-leukocyte-matched child as an insurance policy for a family and even establish a bank of HLA-matched embryos?

PGD has also been used to test embryos before implantation in the uterus so as to prevent parents' passing on serious genetic diseases to their children or their children's children. For example, it has been used to determine whether an embryo in an IVF procedure is a carrier of a recessive condition and should be replaced by a noncarrier embryo in creating a prospective child. For this purpose PGD would employ sex selection in order to prevent the birth of healthy female carriers of X-linked recessive disorders who are at high risk of conceiving affected sons (de Wert 2005). While sex selection for non-medical reasons is widely viewed as an undesirable practice both from the view of ethics and society, employing PGD as a means of sex selection for medical reasons may be considered acceptable, particularly when applied on a case-by-case basis. In this second application of PGD, note that the health benefit is transgenerational and is implemented for a "third party," and therefore differs again from the usual medical model (de Wert 2005).

Finally, in closing this chapter, it is of interest to note that preimplantation genetic diagnosis is also being used for a non-medical reason by a small percentage of those seeking this procedure of embryo selection. These prospective parents are seeking to secure some particular physical or mental trait in their offspring; that is, they are seeking to design their babies in some degree beyond choosing their sex and avoiding transfer of a genetic disease. They may seek to choose their baby's eye, skin, or hair color, or seek a given height or superior intelligence. In the United States there are no legal restrictions preventing private fertility clinics from offering a list of such traits in the offspring of their clients or from implanting a woman with numerous embryos. These services present evident ethical dilemmas, such as that some group might plan to mould the perfect race or that selection of traits on an broad scale could reduce the variability in the gene pool.

References and Suggested Readings

Fertility

Kaplan, A. (2004). *Lecture*. Washington, DC: University of Pennsylvania Alumni Club.

Kolker, A., & Burke, B. M. (1998). *Prenatal testing: A sociological perspective. With a new afterword*. Westport, CT: Greenwood Publishing.

McGee, G. (2003). *Beyond genetics: Putting the power of DNA to work in your life*. New York: Morrow.

National Collaborating Center for Women's and Children's Health. (2004). *Fertility: assessment and treatment for people with fertility problems* (Clinical guideline. Commissioned by the National Institute for Clinical Excellence). London: RCOG Press.

President's Council on Bioethics. (2003, December). The President's Council on Bioethics: Choosing sex of children. *Population and Development Review, 29*(4), 751–760. Excerpted from President's Council on Bioethics, *Beyond therapy: Biotechnology and the pursuit of happiness*, 2003.

Extension of Human Life

Alliance for Aging Research. (2004). *The fight over the future*. Washington, DC: SageCrossroads.

Barry, R. L. (1991). *Set no limits: A rebuttal to Daniel Callahan's proposal to limit health care for the elderly*. Chicago: University of Illinois Press.

Callahan, D. (1995a). *What kind of life: The limits of medical progress*. Washington, DC: Georgetown University Press.

Callahan, D. (1995b). *Setting limits: Medical goals in an aging society, with a "Response to My Critics*." Washington, DC: Georgetown University Press.

Callahan, D. (1998). *False hopes: Why America's quest for perfect health is a recipe for failure*. New York: Simon and Schuster.

Callahan, D. (2003). *What price better health? Hazards of the research imperative*. Berkeley, CA: University of California Press.

Overall, C. (2003). *Aging, death, and human longevity: A philosophical inquiry*. Berkeley, CA: University of California Press.

Right to Die/Obligation to Live

Glick, H. R. (1992). *The right to die: Policy innovation and its consequences*. New York: Columbia University Press.

Glick, H. R. (1994). The impact of permissive judicial policies: The U.S. supreme court and the right to die. *The Political Research Quarterly, 47*(1), 207–222.

Glick, H. R., Cowart, M. E., & Smith, J. D. (1996). Implementation and impact of the patient self-determination act: Preliminary survey and proposals for change. *Journal of Applied Gerontology, 15*(1), 38–56.

Hoefler, J. M. (1994). *Deathright: Culture, medicine, politics, and the right to die*. Boulder, CO: Westview.

Genetic Testing, Gene Therapy, and Organ Transplantation

Boyd, C. S. (2007). *Private correspondence, C.S.Boyd*, Thomas Jefferson University, Medical Center, to author.

de Wert, G. (2005). Preimplantation genetic diagnosis: The ethics of intermediate cases. *Human Reproduction*, *20*(12), 3261–3266.

Devolder, K. (2005). Preimplantation HLA typing: Having children to save our loved ones. *Journal of Medical Ethics*, *31*, 582–586.

Pennings, G., Schots, R., & Liebaers, I. (2002). Ethical considerations on preimplantation genetic diagnosis for HLA typing to match a future child as a donor of haematopoietic stem cells to a sibling. *Human Reproduction, 17*(3), 534–538.

U.S. Government Accountability Office. (2007). *End-Stage renal disease: Characteristics of kidney transplant recipients, frequency of transplant failures, and cost to medicare* (GAO-07-1117). Washington, DC: U.S. Government Accountability Office.

U.S. Government Accountability Office. (2008). *Organ transplant programs: Federal agencies have acted to improve oversight, but implementation issues remain* (GAO-08-412). Washington, DC: U.S. Government Accountability Office.

University of Pennsylvania. (2003, November/December). *Pennsylvania Gazette* (pp. 48–52). Philadelphia: University of Pennsylvania.

Claims of Race Discrimination

Duster, T. (2005, February). Race and reification in science. *Science*, 307, 1051–1052.

Siminoff, L. A., Burant, C. J., & Ibrahim, S. A. (2006). Racial disparities in preferences and perceptions regarding organ donation. *Journal General Internal Medicine, 21*, 995–1000.

Washington, H. (2007). *Medical apartheid*. New York: Doubleday.

Appendixes

Appendix 1: Selected List of Journals in the English Language for the Study of Health Demography

American Journal of Epidemiology
American Journal of Public Health
Biogerontology
Bulletin of the World Health Organization
Demography
Demographic Research
European Journal of Ageing
European Journal of Epidemiology
European Journal of Population
Experimental Biology
Family Planning Perspectives
Genus
Geriatrics
Global Public Health
Human Biology
International Journal of Epidemiology
Journal of Aging and Health
Journal of the American Medical Association
Journal of the American Public Health Association
Journal of Clinical Epidemiology
Journal of Epidemiology and Community Health
Journal of Experimental Gerontology
Journal of Health and Social Behavior
Journal of Population Research
Journal of Social Science and Medicine
Journal of Theoretical Biology
Journal of Woman and Aging
Journals of Gerontology

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4, © Springer Science+Business Media B.V. 2012

Series A: Biological Sciences
Medical Sciences
Series B: Psychological Sciences
Social Sciences

Lancet
Milbank Quarterly
Nature
New England Journal of Medicine
Popnet Europe
Population (English Edition; INED, Paris)
Population and Development Review
Population Research and Policy Review
Population Studies
Rejuvenation Research
Research on Aging
Science
Scientific American
Social Biology
Social Science and Medicine
Studies in Family Planning
Vienna Yearbook of Population Research

General Internet Sources of Data on Health Demography

Listings 1–4 were prepared by Zuali H. Malsawma of the Population Reference Bureau and published in *Population Today 30*(8), Nov.–Dec. 2002.

1. Population, health, and nutrition (DOLPHN) database (USAID), *www.phnip.com/dolphn.* DOLPHN provides data on demographic and health trends for use in managing the global health programs of the U.S. Agency for International Development (USAID). Indicators are given for such categories as child survival, demographic characteristics, family planning, HIV/AIDS prevention, infectious disease control, maternal health, and socioeconomic status.
2. Health InterNetwork, *www.healthinternetwork.org*. The United Nations and the World Health Organization sponsor access to major journals in biomedical and related social sciences.
3. Reproductive Health Gateway, *www.rhgateway.org*. This is an information portal that indexes over 65 international websites providing material on reproductive health. Topics cover adolescent health, family planning, HIV/AIDS, maternal health, and training.
4. YouthNet (Family Health International), *www.fhi.org/en/youth/youthnet/ynetindex.html.* YouthNet is a global program sponsored by USAID to improve

reproductive health and prevent the spread of HIV/AIDS among persons aged 10 to 24. It features information on family life and sex education.

5. Population and Health InfoShare, *www.phishare.org*. InfoShare is an electronic library of documents supplied by member organizations working in reproductive and child health, HIV/AIDS, and population. It provides a platform for sharing and accessing research findings relevant to the Less Developed Countries and enables research groups in these countries to reach a global audience. It is supported by a grant from the Bill and Melinda Gates Foundation.

Appendix 2: Review of General Analytic Devices and Special Survey Methods

Several analytic techniques that are widely used in research on health demography are summarized in this appendix. They have generally been briefly discussed in this book earlier. The techniques described here are ecological analysis, cohort analysis, multivariate logistic regression, use of proportional hazards models, grade-of-membership analysis, and meta-analysis. I discuss also some special methods of collecting sensitive data and the survey method called multiplicity, or network survey design. Some other general techniques that appear from time to time in health studies – odds ratios and propensity scores – are also described. None of these methods are confined to health research but they are part of the armamentarium of the health researcher. This discussion is not intended to provide working instructions as to how to apply the methods described. It is intended to familiarize readers with their general nature and to direct them to sources of further information.

General Analytic Devices

Ecological Analysis

Numerous epidemiological studies have used ecological analysis as the method of demonstrating a relation between a health characteristic and certain explanatory variables. In ecological analysis the units of observation are small geographic areas composing the total area under consideration, such as counties, health districts, and census tracts, rather than individuals. The correlation between a health characteristic and the explanatory variables, calculated for these small geographic units, is typically interpreted to represent the correlation between the variables for individuals in the broader area. The “area” method, rather than the “individual” method, may be employed both for convenience and for necessity; the data to be managed may be vastly reduced by collapsing them in this way and/or the required data may be unavailable for individual respondents.

Because of the extension of the analytic conclusions from areas to individuals, the validity of ecological correlation has been questioned. It can be demonstrated, in fact, that ecological correlation can give an erroneous view of the relationship between variables. To the extent that the constituent areas are small, numerous, and internally homogenous with respect to the relevant health characteristics – that is, the within-area variation is small compared to the between-area variation – the extension of the analytic conclusions from areas to individuals should be more valid. Further improvement in overcoming the limitations of ecological correlation could be achieved by weighting the small area data according to the size of the population in these areas.

Cohort Analysis

The reader has encountered several instances of the application of cohort analysis in this book, such as the analysis of data obtained from longitudinal surveys and the use of generation life tables. The purpose of this section is to describe it as a method for general use in studies of health phenomena. Cohort analysis has been widely applied in health studies to reflect the changes over the life course or some part of it in the health conditions of the population and to link these health conditions with changes in the life styles of individuals, the environmental situation, and the demographic, social, and economic characteristics of the population. A special purpose has been to obviate the possible misinterpretation of the role of the explanatory factors in health outcomes that could result from cross-sectional analysis. As we saw, the age-related changes in the prevalence or incidence of a health condition over the age scale for a birth cohort may not be reflected accurately in the record of the health condition over the age scale in a particular year or short span of years. The cross-sectional data are composed of numerous birth cohorts and therefore confound any interpretation of how health conditions or other characteristics change over the actual life course.

Cohorts differ in their health experiences and potential for longevity. These differences result from the joint influences of genetic, environmental, and stochastic factors acting differently on each cohort. Environmental factors broadly conceived, including lifestyle, socioeconomic status, access to health care, and the physical environment, modify the health situation of a cohort greatly as it ages. These influences modify the expression of the genetic influences affecting the cohort and its initial demographic characteristics. Stochastic influences, or chance events, further influence the role of the genetic and environmental forces that affect the cohort and contribute further to the differences in the health situation among cohorts over the life course.

For example, there may be pronounced cohort differences in the prevalence of specific chronic diseases such as heart disease or lung cancer, particularly over the long term. In the United States, cohorts born in the 1920s and 1930s have high levels of risk for lung and stomach cancer as a result of high levels of smoking. Cohorts born in later decades have a lesser risk for incurring these conditions as a result of decreasing levels of cigarette use.

Cohort analysis also allows for the shifting influence of unobserved heterogeneity of risk and exposure in each cohort. Each cohort includes individuals of differing health and survival potential, so that the composition of the cohort with respect to its health status changes as the cohort ages and members exit from it through death. Weaker individuals drop out of the cohort in its earlier years while the stronger members remain, with the result that the "average" health status and mortality potential of the cohort continually change.

Cohort studies are not immune from the same general types of errors as all survey data, namely nonreporting and misreporting. Some survey questions are not answered and many more are answered inaccurately. The frequency of a particular event or activity may be overreported or underreported, with resulting bias in the tabulated data. Some health conditions carry a stigma and may be underreported. Cohort studies also suffer from dropouts of participants, apart from dropouts due to death. Individuals drop out of a cohort study for various reasons, including sickness, failing interest, change of location, and institutionalization, and these changes may introduce various biases in the data that remain to be considered in the analysis.

The complex problem relating to the decomposition of differences between the values of a health indicator at different dates or at different ages into the age effect, the cohort effect, and the period effect are discussed in Chap. 5.

Multivariate Logistic Regression

A common method of identifying variables associated with health outcomes is to apply regression analysis, that is, to statistically measure the relation of the disease outcome and one or more lifestyle, behavioral, and/or environmental factors. Regression analysis is much less expensive and time-consuming than randomized clinical trials and can provide valuable practical information for health applications.

In regression analysis the dependent variable may be either an unrestricted absolute variable, or a polytomous (usually dichotomous) variable, that may be related to one or more independent variables. In health studies employing regression analysis the dependent variable is often expressed as a dichotomous variable, that is, as the probability of the occurrence of an event. If we let P(t) be the probability that an individual experiences an event at time t, given that the individual is at risk of experiencing the event at time t, and if we let x_1 and x_2 be two explanatory variables, in a first approximation we can express $P(t)$ as a linear function of the explanatory variables,

$$\mathrm{P(t)} = \mathrm{a} + \mathrm{b_1 x_1} + \mathrm{b_2 x_2} + \mathrm{e} \tag{A.1}$$

For example, $P(t)$ could be the probability of entering a hospital at time $t = 1, 2, 3$, etc., x_1 could be age, and x_2 could be the number of chronic health conditions among a specified list of conditions. This equation has the limitation that $P(t)$ can

in fact take on values only between 0 and 1, while the equation can produce any real number for $P(t)$. To deal with this problem, we can make a logit transformation of $P(t)$,

$$\ln[\mathrm{P(t)} \div (1 - \mathrm{P(t)})] = \mathrm{a} + \mathrm{b_1 x_1} + \mathrm{b_2 x_2} + \mathrm{e} \tag{A.2}$$

A logit transformation expresses the natural logarithm of the ratio of a proportion to the complement of the proportion. Since $P(t)$ varies between 0 and 1, logit $P(t)$ varies between minus infinity and plus infinity. The logit expresses an odds ratio in terms of the natural logarithms rather than the natural numbers.

In the multivariate logistic regression equation A.2, the natural logarithm of the odds ratio of the dependent variable is "predicted" by a linear function of the independent variables. The regression coefficients b_1 and b_2 give the change in the logit, or "log odds," for each unit increase in x_1 and x_2, respectively, holding the other factor constant. The logistic equation can be generalized to include k explanatory variables:

$$\ln[\mathrm{P(t)} \div (1 - \mathrm{P(t)})] = \mathrm{a} + \mathrm{b_1 x_1} + \mathrm{b_2 x_2} + \cdots \mathrm{b_k x_k} + \mathrm{e} \tag{A.3a}$$

or in exponentiated form:

$$\mathrm{P(t)} \div 1 - \mathrm{P(t)} = \exp(\mathrm{a} + \mathrm{b_1 x_1} + \mathrm{b_2 x_2} + \cdots \mathrm{b_k x_k} + \mathrm{e}) \tag{A.3b}$$

Computer software for logistic regression is available from several companies. The SAS procedure is called LOGISTIC.

A principal concern in applying regression analysis is the interpretation of the results as merely an association or a causal relation. It is difficult to prove a causal relation even though independent and dependent variables may be significantly associated. The association may have arisen by chance or the variables may each be associated with a third variable, not encompassed in the analysis. Variation may arise from the effect of numerous conditions not explicitly allowed for in the analysis. For example, people who eat leafy vegetables may have better health, not because they eat leafy vegetables because they are more health-conscious and exercise more. People who live near many power lines may have higher cancer rates, not because of the electromagnetism emitted from the power lines but because they live in areas of dense traffic, or are poor and do not eat a healthy diet.

Proportional Hazards Models

Proportional hazards models are ways of applying regression analysis to time-dependent variables, that is, of estimating the effects of different covariates on the time-to-failure of a system or part of a system, or time to incurring a chronic disease or time to death. Many mathematical functions have been applied to describe the age variations of chronic disease and the relation between a disease and other

time-dependent, variables. The three most commonly used multivariate models are the exponential model, the Gompertz model, and the Weibull model, according to Allison (1984).

We define $P(t, t + s)$ as the probability that an individual experiences an event in the interval from t to $t + s$, given that the individual is at risk of experiencing the event at time t. This is the discrete time hazard rate. Alternatively, $h(t)$ is the probability that an individual experiences an event at time (t), given that the individual is at risk of experiencing the event at time t. This is the continuous time hazard rate. In relating the hazard rate to one or more explanatory variables, it is conventional to calculate the natural logarithm of the hazard rate.

In the exponential function, with two explanatory variables, the natural logarithm of $h(t)$ is set equal to a linear function of the explanatory variables,

$$\ln \mathrm{h(t)} = \mathrm{a} + \mathrm{b_1 x_1} + \mathrm{b_2 x_2} \tag{A.4}$$

where a, b_1, and b_2 are constants to be estimated. Such a model is useful for relating health characteristics to explanatory variables at some particular instant in time. It is unrealistic, however, for analyzing morbidity over time or over the age scale because in fact the hazard rate changes with time and age.

A function that allows the hazard rate to change with time and age is the Gompertz function. The regression equation is

$$\ln \mathrm{h(t)} = \mathrm{a} + \mathrm{b_1 x_1} + \mathrm{b_2 x_2} + \mathrm{ct} \tag{A.5}$$

where c is a constant that may be either positive or negative. If the logarithm of the hazard rate increases or decreases linearly with the logarithm of time (or age), we can express this relation in a regression equation based on the Weibull model:

$$\ln \mathrm{h(t)} = \mathrm{a} + \mathrm{b_1 x_1} + \mathrm{b_2 x_2} + \mathrm{c} \ln \mathrm{t} \tag{A.6}$$

where c is a constant greater than -1. The exponential model is a special case of the Gompertz and Weibull models. These models are solved by maximum likelihood procedures.

As mentioned, the exponential, Gompertz, and Weibull models are forms of proportional hazard models. The proportional hazards model describing two time-constant variables, may be written as

$$\ln \mathrm{h(t)} = \mathrm{a(t)} + \mathrm{b_1 x_1} + \mathrm{b_2 x_2} \tag{A.7}$$

where $h(t)$ is the hazard rate and a(t) is any function of time. This model is called a proportional hazards model because, for any two individuals at any particular time, the ratio of their hazard rates is a constant, although it may vary for the explanatory variables. That is,

$$\mathrm{h(t)_i} \div \mathrm{h(t)_j} = \mathrm{c} \tag{A.8}$$

where i and j refer to different individuals and c does not depend on time but may depend on the explanatory variable. Although the hazard rates are no longer proportional when time-varying explanatory variables are introduced, the model is still called proportional hazards.

Cox (1972) proposed the method of partial likelihood as a way of solving the model in (A.7). Partial likelihood solves for the coefficients b_1 and b_2, discarding the information about the function a(t) and treating the information about b_1 and b_2 as if the model were an ordinary likelihood function. The Cox method is widely applied in estimating regression models with continuous time data. Software for computer application of survival analysis is available from SAS. The SAS procedures LIFEREG, PHREG, and LIFETEST can be employed to solve for the constants in the various proportional hazards models.

Grade-of-Membership Analysis

Grade-of-membership (GoM) procedures are multivariate classification techniques based on "fuzzy" data sets. Fuzzy data sets are groupings of data based on criteria that are partly ambiguous or incomplete. An example of a fuzzy classification is the classification of persons according to race in the U. S. census. Many individuals are of mixed race, but they choose to assign themselves in a self-enumeration canvas to a single race category or, by default, are assigned to a single race category by the enumerator. Similarly, disability is a characteristic of individuals that is associated with or composed of several fuzzy data sets. In grade-of-membership analysis the individual is "split up" into several classes, with weights or scores, based on the "components" of the fuzzy data set, *e.g.*, races in the mixture or types of disability. The scores for each individual range from 0 to 1 and sum to 1. Such scores or proportions are the grades of membership. A grade of membership is, therefore, a measure of the degree to which an individual is a member of a particular fuzzy set. The grade-of-membership procedure determines, at the same time, both the identity of the fuzzy subgroups in the population and the profiles of attributes that define the subgroups.

The method calls for a number of grade-of-membership computer runs or iterations. For each run, the number of pure types (K) is predetermined by the analyst. The procedure is applied iteratively until the model converges. The parameters of the grade-of-membership function are estimated by maximizing a multinomial likelihood function. The choice of K, K + 1, or K − 1 groups is based on the goal of minimizing the heterogeneity and maximizing the homogeneity within groups. At convergence the solution for the values of the variables represents the maximization of the within-group homogeneity and the minimization of the between-group homogeneity.

The grade-of-membership method was introduced by Woodbury and Clive (1974) and Woodbury and Manton (1983) as a way of studying the symptoms of clinical conditions. Portrait et al. (2001), applying their method to the concept of

health status, conclude that underlying health status can be described by six health dimensions ("pure types"). Portrait et al. (2001) used estimates of the dimensions of health derived by the GoM method to apportion total life expectancy at selected older ages into components denoted as "health expectancies" (expectancies by cause of death). For example, the latter authors give the following health expectancies (in years) at age 65 for their sample of the population of Amsterdam during the 1990s:

	Females	Males
Total life expectancy:	20.3	14.9
Type 1(COPD, cancer):	2.2	1.3
Type 2 (Other):	3.6	2.0
Type 3 (Cognition):	2.2	1.2
Type 4 (Arthritis):	2.8	1.1
Type 5 (Cardiovascular):	2.1	2.3
Type 6 (Healthy):	7.4	7.0

Source: Portrait et al. (2001), Figure 2

The results indicate that each sex may be expected to live about 7 healthy years and that the additional years females "enjoy" over those of men are very likely to be unhealthy years.

Meta-Analysis

Meta-analysis is the statistical analysis of the results of a group of sample studies, all designed to test a given hypothesis or treatment of a health condition, for the purpose of arriving at an integrated conclusion. It aims to interpret or reconcile conflicting or indeterminate conclusions even where the primary results do not permit firm conclusions. By pooling the results of several sample studies, the analyst greatly increases the effective size of the sample. Combined analysis of the primary results reduces the size of the confidence interval and adds to the power of the findings (*i.e.*, reduces the chance of false positives and false negatives).

Meta-analysis may include studies with different research designs, different concepts, and different measures of the size of the effect of the dependent variable. The dependent variable may be the prevalence of some characteristic in an absolute or ratio form, or the relative risk or the odds ratio if the outcome of the study is a dichotomous variable. A common metric (a scale-free form) is required in the effect-size of the dependent variable before the results of the different studies can be combined. The standardized mean difference and a regression coefficient are possible common metrics. Meta-analysis seeks to evaluate the overall significance of the effect size of the many studies included. For this purpose, it combines the individual estimates of the many studies in statistically appropriate ways.

In the simplest approach to meta-analysis, the analyst counts the number of studies with positive results and the number of studies with negative results, and determines whether the difference is significant. This simple procedure does not take account of the size and variances of the samples used in the studies. The models usually employed in meta-analysis undertake to make these allowances and so are more complex.

By its very nature meta-analysis faces a number of problems. There are three common concerns: The comparability of the data and of the study designs; the omission of some relevant studies; and the variability in the quality of the studies. As a result, it is a controversial research method. Critics argue that meta-analysis cannot overcome the defects of poorly-designed studies or eliminate response bias and that, if the individual studies show marginally significant results, combining them in a meta-analysis does not necessarily produce statistically significant results. Analysts differ on whether to include studies of marginal quality. Some researchers would include weaker studies, accompanied by a statement or measure of how this may affect the quality of their conclusions. In spite of the criticisms, meta-analysis is a popular way of trying to find useful results in those cases where a body of studies does not show that an effect is statistically significant, or where studies yield inconsistent results even though they may be statistically significant.

Propensity Score

Matching with propensity scores is often used in the analysis of health data to reduce or eliminate the effect of selection bias when estimating the effect of treatments or exposures on outcomes in observational data. A propensity score is the probability that an individual will engage in a certain activity or be in a certain state, as determined usually by a logistic regression relating the probability and a series of variables believed to be associated with this activity or state. Propensity scores are calculated so as to form matched pairs of treated (or exposed) and untreated (unexposed) subjects on the basis of similar propensity scores.

For example, to control bias as to whether or not a couple selects a hospice for a dying spouse in a study of mortality outcomes of the surviving spouse, Christakis and Iwashyna (2003) computed propensity scores for every couple, both those where the dying spouse enters a hospice and those where the spouse does not. The selection of a hospice can be confounded by factors that are related to the mortality outcomes of the surviving spouse (*e.g.*, patients with healthier partners may be more likely to enroll in a hospice). To control such bias, variables that can affect the decision to use a hospice must be identified (*e.g.*, for the patient, his or her diagnosis, co-morbidity score, and duration of illness; for the surviving spouse, age, race, co-morbidity score) and measured. Selection bias may then be reduced by matching those couples choosing a hospice and the controls (*i.e.*, those couples not choosing a hospice) according to their propensity scores. This example of the application

of the propensity score is described more fully in Christakis and Iwashyna (2003). The general method is described in Rosenbaum and Rubin (1984) and Smith 1997).

Special Survey Methods

Techniques for Collecting Sensitive Data

In research on population health involving surveys, it is often necessary to secure information on sensitive subjects, such as having had an abortion, using illegal drugs, or driving while intoxicated. In such cases, it may not be wise to ask a direct question, even if self-enumeration is the method of collection. Various approaches have been devised to assure privacy of responses from the interviewer and all other persons involved in handling the survey data and to encourage accurate responses to sensitive questions. We describe two such procedures as illustrations of how this problem has been handled – the randomized response technique and the three-card method. In Chap. 2 adaptive sampling was also described.

Randomized Response Technique

The older procedure designed to secure more complete and accurate responses on sensitive subjects is the randomized response technique. The technique has more than one variation. Commonly, the sensitive question is asked in conjunction with another innocuous question, and the respondent determines which question to answer, using some probability device under his or her control. For example, if the probability device is the toss of a coin, heads could mean that he or she should respond to the sensitive question. The enumerator would not know the choice in the individual case, but the probability of picking the sensitive or the innocuous question over all respondents is known. These theoretical probabilities can be compared with the actual percentages to determine the relative frequency of the responses to the sensitive and innocuous question. The method has some limitations. The randomized response technique requires respondents to perform complex randomization tasks which the interviewer has to try to explain to them. Furthermore, the method does not allow a check of respondent answers through further questions. (See Shimizu and Bonham 1978.)

Three-Card Method

Another procedure for collecting sensitive data that assures privacy of response is the three-card method. This is also a survey-based indirect estimation technique but it seeks to avoid the complex design of the randomized response technique.

The method was devised to provide estimates of the proportion of illegal aliens among the foreign-born population, but it has the potential for wider use, including use in securing information on various sensitive health problems (Droitcour et al. 2001).

In this method, the population is divided into three representative samples, each of which will answer the key question regarding the respondent's health but with a slightly different arrangement of the mutually exclusive categories for responding to the key question. A set of three cards, each containing all the mutually exclusive categories for the question but arranged in three different groups in each sample (boxes A, B, and C), is given to the respondent. The respondent is asked to select one of the three groups on the card. Say, the survey is about securing information on sexually transmitted diseases. The sensitive category is listed on one of the cards but it is always grouped with one of the nonsensitive categories (box A, B, or C). (One of the nonsensitive groups is listed separately on each of the cards (box A, B, or C).) The investigator can estimate from the combined responses the proportion of persons who fall in each of the categories. Each sample yields directly the separate proportion in one of the nonsensitive groups and the proportion for the sensitive group is obtained by subtraction of all these percentages from 100.0.

The method is designed to apply to questions with a limited number of response categories in order to determine the proportion of persons in only one of these – *e.g.*, illegal aliens among a brief list of citizenship statuses (*i.e.*, citizen, legal resident alien, refugee or asylee, illegal alien, other. With respect to health, the application of the three-card method is more limited because it cannot be used with the numerous health categories that are possible. The health categories may be broadly grouped, however, so as to reduce their number: For example, the categories could be sexually transmitted diseases (*i.e.*, the target condition), musculoskeletal diseases, cardiovascular diseases, cancer, other conditions, none of the others. The issue of the proportion of all pregnancies terminating in abortion can be structured in a few pregnancy outcomes – miscarriage, abortion, late fetal death, and birth. The results could be summarized as follows:

$$1.00 - .35 - .25 - .15 = .25$$

All pregnancies − births − miscarriages − fetal deaths = abortions

Survey Design for Rare Events: Multiplicity/Network Sampling

Ordinarily the data secured in large sample surveys assign everyone to some broad category of a characteristic, such as an age group, marital status, or income group. Sometimes we need data on some uncommon or rare health condition and its costs. This information may be needed for health planning and program evaluation. For example, we may want to know how many people have cancer and the costs of the treatment. It is possible to secure such data either from medical-care providers or

households, but it is difficult to secure the required information from the former and health analysts have therefore resorted to household surveys.

Household surveys utilizing the conventional counting rule confine themselves to the sample households selected by the sampling design. This design suffers from some serious statistical problems in those cases where data on uncommon or rare health conditions are being obtained. Serious illnesses have low prevalence ratios and not many cases are found even in large samples, so that the results have large sampling errors. Second, the institutional population is excluded and this population may be particularly affected by the health condition. Finally, nonresponse may be excessive, especially among aged persons.

Multiplicity sampling, or network sampling, is a survey methodology designed to improve the efficiency of surveys in securing data of this kind. In multiplicity sampling the members of the sample households report on the occurrence of a condition or event among a specified range of relatives resident in other households, such as siblings, parents, and children. In the basic application, only the original sample household is interviewed but the effective size of the sample is extended at little cost by securing information about rare health conditions among the relatives' households (Sirken 1970; Levy 1977). For example, if a sibling counting rule is combined with the conventional counting rule to secure information regarding the prevalence of diabetes, the diabetes patient can be reported not only at his/her residence but also at any household including one or more of his/her siblings. (If the relatives' households are interviewed in addition, their responses may be used to evaluate the original responses.) Multiplicity counting rules may be useful for estimating prevalence of a wide variety of health conditions, diseases, disabilities, and impairments.

Network sampling does not necessarily reduce the problem of response bias since the respondents have to know about the health conditions of their relatives and be willing to report them. This process can add to the complexity of the survey and its costs. Well-trained interviewers are needed to handle a somewhat complex questionnaire and the unusual interview situation in which respondents are asked about relatives. Inasmuch as the information on health is obtained from relatives, there is an increased possibility of obtaining erroneous data, especially since only one person responds for the entire sample household.

In a study evaluating the efficiency of multiplicity sampling to derive the proportion of cancer cases selected from the tumor registries at two Illinois hospitals, Czaja et al. (1986) found that multiplicity counting rules always yielded more cases with the desired characteristics than the conventional counting rule, and that multiplicity counting rules had slightly higher reporting biases but consistently lower mean squared errors than did the conventional rule unless the sample is very large. If the condition is very rare (say less than 0.5%), it may be necessary to broaden the scope of the network counting rules to include other relatives or use other methods of collection such as administrative records or oversampling of segments of the population.

Network sampling requires developing multiplicity estimators taking account of the number and characteristics of all the relatives who are eligible to be included.

Estimates of the total number of persons with a health condition are developed by weighting the responses with the respondent's sample weight and the number of persons eligible to report on each person identified as having the health condition (Sirken 1970).The sample weighting scheme has to be adapted to allow for the possibility of duplicate reporting on the same person, for example, two households in the sample having the same sibling relative who has a health condition.

References and Suggested Readings

Statistical Modeling and Mathematical Demography

Arminger, G., Clogg, C. C., & Sobel, M. C. (1994). *Handbook of statistical modeling for the social and behavioral sciences*. New York: Kluwer Academic.
Halli, S. S., & Rao, K. V. (1992). *Advanced techniques of population analysis*. New York: Plenum.
Namboodiri, K. (1991). *Demographic analysis: A stochastic approach*. San Diego, CA: Academic.
Schoen, R. (1988). *Modeling multigroup populations*. New York and London: Plenum Press.

Cohort Analysis

Glenn, N. D. (2005). *Cohort analysis* (2nd ed.). Quantitative applications in the social sciences, U.S. Thousand Oaks, CA: Sage Publications.
Halli, S. S., & Rao, K. V. (1992). Age, period, and cohort effects in demography. In *Advanced techniques of population analysis* (Chap. 3). New York: Plenum Press.
Hobcaft, J., Menken, J., Preston, S. (1982, Spring). Age, period, and cohort effects in demography: A review. *Population Index, 48*(1), 4–43.
Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1997). Age, period, and cohort effects. In *Epidemiologic research: Principles and quantitative methods* (Sect. 7.3). Belmont, CA: Lifetime Learning Publishers.
Mason, W. M., & Fienberg, S. E. (Eds.). (1985). *Cohort analysis in social research: Beyond the identification problem*. New York: Springer. See esp. papers by W. M. Mason, & S. E. Fienberg, Introduction, beyond the identification problem; S. E. Fienberg, & W. M. Mason, Specification and implementation of age, period, and cohort models; and W. M. Mason, & H. L. Smith, Age-period-cohort analysis and the study of deaths from pulmonary tuberculosis.
Palmore, E. (1978, September). When can age, period, and cohort be separated? *Social Forces, 57*(1), 282–295.

Proportionate Hazards Models

Allison, P. (1984). *Event history analysis: Regression for longitudinal event data*. Quantitative Methods in the Social Sciences, V. 16. Beverly Hills, CA: Sage.
Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*. Series 534, 187–202.
Smith, P. J. (2002). *Analysis of failure and survival data*. Virginia Beach, VA: Chapman and Hall CRC.

Grade-of-Membership Analysis

Manton, K., & Woodbury, M. (1992). Grade of membership analysis in the epidemiology of aging. In R. A. Wallace, & R. F. Woolson (Eds.), *The epidemiologic study of the elderly* (pp. 333–357). New York: Oxford University Press.

Portrait F., Lindeboom, M., & Deeg, D. (2001). Life expectancies in specific health states: Results from a joint model of health status and mortality of older persons. *Demography, 38*, 525–536.

Woodbury, M. A., & Clive, J. (1974). Clinical pure types as a fuzzy partition. *Journal of Cybernetics, 4*, 111–121.

Woodbury, M. A., & Manton, K. G. (1983). A new procedure for analysis of medical classification. *Methods of Information in Medicine, 21*, 210–220.

Other Techniques

Christakis, N. A., & Iwashyna, T. J. (2003). The health impact of health care on families: A matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses. *Social Science and Medicine, 57*, 465–475.

Guo, S., & Fraser, M. (2010). *Propensity score analysis: Statistical methods and applications*. Sage Publications, Inc. Thousand Oaks, CA: Pine Forge Press.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*, 516–524.

Smith, H. L. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology, 27*, 325–353.

Multiplicity/Network Sampling

Czaja, R. F., Snowden, C. B., & Cassady, R. J. (1986, June). Reporting bias and sampling errors in a survey of a rare population using multiplicity counting rules. *Journal of the American Statistical Association, 81*(394), 411–419.

Levy, P. S. (1977). Estimation of rare events by simple cluster sampling with multiplicity. In *Proceedings of the social statistics section* (pp. 963–966). Alexandria, VA: American Statistical Association.

Sirken, M. G. (1970). Survey strategies for estimating rare health attributes. In *Proceedings, Sixth Berkeley Symposium on Statistics and Probability* (pp. 133–144). Berkeley, CA: University of California Press.

U.S. National Center for Health Statistics. (1977). Multiplicity study of marriages and births in Israel. By G. Nathan, U. O. Schmelz, & J. Kenvin. *Vital and Health Statistics*, Series 2, No. 70. DHEW Publication No. (HRA), 77–1344.

Methods for Collecting Sensitive Data

Beck, F., & Peretti-Watel, P. (2002, May–June). The impact of data collection methodology on the reporting of illicit drug use by adolescents. *Population-E, 57*(3), 571–591.

Droitcour, J. A., Larson, E. M., & Scheuren, F. J. (2001, August 5–9). *The three-card method: Estimating sensitive survey items – With permanent anonymity of response*. In Proceedings of the annual meeting of the American Statistical Association.

Fox, J. A., & Tracy, P. E. (1986). *Randomized response: A method for sensitive surveys*. Quantitative applications in the social sciences, V. 58. Sage Publications, Inc. Thousand Oaks, CA: Pine Forge Press.

Kingsbury, N., Droitcour, J. A., & Larson, E. M. (2003, April 28–30). *Estimating illegal immigrants in a census or survey: The three-card method – an innovative technique*. Paper presented at the Joint-ECE-Eurostat Work Session on Migration Statistics, Geneva, Switzerland.

Shimizu, I., & Bonham, G. (1978). Randomized response in a national survey. *Journal American Statistical Association, 73*, 35–39.

Warner, S. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal American Statistical Association, 66*, 63–69.

Appendix 3: Glossary[1]

Abortion Induced termination of pregnancy.

Acceptor (Nonacceptor) A woman who accepts and employs (fails to accept and to use) contraceptive devices in a survey or trial studying the extent to which family planning programs are effective.

Active (Healthy, Disability-Free) Life Expectancy A measure derived from a table of active (healthy, disability-free) life representing the average number of years of active (healthy, disability-free) life remaining at birth or, alternatively, at various successive ages, to a cohort of births or survivors at specified ages. The base of the measure may be total survivors or active (healthy, disability-free) survivors.

Activities of Daily Living (ADL) A specific list of basic abilities relating to personal care routines. These abilities are eating, dressing, toileting, grooming, transferring into and out of bed, and bathing. The inability to carry out one or more of these routines usually leads to a classification of the person as disabled.

Acute Illness (Disease) An illness or disease of short duration terminating in cure or death, usually readily treated, and caused mainly by external forces, such as a bacterial or viral infection or injury.

Age at Specified Years until Death The age in the life table at which a person has, for example, 10 or 15 years of remaining life.

Age-Bounded Life Expectancy The average number of years of life expected to be lived in an age-group for persons at the initial age of the age group. It is defined formally as the quotient of the difference between T_x and T_{x+i} divided by l_x in the life table.

Age-Dependency Ratio The ratio of the number of persons in the principal nonworking ages – children and elderly – to the number of persons in the principal

[1]The contribution of the Biogerontology Terminology Committee of the Gerontological Society of America, directed by Dr. Leonard Hayflick, to the biogerontological terms in this glossary is acknowledged. The author served on this committee and benefited from the deliberations of the committee. He is alone responsible for all definitions, however.

working ages. One common form of the age dependency ratio is the ratio of the population under 18 and 65 year of age and over to the population 18–64 years of age. The age-dependency ratio is often partitioned into the child dependency ratio and the elderly or aged dependency ratio. They can be further distinguished as familial and societal, depending on whether the numerator ages and the denominator ages represent ages that are matched for dependency within families or are matched for dependency of an age group and the broad working ages.

Age Effect The influence of age on a variable or measure of interest (a rate or ratio) resulting from the fact that the rate or ratio changes with age either at a particular calendar year or over time.

Age-Period-Cohort Effects The separate or combined effects on the level and change in a variable or measure (rate or ratio) resulting from changes from age to age, period to period, and (birth) cohort to (birth) cohort. The analyst's interest may be to separate these effects in order to interpret changes in the variable or the measure.

Age-Specific Death Rate The number of deaths in a year at a given age (or given ages) divided by the midyear population at that age (or those ages) in the year. Conventionally the rate is computed per 1,000 population and is represented by the symbol m_x (or ${}_nm_x$). It is described as a type of central rate because the midyear population is used in the denominator.

Age-Specific Mortality Rate or Conditional Probability of Dying The proportion of a group of individuals that die between exact age x and exact age $x + 1$ (or $x + n$) during a given year (or period of n years) among the survivors to the initial exact age x. It is conventionally symbolized by q_x (or ${}_nq_x$). A probability is a type of rate that defines the population exposed to risk more precisely than a central rate and is described as being in cohort form.

Aging *See* Individual Aging, Biological Aging, Population Aging.

Allele One member of a pair or series of genes that resides at a particular locus (*i.e.*, position) on a particular chromosome.

Allostasis The body's ability to adapt its internal physiologic milieu to match external demands. Adaptive allostasis is a desirable response to stress, particularly among younger persons. Allostatic load, in contrast, is the nonadaptive form representing a combination of biomarkers reflecting dysregulation of bodily organs; it is cumulative and commonly associated with increasing age.

Allostatic Load The cumulative dysregulation of a person's bodily processes, commonly associated with increasing age. It is tracked by use of a package of biomarkers over the life course of the person. Sometimes it is referred to as cumulative allostatic load.

Amniocentesis A type of prenatal diagnosis in which a small sample of amniotic fluid is extracted from the uterus of a pregnant woman to determine possible genetic abnormalities in the fetus.

Angiogenesis The process by which new blood vessels are formed. Gene therapy is being tested for its utility in promoting angiogenesis and its opposite, antiangiogenesis, that is, controlling the growth of blood vessels.

Antagonistic Pleiotropy The phenomenon that some multipurpose genes have beneficial effects early in life but deleterious effects late in life.

Anthropometric Data Bodily measurements of individuals intended to assess the health of the person, such as weight, height, skin-fold thickness, waist circumference, and waist-hip ratio.

Antioxidants Compounds that neutralize oxygen radicals. Some are nutrients like vitamin C and others are enzymes like superoxide dismutase (SOD).

Assisted Reproduction The group of methods designed to treat inability to conceive and bring to term products of conception by couples on their own. These techniques include variations of intrauterine insemination, donor insemination, *in vitro* fertilization, and intracytoplasmic sperm injection.

Asymptomatic Describes a patient free of symptoms perceived by the patient. A patient with glaucoma or an abdominal aortic aneurysm in its earlier stages may be asymptomatic although the condition can be detected clinically.

Autophagy Scavenger and repair operations of genes.

Average (or mean) Observed Life Span – Decedents The average of the "verified" ages at death of the group of persons who died in a given year. A second definition is the average of the "verified" ages at death of the group of persons who died in a given year at age 100 or more.

Average (or mean) Observed Life Span – Survivors The average (mean) of the "verified" ages of survivors to a selected age or age range, such as survivors to ages 100 and over.

Ayurvedic Medicine A form of traditional medicine practiced in India representing a holistic approach to health care. It is based on the principles of Ayurveda, an ancient Hindu health system, and employs yoga, dietary modification, massage, herbal preparations, and other devices.

Biodemography The branch of demography that is informed by biology, including the role of evolutionary biology in human demographic processes, the role of biological factors in human fertility and mortality, especially the age pattern of death rates, the demography of subhuman species for understanding human demographic processes, and the use of biomarkers in the analysis of human aging and longevity.

Biogerontology The subfield of gerontology concerned with biological aging in its dual senses (s*ee below*), both in humans and in subhuman species.

Bioinformatics Information technology as applied to biology, especially the technology used in the collection, storage, retrieval, and analysis of genomic data.

Biological Aging Changes with time in the cells, tissues, and organs of an organism that reflect an increasing vulnerability to challenges, as a result of which the ability of the organism to survive is decreased. The term is also used to refer to the accumulated damage to the building blocks of life (e.g., proteins, fats, carbohydrates, DNA, etc.) that begins early in life and whose source is the product of the biochemistry of life itself. In the latter meaning it corresponds to SENESCENCE. According to most biogerontologists, biological aging is not a disease although disease may be an integral part of biological aging.

Biomarkers Measures of biological changes that characterize the aging process in humans. There are numerous biomarkers, but none has been consistently applicable to a particular individual as a basis of judging his/her biological age. Among the biomarkers are grip strength, blood pressure (e.g., mean arterial pressure), pulmonary functioning (e.g., forced expiratory volume), motor functioning (e.g., ability to touch left earlobe with right hand), kidney function, and bone mass.

Body Mass Index (BMI) An index of protein and fat stores, undernutrition, underweight, or overweight calculated by the formula, weight (in kg) ÷ height squared (in m).

Burden of Disease A measure of the combined effect of mortality and disease/injuries, more specifically a combination of years of life lost prematurely (YLL) and years of life lived with a disability (YLD).

Caloric Restriction An experimental intervention in animal behavior involving a sharp reduction in calories while maintaining the necessary level of nutrients, and being studied to determine its impact on longevity. In laboratory settings, the life spans of animals have been extended by caloric restriction but the evidence for its efficacy has not yet been demonstrated for humans.

Case-Control Study A study that compares two population groups, one with the study characteristic (e.g., having a particular disease) and another, a control group free of this characteristic, on the basis of a variety of risk factors and demographic/socioeconomic characteristics. Inasmuch as the study examines the record of the past to explain a current outcome from possible historical causes, it is usually considered a type of retrospective study.

Case-Fatality Rate The percentage of the population with a disease in given year and area dying from that disease during the year. The denominator could be the population with the disease at mid-year or at any time during the year. One could also construct a case-fatality rate with a numerator based on the number dying of the disease by a specified number of years later, but this is not conventional.

Case-Survival Rate The proportion of a population with a disease in a given year and area living with the health condition a specified number of years later. The numerator excludes deaths to the original group, new cases with the disease, new immigrants, and emigrants. Its complement approximates one version of the Case-Fatality Rate.

Cause-Eliminated (or Deleted) Life Table A life table constructed on the assumption that a particular cause of death or group of causes of death has been eliminated. One design for constructing such a table is simply to recalculate the age-specific death rates excluding the deleted cause. The new life table will show a higher life expectation, and provide the basis for measuring the years of life added by eliminating the cause in question under the assumption of independence of the cause-specific death rates.

Cause-Specific Death Rate The number of deaths from a specific cause in a year divided by the midyear population. It is generally expressed per 100,000 population.

Cell Line A cell line is living tissue that is being grown in culture for research purposes.

Censored Data Data in clinical trials or survival studies for which the outcome is unknown by the date the study concludes. Some persons survive to the conclusion of the study, neither dying nor developing the disease that is the subject of the study, and it is common in clinical trials to lose track of some patients observed in the study before the study concludes.

Centenarian A person with a verified age of 100 years or more.

Chorionic Villus Sampling A type of prenatal diagnosis in which a small sample of tissue from the outer layer of the placenta of a fetus in a pregnant woman is obtained and examined to determine possible genetic abnormalities.

Chromosome A cellular structure composed of DNA and proteins and containing genes. Humans have 23 pairs of chromosomes in each body cell, one of each pair coming from the mother and the other from the father.

Chronic Illness (disease) An illness or disease lasting over a long period of time, requiring ongoing treatment over this period, typically resistant to cure, and caused essentially by internal causes rather than by an external one, such as heart disease or chronic obstructive pulmonary disease.

Circular Migration The repeated movement of individuals or groups between two or more locations and a return to the original place of usual residence during some specified period. It does not usually involve a "permanent" change to another usual residence, but it may do so, such as when there are several changes of usual residence and then a return to the original place of usual residence.

Cohort A group experiencing the same demographic, socioeconomic, or health event (e.g., birth, marriage, graduating high school) in a year on other brief period which is tracked over time with reference to their experience of another demographic, socioeconomic, or health event (e.g., death, divorce, having a heart attack). The most common type of cohort considered is a birth cohort, or group of persons born in the same year or having the same age at a particular date. *See* Synthetic Cohort.

Cohort Analysis Analysis of changes in a cohort over some stated period of time with reference to some demographic, socioeconomic, or health characteristic. *See* Cohort, Age/Period/Cohort Effects.

Cohort Effect The influence of the characteristics of a cohort on a variable of interest (i.e., a rate or ratio) resulting from the fact that these characteristics vary from cohort to cohort, as reflected in the variable either in a particular calendar year or over time.

Cohort Interquartile Survival Range The difference between the ages to which 25% and 75% of the original cohort of births survive in a cohort (or generation) life table

Cohort or Generation Life Expectancy Average future lifetime for those surviving to exact age x in the life table from an initial cohort of individuals born within a specified interval (e.g., same year) and tracked with respect to their mortality from exact age x to the end of life. Usually refers to the figure at birth and is based on the age-specific mortality rates for successive calendar years and successive ages beginning with the infant mortality rate in the birth year of the cohort.

Cohort (or generation) Life Table A life table based on the age-specific mortality rates corresponding to the successive ages and successive calendar years for the birth cohort of the particular life table.

Comparative Effectiveness Research "Conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat, and monitor health conditions in real world settings." – Federal Council on Coordinating Comparative Effectiveness Research.

Competing Risks Describes the situation where the individual is subject to more than one type of risk from a class of events and the occurrence of one event to the individual removes him or her from the risk of experiencing another event of the same type or other type in the class of events. Consider, for example, the case of death from various causes; the individual who dies of a stroke is no longer at risk of dying from a stroke, cancer, or other cause.

Complex Humanitarian Emergency A relatively acute situation in which a large civilian population is affected by war, civil strife, or a violent attempt to restructure the state, leading to large-scale population displacement, deterioration of living conditions (e.g., shortages of food and potable water), and excess mortality.

Complex Survey Design A survey sample design that progresses through various stages and involves stratification of the sample population, such as sample selection of areas, then sample selection of household (or establishment) clusters within areas, and finally sample selection of individual households (or establishments) within clusters.

Compression of Mortality The progressive reduction of the range of ages over which deaths are concentrated. A variety of measures have been used to gauge the compression or expansion of mortality. They differ as to the degree and direction

of the concentration of deaths, but for the most part they show that a considerable compression of mortality has occurred in the United States in the last century and the process continues at the present time.

Compression of Morbidity The progressive reduction of the gap between a population's life expectation and its average age of onset of chronic disease. If the latter age rises or the former age declines, there is a compression of morbidity. The term usually refers to the situation where the average age of onset of chronic disease rises faster than life expectancy. It is difficult to determine the average age of onset of morbidity. The contrasting situation is an expansion of morbidity, where the gap between a population's life expectation and its average age of onset of chronic disease widens.

Contraceptive Prevalence Ratio The percent of currently married women, or women living in a stable *de facto* union, of reproductive age (the range 15–49 years of age) who use a contraceptive method, usually at a midyear date.

Control Group In epidemiological studies, a group of persons recruited into the study for comparison with the group having a particular health condition or receiving a treatment. The control group receives no treatment, receives a dummy treatment (a placebo), or receives a treatment with known effect.

Cormic Index The ratio of leg length to trunk length, or sitting height to standing height. Used to adjust the body mass index for differences in body shape.

Cost-Effectiveness Analysis (or Comparative Effectiveness Analysis) Research into which medical procedures, either diagnostic or therapeutic, are more effective, as measured by the value of the health procedures. The value of a procedure can be measured by improvements in health outcomes, say by relating quality-adjusted life years (Qualys, *see*) added by the procedure to the cost of the procedure.

Cross-Sectional Survey A survey of a group of people at a single date. It contrasts with a Longitudinal Survey (*See*).

Cryopreservation Preservation of sperm, eggs, or embryos by freezing and storage for possible later thawing and use in *in vitro* fertilization.

Daily Crude Mortality Rate Deaths per 10,000 population per day. This measure is used to determine comparative mortality levels among refugees and internally displaced persons.

Death Rate The number of deaths in a year divided by the population exposed to the risk of death (typically expressed in person-years). When the calculation refers to single calendar years, the midyear population can be used as an approximation to the population exposed to risk. When the time interval is longer, the population at the midpoint of the interval can be multiplied by the length of the interval to derive an acceptable estimate of the population exposed to risk. Conventionally the rate is computed per 1,000 population and is represented by the symbol, m.

Degenerative Diseases Chronic diseases, usually of later life, that typically have a substantial biological or genetic basis and are resistant to treatment and cure (for example, cancer, cardiovascular diseases, diabetes, cirrhosis of liver). They are also referred to as endogenous or intrinsic diseases.

Design-Based Estimate An estimate from a national sample survey for a "domain," such as a small geographic area or sociodemographic group in a small geographic area, derived solely from the sample units within the domain (allowing for adjustment of such populations to independent estimates for larger areas).

Disability-Adjusted Life Years (DALY) A measure of the burden of disease representing commonly a weighted combination of years of life lost because of premature death (YLL) and years lived with disability (YDL). The term is also applied to a measure of years lived with disability only.

Disability-Free Aging The process of growing old without a disability (e.g., inability to perform an ADL)

Disease An interruption, cessation, or disorder of body systems, or organ structure or function.

Dissimilarity Index (DI) A measure of inequality of two distributions with respect to some variable of interest. The formula is $1/2$ (sum of the absolute differences between two distributions expressed as percents of their totals).

DNA (Deoxyribonucleic Acid) The molecule that carries the genetic code for all life forms (except for a few viruses). It consists of two long twisted chains (a double helix) composed of nucleotides. Each nucleotide contains one of four nitrogenous bases, one phosphate molecule, and the sugar molecule deoxyribose. The bases in DNA nucleotides are adenine (A), thymine (T), guanine (G), and cytosine(C).

Dominant Trait A trait that is expressed in anyone who inherits the gene for the trait. More formally, dominant refers to one of a pair of alternative alleles that masks the effect of the other allele when both are present in the same cell or organism. *See* Allele.

Dual Systems Analysis Case-by-case matching of records from two collection systems for the purpose of evaluating one of the sets of records by the other. The collection systems may be individual records from a census, administrative records from a health insurance system (e.g., Medicare), records from a vital statistics system, or other collection system. One of the two sets of records is assumed to be the standard for evaluating the other.

Edema Excessive accumulation of body fluids, usually in the feet or legs, often resulting from or associated with undernutrition and resulting in swelling of the area.

Elderly Support Ratio The ratio of the number of persons in the principal working ages to the number of persons in the principal elderly nonworking ages. A common form of the elderly support ratio is the ratio of the population 18–64 years of age to the population 65 years and over, expressed as a percent.

Endemic Characteristic of a disease that is concentrated in a local geographic area or population group.

Endogenous /Exogenous Deaths A classification of deaths according to cause into two broad classes, distinguishing those deaths primarily resulting from the intrinsic (i.e., aging and genetically-related) causes and those resulting from extrinsic or external causes. The latter causes include infectious and parasitic diseases, other acute conditions, and adverse effects (accidents, homicide, suicide), and the former causes include most cardiovascular diseases, malignant neoplasms, diabetes, and other chronic diseases and disorders. Some demographers have referred to endogenous mortality as senescent mortality and to exogenous mortality as background mortality. Biologists favor the terms intrinsic and extrinsic for these two categories of deaths.

Environmental Racism Selective location of environmental hazards such as waste dump sites, nuclear power plants, chemical plants with toxic by-products, and other locally unwanted land uses, in racial-minority neighborhoods.

Enzyme A protein that promotes a specific biochemical reaction in the body without itself being permanently changed or destroyed. Enzymes, for instance, are involved in blood clotting and initiate the process of breaking down food.

Epidemiology The field of public health that is concerned with the causes, distribution, and control of health events (e.g., diseases) in a population. By extension, it encompasses the statistical analysis of the determinants and geographic distribution of health related conditions and processes, such as aging.

Epidemiological Transition The shift in causes of deaths and diseases from the infectious diseases to the chronic, degenerative diseases that took place in the second half of the twentieth century in the more developed countries.

Epigenetics Roughly, environmentally imposed modifications to the genome that in turn determine the phenotype. Alternatively, a study of the processes that account for the unfolding of the genetic program and that are jointly responsible for the development of an organism. It encompasses the way genes are expressed in different cell types and inheritable information superimposed on DNA. Authorities disagree on whether the concept is limited to changes that are somatically heritable, that is, changes that appear in later generations. Epigenetic stability refers to the stability of cells to maintain their function without mutation or without changing as a result of incorrect expression of control mechanisms.

Event-Exposure Method A method of constructing a multiple decrement-increment life table in which the increments and decrements other than death at each age are explicitly allowed for by probabilities of the event occurring to the initial population at each age.

Exogenous Deaths Those causes of death that result from factors external to the individual, such as deaths from violence (accidents, homicide, suicide) and

infectious and parasitic diseases. This is the narrow definition. A broader definition includes, in addition, those deaths that are caused essentially by infectious agents or abuse of toxic substances, such as lung cancer, cirrhosis of liver, and rheumatic heart disease. *Compare* Endogenous Deaths.

Exponential Curve A mathematical expression for a constant rate of change over time or age, either growth or decline, as in $P_1 = P_0e^{rt}$, where P = population at two dates, t = time, a variable, and r = rate of change, a constant.

Extreme Aged Persons aged 95 years and over or 100 or more. Also referred to as Advanced Aged. Preferences vary among demographers and gerontologists in their use of this term.

Extreme Longevity State of having lived to at least 95 or 100 years of age.

Extrinsic Diseases *See* Exogenous Diseases.

Failure Rate *See* Hazard Rate.

Family Planning A family or community program designed to control the number and spacing of children. It encompasses contraceptive use, issues of unwanted pregnancies, and programs to deal with infertility.

Famine An extreme and widespread shortage of food in an area resulting in widespread undernutrition and starvation in a population.

Famine Edema Bilateral dependent pitting edema resulting from undernutrition. As an important sign of severe undernutrition, the condition carries a high risk of death. It is to be distinguished from edema resulting from other health conditions. *See* Edema.

Food Insecurity The condition where households or groups of households (*i.e.*, regions, provinces) lack sustainable physical or economic access to safe, nutritious or socially acceptable food for a healthy and productive life. Food insecurity is the proximate cause of undernutrition. *See* Undernutrition.

Force of Mortality *See* Hazard Rate.

Fourth Age In Peter Laslett's book, *A Fresh Map of Life*, the ages from 75 to 100, typically ages free of family and work responsibility but ages of rapidly diminishing physical and mental health and increasing dependency.

Frail Span/Non-Frail Span The period during which an individual survives in a condition of frailty. Of the total number of years of life remaining at a given age according to the current life table, the number of years the population on average will spend in a frail condition. The other years are the non-frail span years. *See* Health Span.

Frailty or Frailty Syndrome A health state characterized by greatly increased vulnerability to stressors, age-associated declines in the functioning of multiple physiological systems, and a compromised ability to maintain homeostasis. It is

clinically identified as a combination of three or more of the following criteria: unintentional weight loss, low physical activity level, slow walking speed, low energy, and diminished strength.

Free Radicals Molecules with unpaired electrons that react readily with other molecules. Oxygen free radicals, produced during metabolism, damage cells and may be responsible for aging in tissues and organs.

Gene A segment of DNA that is located on a chromosome at a specific location and that contains the "code" for a specific protein. It can have many different forms called Alleles.

Gene Expression The process by which the information contained in the genes is transcribed and translated into proteins.

Generation or Cohort Life Expectancy Average future lifetime for those surviving to exact age x in the life table for a cohort of individuals born within a specified interval (e.g., same year). Usually refers to the figure at birth and is based on the age-specific mortality rates for successive calendar years and successive ages beginning with the infant mortality rate in the birth year of the cohort.

Gene Therapy The treatment of disease by the transfer of healthy genes from one individual to an unhealthy person or from a person to other parts of that person's body to replace or supplement malfunctioning or diseased genes or to block genes from producing unwanted products.

Genetics Science of inheritance, seeking to explain the differences and similarities between related organisms and the way characteristics are passed on from parents to children.

Genome The whole set of genes carried by a species. The human genome has about 30,000 genes.

Genotype The genetic makeup of an organism. It consists of the complete set of alleles at all loci.

Geographic Information System (GIS) A computer-automated system of compiling and mapping statistical data identified by geographic coordinates, for specified sets of geographic areas identified by geographic coordinates. Geographic information systems represent the marriage of mapping procedures, statistical data, and computer technology.

Glycemic Index An index assigned to a food, typically carbohydrate-rich foods, based on the average increase in blood glucose levels resulting when the food is eaten.

Glycation (or Glycosylation) The process by which glucose links with proteins and causes these proteins to bind together. In some circumstances, this can result in stiffening of tissues and some of the physiological problems associated with aging.

Gompertz "Law" of Mortality A generalization about the rate of increase in the death rate as age increases, propounded by Benjamin Gompertz, British actuary. The "law" states that, over most of the adult years, the death rate rises by a constant factor for successive equal age intervals. Gompertz's "law" is symbolized by $\mu_x = Be^{cx}$, where μ_x represents the age-specific mortality rate and B and c are constants. Gompertz's "law" is a rough empirical generalization for which numerous modifications have been offered since it was first propounded.

Gross Domestic Product The total value of goods and services, usually for an entire nation.

Harmonization In survey research on international variations, refers to the adjustment of the data, definitions, and measures so as to achieve comparability and consistency between countries.

Hazard (or failure) Rate The relative rate of decline of the survival (reliability) function. With respect to mortality, the instantaneous probability of dying, representing the chance of dying in an infinitesimally short period. It is also called the force of mortality. If l(t) is the number of survivors to time t, then the hazard rate, (h(t)), is given as $[l(t) - l(t + \Delta t)] \div l(t)\,\Delta t$ as Δt tends to zero. More commonly h(t) is formulated in terms of S(t) rather than l(t) since the concept does not apply only to life table analysis. The hazard rate (force of mortality) may also be represented as: $h(x) = -d[\ln S(x)] \div dx$.

Health Condition A departure from a state of physical and mental well-being.

Health Dependency Ratio Refers to a class of ratios relating persons of working age and older who are unable to work because of a health condition to persons in the age group who may expected to support them, whether as their children or the broad range of working-age adults. Example are the ratio of persons who are 18 years and over who are disabled to persons 18–64 years, persons of working age who have a work disability to persons 18–64, persons 85 and over to persons 50–64 years of age.

Health Span The period of years at any age that an individual is expected to live without a serious chronic illness, disability, or ADLs/IADLs. It has also been called healthy life span. Health span is often assumed to be the same as disability-free years, active life expectancy, or healthy life expectancy. As with the term life span, this term may be applied to individuals as well as to population groups, but more commonly it is the average number of years that a population experiences a healthy state as a component of the average years of life expectancy.

Health Transition A broad shift in the health characteristics of a society, such as the current emergence of new infectious diseases or the reemergence of old diseases. More generally, the social, cultural, economic, behavioral, and psychological changes that are associated with changes in the health and mortality of a society.

Healthy Aging A theoretical concept referring to the process of growing old without serious health conditions. Preferred to the terms successful aging, productive aging, or active aging. Because it rarely describes any real person, however, it has its limitations as a description of later life. *See* Disability-Free Aging and Successful Aging.

Healthy (Disability-Free; Active) Life Expectancy A measure derived from a table of active or healthy life representing the average number of years a hypothetical cohort would live before incurring a disability (*e.g.*, inability to perform one or more of the "activities of daily life").

Healthy Life Span The part of an individual's total life span that he or she spends in a independent or active state. To be distinguished from Healthy Life Expectancy and Frail Span (s*ee*).

Healthy-Survivor Effect The phenomenon whereby a birth cohort includes a larger proportion of healthy survivors at later ages than at earlier ages because of the lower survival of members of the cohort with poorer health.

Heat-Shock Proteins A family of proteins involved in maintaining cell and somatic homeostasis and regulating the immune system. *See* Homeostasis.

Heritability/Heritability Index In the broad sense the degree to which the phenotype is genetically determined. It is measured by the heritability index, the ratio of genetic variance to the total phenotypic variance. The phenotypic variance is the sum of the genetic variance and the environmental (or nongenetic) variance. Heritability in the narrow sense is the degree to which a trait is passed on from parent to offspring; it is measured by the ratio of additive genetic variance to total phenotypic variance. Additive genetic variance is the mean effect of substituting one allele for another at a given locus, or at many loci for cases of polygenic traits. *See also* Phenotype/Phenotypic Variation.

Heritability Equation The basic equation is: $\mathrm{Var\,P} = \mathrm{Var\,G} + \mathrm{Var\,E}$, that is, the phenotypic variance equals the sum of the genotypic variance and the environmental variance. This formulation assumes that the variances are additive and that there is no covariance, or interaction between the genotypic variance and the environmental variance. This is an unrealistic formulation given the considerable interaction of these factors in non-Mendelian traits. A revision of the formula taking the additional factors into account is: $\mathrm{Var\,P} = \mathrm{Var\,G} + \mathrm{Var\,E} + 2\,\mathrm{Cov\,(G, E)} + \mathrm{Interaction\,(G, E)}$.

Heterogeneity The variation with respect to demographic, socioeconomic, and health characteristics among the members of a birth or other cohort, or other population group usually unobserved, leading to changes in its demographic, socioeconomic, and health composition as it ages, as a result of different rates of survival for the component and Unobserved Heterogeneity segments of the cohort and shifts in the socioeconomic status of the members.

Heterogeneity of Frailty The unobserved variation with respect to health among the members of a birth cohort, leading to different rates of survival for the compo-

nent health segments of the birth cohort and changes in the health composition of the birth cohort as it ages. *See also* Heterogeneity and Unobserved Heterogeneity.

Hispanic Paradox The counterintuitive finding that the Hispanic population in the United States has lower mortality rates than the non-Hispanic population although its socioeconomic status is lower.

Homeostasis The condition of equilibrium of physiological processes or systems in the body. Also the tendency of the body to maintain internal equilibrium by adjusting its physiological processes or systems as demands are made on one or another of them.

Hormesis A term employed primarily in toxicological studies to denote a generally favorable biological response to low exposures to toxins and other stressors but the opposite effect with large doses. Adapted for gerontological use, the term denotes the positive effect of mild stress on health in stimulating a person to adapt to their environment and to protect himself or herself from harm, but the negative effect of continuous and intensive stress on health. Another concept of hormesis interprets it as a physiological strategy that improves an organism's stress response, increasing stress resistance, such as exercise, caloric restriction (and its mimetics), and some pharmacological products.

Human Capital The capabilities and potentialities of humans as workers, homemakers, family members, and caregivers, etc., including their physical and mental health and their knowledge. It is measured in terms of health indexes, years of school completed, and similar measures of health and education.

Human Cloning The process of copying some part of human DNA, quite commonly carried out in the study of gene functions and in the application of gene therapy and recombinant DNA technology. These uses encompass therapeutic cloning. The term refers also to copying an entire human being as a reproductive method. Such asexual reproduction using somatic cells is intended to produce a genetically identical human being.

Human Development Index (HDI) A standard measure of human well-being, varying from 0 to 1 and based on life expectancy at birth, the literacy ratio, the enrollment ratio, and a measure of the material standard of living. It is employed by the United Nations Development Program to distinguish countries according to low, medium, and high development.

Iatrogenic Disease A health condition caused by medical treatment or the physician.

Idiopathic Refers to a disease of unknown cause. It is also used less frequently to refer to an intrinsic (or endogenous) disease.

Imaging Devices Refers to the variety of devices used to record images of the interior of the body, such as computerized axial tomography (CAT scan), magnetic resonance imaging (MRI), Doppler ultrasound, functional magnetic resonance

imaging (fMRI), positron emission tomography (PET scan), and single-photon emission computed tomography (SPECT). Some devices are designed to show body structures; others body functions.

Imputation The process of filling in missing subject entries for persons missed in censuses and surveys or filling in missing entries for missing subject items for persons included in surveys or censuses. *See* Multiple Imputation.

Incidence Rate The number of new cases of a disease or other condition in a population within a year or other short specified period per 1,000 or 10,000 in the population.

Individual Aging (or aging) The process by which individuals gain a year of age with the passage of a calendar year, i.e., gain x years of age with the passage of x years. The demographic factor mainly affecting changes in the numbers in any aging cohort is the level of age-specific mortality. *See* Biological Aging and Population Aging.

Induced Pluripotent Stem Cells (IPS Cells) Somatic cells "trained" to work like stem cells, i.e., cells that have the ability to differentiate into a variety of body cells (pluripotency).

Infant Mortality Rate Number of infant deaths in a year per 1,000 births in the year.

Infecundity Physiological inability to parent a living child.

Inflammation A nonspecific defensive response of the body's immune system to injury or disease, usually characterized by redness, heat, swelling, and pain.

Injury An injury is an acute physical disorder resulting from an accident or deliberate attack against a person. Injuries, unlike diseases, must be defined simultaneously by the causative event and the resulting pathology. The most common operational definition of an injury refers to those events coded to those pathologies included in Chapters XIX and XX of the International Classification of Diseases-10.

Instrumental Activities of Daily Living (IADL) Certain more complex routines associated with independent living, such as using the telephone, shopping for groceries or clothing, preparing a meal, managing one's finances, and doing light housework.

Internally Displaced Person A person who has been displaced from his usual residence in his country of nationality, is living in another part of the country, and is unable to return to his usual residence because of civil war or civil strife.

Interquartile Survival Range The difference between the ages to which 25% and 75% of the original cohort of births survive in a life table. The measure can be applied to either a period life table (i.e., a life table based on the age-specific mortality rates of a particular year or few years) or a generation or cohort life table

(i.e., a life table based on the age-specific mortality rates corresponding to a real birth cohort. The interquartile range is less commonly computed on the basis of an age distribution of observed deaths.

Intrinsic Mortality See Endogenous Mortality.

Intracytoplasmic Sperm Injection A form of *in vitro* fertilization in which a single sperm is injected into a single egg. *See In Vitro* Fertilization.

In Vitro Fertilization A type of assisted reproduction in which a woman's eggs are collected and fertilized with a man's sperm outside the body, and some of the resulting embryos are inserted in the woman's body with the goal of achieving pregnancy.

Least Developed Countries A select group of Less Developed Countries that have social and economic indicators that are less favorable for public welfare than those of the other Less Developed Countries.

Lexis Diagram A graph relating time and age and thus illustrating how, with the passage of time, the age references of a birth cohort change. If the diagram is shown in three dimensions, actual population values may be depicted.

Life Expectancy Average future lifetime for those surviving to exact age x in the life table, conventionally symbolized as e_x. Usually refers to the value at birth and assumes that current age-specific mortality rates remain constant; in this use the term represents Period (or Calendar Year) Life Expectancy at Birth and is equivalent to the Mean Age at Death (or Mean Length Of Life) in the corresponding life table. *See* also Period (or Calendar-Year) Life Expectancy and Cohort Life Expectancy.

Life Span The verified age at death of an individual, which can range from minutes after a live birth to the age at death of the world's longest-lived individual. A second definition, widely used by demographers and biostatisticians, refers not to individuals but to a species. It is the age to which a substantial number of members of the species can survive under optimum conditions and is commonly presumed to be in the range of 95 to 104 for humans. In this sense the term is to be distinguished from the term life expectancy. (See Life Expectancy, Average (mean) Observed Life Span, Maximum Observed Life Span.)

Life Span Potential The life span of a given individual under the theoretical assumption that he or she experiences ideal living conditions throughout life. The term is useful in distinguishing between the presumed additions to the length of life resulting from modifications in environment and life style and behaviors (e.g., diet, exercise, caloric restriction), and medical interventions (e.g., heart bypass and dialysis). The former practices enable individuals to achieve the life span potential more closely that they had at conception and the latter procedures add survival time beyond the life span potential to an individual who would otherwise have died. In an alternative interpretation, the term encompasses the survival time gained from all these factors, including medical interventions.

Life Table A table describing quantitatively the course of mortality throughout the life cycle. It consists of several functions, including mainly age-specific mortality rates, probabilities of death, survivors of an initial hypothetical cohort of births to various exact ages, deaths between the exact ages, and life expectation at each exact age.

Logistic Regression A form of linear multivariate regression in which the independent variable is expressed as a logarithm, usually in the form of a logit, i.e., the logarithm of the ratio of a percent to its complement.

Logit *See* Logistic Regression.

Longevity A general term representing a variety of measures used to gauge the length of life. Also the state of having lived a particular number of years.

Longitudinal Survey A survey that is conducted at various dates so as to secure data for the same cohort, usually the cohort that enrolled in the survey at the initial date. This type of survey includes panel surveys and differs from cross-sectional surveys.

Long-Term Care Generally defined as the provision of health care, personal care, or social services over a substantial period of time to individuals who have functional limitations. It encompasses home care, community services, and institutional care. Most persons in long-term care settings in the United States reside in nursing homes.

Makeham's "Law" of Mortality A variation of the Gompertz "law" of mortality that includes a constant, or age-independent, term to allow for extrinsic, or exogenous, causes of death. It is represented as: $\mu_x = A + Be^{cx}$, where A is the constant term and Be^{cx} is the Gompertz function.

Malnutrition A general term for inadequate nutrition, including undernutrition and overnutrition. *See* Undernutrition; Overnutrition.

Maternal Mortality Rate Number of deaths in a year due to puerperal causes per 10,000 or 100,000 births in the year.

Maximum Life Expectancy The highest life expectancy observed in any country in the world in a given year.

Maximum Life Span *See* Maximum Observed Life Span and Maximum Recorded Life Span.

Maximum Observed Life Span Highest verified age for a living member of a species. In the case of humans the maximum observed life span as of October 1, 2008 is defined by the age of Edna Parker who was 115 years, 164 days, and lives in the United States.

Maximum Prevalent Life-Span *See* Maximum Observed Life Span.

Maximum Recorded Life Span Highest verified age of a (living or dead) member of a species. In the case of humans, the maximum recorded lifespan is defined at

this time by the age at death of Jeanne Calment, who died in 1997 at the age of 122 years, 164 days.

Mean Age at Death The arithmetic mean of the ages of decedents in a given year either in the observed population or in a given life table population. In the life table it is equivalent to life expectancy at birth. It is also called the Mean Length of Life.

Mean Life Span *See* Average or Mean Observed Life Span.

Medical Sociology The branch of sociology concerned with the demographic, socioeconomic, and behavioral factors in the etiology of mortality and disease, the organizational dynamics of health institutions, the social construction of illness by a society, and social movements and socioeconomic policies related to disease and mortality.

Mendelian Diseases A disease that depends on the genotype at a single locus; that is, the disease is determined by a single gene. Single-gene diseases are uncommon to rare, treatment is usually unknown or difficult, and cures have not generally been found for them. An example of a Mendelian disease is cystic fibrosis.

Median Age at Death The median of the ages of decedents in a given year either in the observed population or in a given life table population. In the life table it corresponds to the age to which 50 percent of the original birth cohort survives. It is also called Median Length of Life.

Menarche The Gist menstructured period, corresponding to the period during adolescence when females begin to develop adult sex characteristics and become capable of reproduction.

Menopause The period characterized by the permanent cessation of menstruation and reproductive effectiveness in the human female, triggered by hormonal changes.

Meta-Analysis A quantitative analysis of the pooled results of several independent sample surveys relating to a specific health finding and intended to make more informative and accurate inferences regarding that health finding. Some studies of the same health outcome may have to be excluded from the meta-analysis because of incompatibility with the body of studies on the subject, e.g., variation in concepts, problems with sample selection, and so on.

Metabolic Syndrome A cluster of biomarkers characterized by high levels of blood glucose and triglycerides, high insulin resistance, low levels of high density lipoproteins, high blood pressure, an increased tendency to form blood clots, and abdominal obesity. Having three of these warrants a diagnosis of metabolic syndrome. These risk factors collaboratively increase the risk of developing cardiovascular disease and diabetes.

Methylation The process by which methyl groups (CH_3) attach to DNA.

Microdata Data collected in censuses, sample surveys, or administrative records for individuals. It contrasts with macrodata, that is, aggregated data or data tabulated for groups of individuals.

Migrant A person who has changed his/her usual residence between two specified dates between two specified area units. The term encompasses internal migrants and international migrants.

Mid-Upper Arm Circumference (MUAC) An index of peripheral protein and fat stores and of acute undernutrition measured on a straight left arm (for right-handed persons) midway between the tip of the shoulder and the tip of the elbow.

Mitochondria Cell organelles that produce biochemical energy by metabolizing glucose and other sugars. A high level of free radicals is produced during this process, which damages the DNA contained in the mitochondria.

Mitosis The process of replication of DNA by its division into two equal parts, generating two identical "daughter" cells from one "mother" cell.

Modal Age at Death The age at which the maximum number of deaths occurs in the observed population in a given year or in a given life table population. The curve of age-specific deaths is bimodal, but the modal age at death usually refers to the adult age, where the diminishing number of survivors is offset by rising death rates.

Model-Based Estimates An estimate for a "domain" (a subnational area or a socioeconomic group within such an area) derived by combining a sample survey estimate based on a national sample survey ("design-based estimate") and an independent estimate (s) derived from administrative data or regression analysis. Often, when a national sample is taken, the direct estimates from the sample for small geographic units within the country lack adequate precision because the domain and the sample for the domain are too small. To derive more reliable estimates where the sample domain is small, independent estimates may be generated and combined by some weighting procedure or otherwise with the sample survey estimates.

Model Life Tables A set of life tables designed as generalized patterns of lifetime mortality and survival and prepared with values of life expectancy at birth at specified standard levels and at a fixed interval (*e.g.*, 60 years, 65 years, 70 years, etc.)

Model Stable Populations A set of stable population age distributions designed as generalized patterns of age structure and prepared to correspond to specified levels of growth rates and mortality. With a fixed growth rate and a fixed mortality level, each age distribution remains unchanged over time. These characteristics define a Stable Population (*see*).

Molecular Biology The branch of biology that deals with living tissue at the molecular and cellular levels.

Money Value of a Person A monetary value placed on a person who dies prematurely and representing the value of the years of work lost. Calculated on the basis of generalized data on age at death, sex, average age at retirement, education, employment ratios, and other relevant characteristics.

Morbidity A generic term for sickness that includes injuries and impairments as well as diseases.

Mortality Crossover The shift from an excess of age-specific mortality of one group over another to an excess of age-specific mortality of the second group over the first. The two groups may be sexes, races, ethnic groups, or national populations. This shift has been observed to occur in comparisons of the mortality of various groups at the various parts of the age distribution, but it is commonly observed to occur at the upper end of the age span.

Mortpak A software package for demographic measurement in the Less Developed Countries, with special emphasis on mortality measurement. It was developed in 1988 by the United Nations' Population Division.

Multicollinearity The inclusion of two or more independent variables in a multiple regression analysis that are highly correlated, with the result that the interpretation of the association between independent and dependent variables is severely compromised.

Multiparous Having given birth to more than one living child.

Multiple Imputation The process of filling in missing data by use of several multiple regression equations, thereby deriving multiple entries for the missing cells.

Multiple Increment-Decrement Life Table A type of life table in which at least one other factor in addition to total deaths is included and which allows for both incremental and decremental factors of change in the initial cohort. Example: A disability life table that allows for death, disablement, and recovery from disablement.

Multiplicity Counting Rule A method of collecting data in surveys by which responses are secured not only for members of a household on select questions but also from these members about close relatives residing in a different housing unit. It is a device for effectively enlarging the sample when the event or characteristic is relatively rare while saving the expense of enumerating additional households.

Multistage Cluster Sampling A type of complex sample design in which sample selection passes through a series of stages, with areas being selected first and then sample clusters of households being selected randomly from the sample areas.

Multistate Life Table A type of multiple increment-decrement life table in which transition probabilities are employed instead of event-exposure rates to make possible and efficient numerous changes of states, including death, and based on principles of Markov processes. Because of the complexity of such tables, computer programs and matrix algebra are used as tools in their derivation.

Mutation A permanent change in a DNA sequence within a gene or chromosome of an organ resulting in a change in a trait(s) or characteristic(s) of the organism.

Negative Pleiotropy *See* Antagonistic Pleiotropy.

Net Worth The value of an individual's or household's assets minus debts. Assets include banks accounts, value of home and other property, saving certificates, stock, bonds, and so on.

Neurological Disorders Disorders of the neurological system, including most commonly Alzheimer's disease, Parkinson's disease, and epilepsy.

Non-Mendelian Diseases Diseases that depend on two or more genotypic loci (i.e., involving many genes) and are influenced by environmental factors. An example of such a disease is lung cancer.

Normal Aging Average deterioration in organs and physiological systems shown by persons with increasing age. Biomarkers with range values are often used as a standard for normal aging in evaluating the health of individuals at particular ages; they usually reflect the typical range for particular biomarkers shown by persons at that age.

Nosocomial Infection An infection acquired in a hospital by a patient apart from the primary cause for treatment in the hospital.

Nosology The branch of biostatistics that deals with the classification of diseases.

Nucleotide The building block of DNA, consisting of deoxyribose (a sugar), phosphate, and one of four nitrogenous bases.

Nulliparous Never having given birth to a live child.

Obese Under current recommendations, having a body mass index of 30.0 or more. (*See* Body Mass Index)

Odds Ratio In clinical studies or in logistic regression analysis, a measure of the relative risk of experiencing an event, incurring a disease, or experiencing a different outcome from a treatment. In clinical studies the odds ratio is defined as the ratio of the odds that the cases that incurred a disease were exposed to the disease to the odds that the cases that did not incur the disease were exposed to the disease. The odds of an event is the ratio of the number of ways an event can occur to the number of ways it cannot occur, or the probability of success divided by the probability of failure. An odds ratio of 1 between two "treatment" groups indicates that the risk of an adverse outcome is the same in each group. Similarly an odds ratio of 2 indicates that the risk of an adverse outcome is twice that for the first group compared to that for the second group.

Oncogene A gene that causes normal cells to change into cancerous tumor cells. In other words, it is a region of DNA in a human cell that can effect uncontrolled growth of the cell if not suppressed by a growth-suppressor gene.

Overnutrition A type of malnutrition involving intake of an excessive volume of food or the wrong types of food.

Oversampling In survey sampling, the practice of increasing the size of the sample for some sub-group of the population in order to obtain estimates with acceptable sampling errors, and then adjusting the sampling weights accordingly.

Overweight Under current recommendations, having a body mass index of 25.0 or more.The term "overweightness" is used here to refer to the state of being overweight. *See* Body Mass Index.

Oxidative Stress The effect of oxygen free radicals in linking with cells and altering their functions negatively.

Pandemic Characteristic of a disease that has spread over a wide geographic area and that has affected a substantial proportion of the population.

Parameter A constant(s) in a model estimating equation linking the variables and the other constants. For example, the equation $Y = a + bX$ has two parameters, a and b, and two variables, X and Y. The term is also used to refer to the principal mathematical characteristics of a geometric figure or an object, for example, the values for the length and width of a rectangle.

Partial-Birth Abortion Procedure usually performed in the second trimester of uterogestation in which the life of the fetus is terminated after partial delivery through the birth canal. This procedure is medically called intact dilation and extraction and is usually performed to save the life of the mother. The procedure is banned by many U.S. States, and U.S. Federal law.

Period Effect The effect on a variable (rate or ratio) of the changes it undergoes from one calendar year to another, usually reflecting major social events.

Period or Calendar-Year Life Expectancy Average future lifetime for those surviving to exact age x in the life table, assuming that current age-specific mortality rates remain constant. Usually refers to the figure at birth and in this form is equivalent to the Mean Age at Death (or Mean Length of Life) in the corresponding life table.

Personhood A state where a human being that can live independently of its mother and has integrated functions permitting its own survival. Personhood requires such characteristics as consciousness, self-awareness, and the ability to direct one's attention and act purposively. This state can be said to be achieved at the point of birth.

Persons of Concern The several categories of displaced persons classed together by the United Nations for statistical analysis, namely, refugees, internally displaced persons, returning refugees, asylum-seekers, and stateless persons. *See* Refugees and Internally Displaced Persons.

Person-Years Lived The sum of the amounts of time lived by the members of a population during some specified time period, usually a year.

Phenotype/Phenotypic Variation The observable physical (including disease) or biochemical traits of an organism; these can be determined both by genetics and environment. Phenotypic variation equals the sum of genetic variance and environmental variance.

Physiological Biomarkers Measures of the functioning of the various physiological systems intended to assess the health of the person, such as systolic blood pressure, total cholesterol, cortisol, C-reactive protein, and creatinine.

Placebo A dummy, false, or inactive treatment given to members of the control group in clinical trials. It may consist of sugared water or a fake procedure and is intended to mimic the active treatment given to the experimental group and to aid in the measurement of the effect of active treatment over and above the effect of inactive treatment.

Polygenic Describing a group of nonallelic genes that individually have a small effect but together are responsible for much phenotypic variation.

Population Aging Changes in demographic indicators reflecting a rise in the age of a population, such as a rising proportion of persons 65 years and over, a rising mean age, or a rising ratio of persons 65 years and over to persons under 18 years. This is a characteristic of populations and is affected by the demographic factors of fertility, mortality, and migration. It is possible for a population to grow younger with the passage of time. *See* Population Younging.

Population Younging Changes in demographic indicators reflecting a decline in the age of a population, such as a declining proportion of persons 65 years and over, a declining mean age, or a declining ratio of persons 65 years and over to persons under 18 years. It is possible for a population to grow older and younger at the same time, as when the proportion of elderly persons and children both increase.

Preimplantation Genetic Diagnosis (PGD) Laboratory examination of fertilized embryos created by *in vitro* fertilization for genetic abnormalities prior to implantation in the mother's uterus. The screening of the embryos for mutations and other genetic disorders may be intended to assure genetic compatibility with a preexisting offspring who may be suffering from an untreatable disease and who may be saved by gene therapy with the participation of a newborn sibling.

Prevalence Ratio The number of persons having a disease or other health condition at a specified date per 1,000 or 10,000 in the population at that date. Alternatively, the number of person even having the disease during a year per 1,000 or 10,000 person at midyear.

Prevalence-Ratio Method A method of constructing a multiple decrement-increment life table in which prevalence ratios are employed to allow implicitly for the increments and decrements other than deaths.

Productivity Output per unit of labor or production of goods and services per hour.

Proportional Hazards Model The model usually assumed for the survival curves of two populations being compared in clinical studies of different treatments, and

for the solution equations, with time varying variables and "constraints" namely, that the probability of dying at time t for individuals who have lived up that time is a constant proportion between the two test groups. *See* Hazard Rate.

Prospective Proportion of Elderly Persons The proportion of persons in the observed population over the age corresponding to a specified years until death (e.g., 10 or 15 years) in the current life table.

Protein Cross-Linking The bonding of proteins with genes, with adverse effects on the functions of the genes.

Psychiatric Disorders Disorders commonly viewed as mental and emotional disorders, such as major depression, bipolar disorder, schizoaffective disorder, and schizophrenia.

Puberty The period during adolescence when males develop adult sexual characteristics and become capable of reproduction.

Quadrate Survey Method In survey sampling, a form of area probability or spatial sampling that divides the total area equally in small square blocks and then counts the population in a random sample of the blocks.

Quality-Adjusted Life Years (QALY) A combination of years of life lost (as compared to some standard goal) and years of unhealthy or disabled life.

Randomization Selection according to a probability distribution, that is, such that any member of the population is as likely as any other member to be selected for a study, and any given individual is as likely to be selected for one sample group as another (i.e., treatment or control group). It is usually accomplished with tables of random numbers. In *stratified randomization*, the population is first divided into strata and random selection is carried out within these strata. The purpose of stratified randomization is to be sure that sufficient cases are selected for each stratum so that the results for these strata can be analyzed separately.

Randomized Clinical Trial An experimental, longitudinal (panel) study designed to test the efficacy of a specific drug or other treatment in which people are assigned randomly to one of two or more groups, one the experimental group that receives the treatment being evaluated and the other the control group that receives an alternative treatment. The alternative "treatment" may be a placebo or no treatment at all. Random assignment to the groups on a double-blind basis (i.e., not known to analyst and subject) is intended to reduce selection bias and to maximize their similarity so as to distinguish them primarily on the basis of treatment given. As the groups are tracked through time, differences in the efficacy of the treatment are observed and compared.

Randomized Response Technique A technique of collecting information on sensitive subjects designed to conceal the answer of the respondent to the sensitive question from the interviewer or other data-collecting agent. It involves a probability scheme by which it is determined whether the respondent selects either the sensitive question or an alternative nonsensitive question for reply.

Recessive Trait A trait that is expressed only when an individual inherits two identical copies of the gene causing the trait; hence, it is not expressed when it is present with a dominant allele. More formally, recessive refers to that one of a pair of alternative alleles whose effect is masked by the action of the second allele when both are present in the same cell or organism.

Recombinant DNA Genetically engineered DNA that contains DNA from two different sources, prepared by splicing genes from one species into the cells of an organism of a different species.

Record Linkage (or Case-By-Case Matching of Records) An integrated file derived by linking two or three different files. Linking files involves matching them on a case-by-case basis. Linking a record with other records provides a powerful and efficient means to augment their analytical potential for research.

Rectangularity A distribution model for a variable in which the variable is assumed to change linearly, or in even increments or decrements in a particular range.

Rectangularization of the Survival Curve The tendency of the survival curve to become more rectangular in shape at the higher ages, reflecting a compression of mortality in a narrower range of ages. It can be measured in several ways, such as by a reduction in the interquartile survival range, a reduction in the standard deviation of life table deaths around the modal age of death in the life table, or a rise in the ratio of the average (negative) slope of the survival curve after the modal age of death to the average slope of the survival curve in the adult ages prior to the modal age of death.

Refugee A person who is outside his or her country of nationality and who is unable or unwilling to return to that country because a well-founded fear of persecution. Fear of persecution may be based on the person's race, ethnicity, religion, nationality, or political views.

Regenerative Medicine The branch of medicine concerned with producing more healthy body parts or generating body parts through gene therapy, stem cell transfer, and recombinant DNA technology, It includes such techniques as angiogenesis and osteogenesis.

Relative Interquartile Survival Range The interquartile survival range as a percentage of the second quartile (or median) value. It is commonly used as a measure of dispersion of a distribution adjusted for the general level of the distribution. *See* Interquartile Survival Range.

Remaining Life Expectancy at Age x Life expectancy at any age x above age zero (*i.e.*, birth) in a life table. The figures may be derived from a period life table or a generation (cohort) life table.

Reproductive Aging The process by which reproduction become less efficient and successful as the woman gets older during the reproductive period. It is indicated by

the falling percentages of fecund women and the declining rates of natural fertility. In the later phases, it is also indicated by the various signs of the premenopause, such as irregular and longer menstrual cycles, increased frequency of anovulation, and more variable (with lower) estrogen levels.

Reproductive Health "A state of complete physical, mental, and social well-being in all matters relating to the reproductive system and to its functions and processes." – United Nations' 1994 World Plan of Action. It encompasses the ability to bear healthy children, avoid pregnancy losses and unintended pregnancies, regulate fertility, engage in satisfying sexual behavior without fear of disease and unwanted pregnancy, and avoid sexually transmitted diseases. It implies access to contraception and contraceptive services that are affordable and safe and to maternal and child health services that seek to prevent child and maternal mortality.

Response (item) Bias Responses to survey or census questions in the incorrect class of a variable, as when a person who has high blood pressure reports that his or her blood pressure is normal. This type of misclassification is commonly used to describe populations rather than individuals. In this case the group value is deemed to deviate (plus or minus) from the true value for this item. It may be compared with nonresponse bias, which denotes the bias introduced by the omission of responses to a survey question.

Replacement Level (or rate) The number of children the average woman would need to have in her lifetime in order to replace her and her mate. It depends on the sex ratio at birth and the survival rate of females to the mean age at childbearing. It now approximates 2.1 in the industrial countries.

RNA (Ribonucleic Acid) The molecule that transmits genetic information carried by DNA to the protein-manufacturing part of the cell. It consists of a long, usually single-stranded chain of alternating phosphate and ribose units with the bases adenine, guanine, cytosine, and uracil bonded to the phosphate.

Sagittal Abdominal Diameter The length of the straight line through the body from slightly below the navel to the lower back. It is used as a measure of abdominal fat.

Salmon Bias The effect, on the mortality levels of an ethnic group in a country of immigration, of the return movement to the country of origin of some of the older migrants after residing in the country of immigration for a period of time. This return movement tends to lower the mortality of the ethnic group in the country of immigration.

Sarcopenia Loss of lean body mass and of muscle strength. It results from the decrease in the number and size of muscle fibers and the amount of muscle tissue with aging.

Scaling Adjustment, for differences in life span, of the age distributions of variables that are a function of age (e.g., age-specific death rates), for two or more

population groups being compared e.g., mice and humans, to facilitate comparison of the age patterns of the groups. A common method of scaling equates the mean values of the distributions. Some measures of comparison automatically allow for the scaling of the age distributions, e.g., relative interquartile range, coefficient of variation.

Self-Assessed (Self-Reported; Self-Rated) Health Status Census or survey respondents' reports as to their health status, usually in response to a question calling for a choice between poor, fair, good, very good, or excellent health. *Same as* Subjective Health.

Senescence The process by which age-related vulnerabilities manifest themselves as disease or disability. Alternatively, the process by which deleterious alterations in structure and function of organisms, such as dementias of the Alzheimer type, emerge during the later stages of the life cycle.

Separation Factor In mortality analysis refers to the proportions of infant deaths in a given year occurring to births of the prior year and to births of the current year.

Seroprevalence The percent of a population testing positive for infection in a blood test at a particular date, for example, the percent testing positive for antibodies to HIV.

Single Nucleotide Polymorphism (SNP) A variation in a DNA sequence occurring when a single nucleotide, A, C, G, or T, in the genome differs between paired chromosomes in an individual or between individuals; for example, AAGCCTA and AAGCTTA. A nucleotide is a small fragment of DNA consisting of a nitrogenous base (T, A, G, or C), a sugar, and a phosphate. When these molecules are joined together, they constitute the structural center of DNA. *See* Nucleotide.

Single-Payer Insurance System An insurance system in which a single entity, usually a government agency, pays for the costs of medical care for all participants in the insurance pool. It may be combined with multiple health providers and with coverage of only a segment of the population, as with Medicare in the United States. The health insurance system in Great Britain combines the national government as single payer with paid government health providers and universal coverage.

Socioeconomic Status A characteristic of members of a population, assigning them to a category in the socioeconomic hierarchy, usually measured on the basis of educational level, household income, household wealth, or occupation, or some combination of these characteristics.

Somatic Mutation The process by which, as animals age, the genes in somatic cells mutate as a result of the effect of oxygen free radicals, radiation, protein cross-linking, copying errors, and other factors.

Stable Population A population whose age distribution does not change from one date to another. Such an age distribution results from an unchanging birth rate, an unchanging death rate, and an unchanging annual growth rate. In actuality, it arises after such conditions have been in effect for about 70 or 75 years.

Stationary Population A population whose age distribution does not change from one date to another and whose rate of growth is zero, so that the size of the total population and the number in each age group remain unchanged over time. A life table population is an example of a stationary population. A stationary population is a type of Stable Population (*see*).

Stem Cells Unspecialized cells that can become one of a multitude of specialized cells and develop the functions of the specialized cells. An example is the cells of bone marrow from which hematopoietic and lymphopoietic cells develop. Stem cells can also be obtained from embryos and umbilical cord blood. Fetal cells and some adult cells can be "trained" to become stem cells.

Stochastic Factor The factor representing chance occurrence or probability. Stochastic forces play a major role in extrinsic (or exogenous) causes of illness or death and a substantial role in intrinsic causes of death.

Stock vs. Flow Measures Distinguishes measures of prevalence and measures of incidence, or measures relating to persons with the characteristic at some date and those relating to persons taking on the characteristic during a specified period.

Stunting A state of undernutrition, especially in children, characterized by low height-for-age, usually defined as more than two standard deviations below the NCHS/WHO reference median value or below 80% of the NCHS/WHO reference median value. It results from prolonged nutritional deficit and/or disease. Currently, no such definition of stunting in adults is available.

Subjective Health *See* Self-Rated Health.

Subjective Survival A survey respondent's self-assessed years of survival, e.g., years until death, age at death, or probability of survival another 5 years.

Successful Aging A theoretical concept based on the view that observance of the risk factors of aging will enable an individual to live to advanced old age without incurring serious chronic illness. It is also intended to encourage positive risk-factor behavior and life style. However, it has many limitations as an actual description of aging and as a theory of aging, and is an inadequate description of anyone's actual aging experience.

Supercentenarian A person with a verified age of 110 years or more.

Support Ratio The ratio of persons of working age to persons whom they support defined by age. An example is the ratio of persons 18–64 to persons 65 years and over.

Surveillance Continuing scrutiny of a health situation in order to detect changes in distribution or trend of a disease and to institute control measures when needed. The principal general purpose of a surveillance system is to provide up-to-date information in order to assist in decision-making at a community level.

Survival Analysis Application of statistical methods to survivors as they age and progress over time. These methods include the use of regression methods with life tables, such as in proportional hazards models.

Survival Curve The curve (or function) representing the number of survivors to successive ages in the life table, symbolized by l_x. It is also defined as the probability P that the failure time X is greater than time x, designated as (P(X > x).

Survival Rate The proportion of a cohort of individuals aged x (or x *to* $x + n$) surviving to age $x + t$ (or $x + t$ to $x + n + t$). It is usually based on the ${}_nL_x$ function of an abridged life table and is symbolized by ${}_nL_{x+t} \div {}_nL_x$. It is used to measure survival propensities both during past periods and future periods.

Synthetic Cohort A special type of birth cohort in which the entire schedule of data for the ages or age groups refers to the same calendar year or brief group of years and, in combination, constitute a hypothetical lifetime of experience describing that year or years. In demographic studies life expectation at birth of the standard life table and the total fertility rate are the most common summary measures describing synthetic cohorts.

Table of Active (Healthy, Disability-Free) Life An extension of the standard life table incorporating a measure of health status (e.g., percent disabled, percent in fair or poor health, percent with a chronic health condition with limitation of activity) and providing estimates of years of active (healthy, disability-free) life remaining at each age.

Telomere The tip of a chromosome. It is made of densely packed DNA and protects the chromosome against damage. With increasing age, the telomere tends to become shorter.

Third Age In Peter Laslett's book, *A Fresh Map of Life*, the ages from roughly 50 to 75, when work and family responsibilities are diminishing and health is essentially very good or excellent, permitting choices for personal growth and development, including especially the pursuit of higher education.

Total Fertility Rate The total number of children a woman would have in her lifetime assuming that she experienced the age-specific fertility rates of the given year and that she lived to the end of the childbearing period.

Total Life Expectation The sum of the attained age of survivors and life expectation at that age given in a particular life table. For example, total life expectation at age 85 in the United States in the year 2000 was 91.3 years (= 85.0 + 6.3).

Transition Probability The proportion of a population distributed among a number of categories (*e.g.*, not disabled, moderately disabled, severely disabled) at an initial date that will be located in these categories and a death category at a terminal date, usually for a period of 1 year. They are derived for use in the construction of multistate life tables.

Undernutrition *or* Undernourishment Intake of insufficient total amounts of food or insufficient amounts of the nutrients required for health. More specifically, a condition where dietary energy consumption, in the form of proteins, carbohydrates, and fats, and vitamins and minerals, is continuously below minimum dietary energy requirements for maintaining health and carrying out light physical activity.

Universal Health Insurance System A health insurance system that covers all members of a population. It may be combined with multiple health providers and government management or private management of the system, although private management is difficult to envisage with universal health insurance.

Unobserved Heterogeneity The heterogeneity in the characteristics of a group that is not reflected in the observed measure or value. For example, an age-specific death rate for all races conceals the differences in mortality at the age among the various races, or a birth cohort includes a mixture of persons of varying states of health – a mixture that changes over time.

Unmet Need for Family Planning The proportion of currently married women of reproductive age not using contraception but wishing either to prevent unwanted childbearing (after having achieved their desired number of children) or to postpone the next wanted birth.

Utilization (of health care) Per capita use of health-care services, including use of new or existing technology, frequency of visits, type of services sought, and the quality and quantity of health services received. The term is used in analysis of the factors contributing to changes in the cost of health care.

Viability Ability of a fetus to live independently of the mother or to be maintained independently with medical intervention. Some fetuses in the second trimester are viable but most fetuses are not viable until the third trimester.

Voodooism A form of traditional "medicine" practiced in some Caribbean countries, especially Haiti, based on a religion involving a combination of West African animism and Roman Catholic ritual practices.

Waist-Hip Ratio The ratio of the circumference of the body at the waist to the circumference of the body at the hips. It is used as a measure of abdominal fat and obesity.

Wasting A state of undernutrition characterized by low weight for height, low weight for age, or small arm circumference for age. It is caused by acute nutritional deficit and a disease such as diarrhea.

Weibull (or power) "Law" A generalization relating the age trajectory of failure rates for manufactured products. According to this "law" the logarithm of the failure rate is assumed to increase linearly as a function of the logarithm of age: $\mu_x = Be^{c \ln x}$, or $\ln \mu_x = \ln B + c \ln x$.

Wellness A state of positive physical, mental, psychological, and social health.

Xenotransplantation The process of transplanting animal organs into humans.

Zoonoses or Zoonotic Diseases Infections that are naturally transferred from animals to humans.

Appendix 4: Classifications of Countries and Regions

Major area or region
World
More developed countries[a]
Less developed countries[b]
Least developed countries[c]
Other less developed countries
Africa
Sub-Saharan Africa
Eastern Africa
Middle Africa
Southern Africa
Western Africa
Northern Africa[d]
Asia
East Asia
South Central Asia
Southeast Asia
Western Asia
Europe
Eastern Europe
Northern Europe
Southern Europe
Western Europe
Latin America and Caribbean
Caribbean
Central America
South America
Northern America
Oceania

(continued)

(continued)

Australia
New Zealand
Melanesia
Micronesia
Polynesia

Exhibit A.1 Major areas and regions of the world and distribution of the countries of the world as more developed and less developed, according to the United Nations as of 2007
Source: United Nations, Population Division; *UN.org/esa/population*
[a]Comprising all regions of Europe and Northern America as well as Japan, Australia, and New Zealand
[b]Comprising all regions of Africa, Asia (except Japan), and Latin America and the Caribbean, as well as Melanesia, Micronesia, and Polynesia
[c]See Exhibit A.3
[d]The United Nations excludes Sudan from Northern Africa and places it in sub-Saharan Africa; the Population Reference Bureau includes Sudan in Northern Africa

Economies in transition [a]	Former Soviet Republics in Asia	Former socialist economies in Europe [b]
Albania	Armenia	Albania
Armenia	Azerbaijan	Belarus
Azerbaijan	Georgia	Bulgaria
Belarus	Kazakhstan	Czech Republic
Bulgaria	Kyrgyzstan	Estonia
Czech Republic	Tajikistan	Hungary
Estonia	Turkmenistan	Latvia
Georgia	Uzbekistan	Lithuania
Hungary		Moldova
Kazakhstan		Poland
Kyrgyzstan		Romania
Latvia		Russian Federation
Lithuania		Slovakia
Moldova		Ukraine
Poland		
Romania		
Russian Federation		
Slovakia		
Tajikistan		
Turkmenistan		
Ukraine		
Uzbekistan		

Exhibit A.2 Economies in transition, former Soviet Republics in Asia and former socialist economies in Europe
[a]Combination of former Soviet Republics in Asia and former socialist economies in Europe. List is consistent with the region designated as economies in transition, used by World Bank
[b]List is consistent with the region designated as former socialist economies in Europe in C.J.L. Murray and A.D. Lopez (eds.), *The Global Burden of Disease*, WHO, World Bank, and Harvard University 1996

Afghanistan	Madagascar
Angola	Malawi
Bangladesh	Maldives
Benin	Mali
Bhutan	Mauritania
Burkina Faso	Mozambique
Burundi	Myanmar
Cambodia	Nepal
Cape Verde	Niger
Central African Republic	Rwanda
Chad	Samoa
Comoros	Saõ Tomé and Principe
Congo (Democratic Republic of the Congo)	Senegal
Djibouti	Sierra Leone
Equatorial Guinea	Solomon Islands
Eritrea	Somalia
Ethiopia	Sudan
Gambia	Tanzania (United Republic of Tanzania)
Guinea	Timor-Leste
Guinea-Bissau	Togo
Haiti	Tuvalu
Kiribati	Uganda
Laos (Lao People's Democratic Republic)	Vanuatu
Lesotho	Yemen
Liberia	Zambia

Exhibit A.3 List of least developed countries according to the United Nations as of 2007
Source: United Nations, Population Division; Murray and Lopez (1996); *UN.org/esa/population*
Note: Includes landlocked and small-island less developed countries

World
Low- and middle-income countries
East Asia and South Asia
Western Asia and Central Asia
Eastern Europe
Latin America and Caribbean
Middle East and North Africa
Sub-Saharan Africa
High-income countries
Northern America
Northern/Western/Southern Europe

Exhibit A.4 Low- and Middle-income countries and high-income countries according to the World Health Organization as of 2004
Source: World Health Organization, *World Health Report, 2004*

Appendix 5: List of Abbreviations and Acronyms

ADL	Activities of daily living
AID	(U.S.) Agency for International Development
AIDS	Acquired immunity deficiency syndrome
BMI	Body Mass Index
CDC	(U.S.) Centers for Disease Control and Prevention
COPD	Chronic obstructive pulmonary disease
CR	Caloric restriction
CVD	Cardiovascular diseases
DALY	Disability-adjusted life years
DHS	Demographic and Health Survey
DNA	Dioxyribonucleic acid
GAO	(U.S.) General Accountability Office
GBD	Global burden of disease
GDP	Gross domestic product
GDI	Gross domestic income
GIS	Geographic Information System
HALE	Health-adjusted life expectancy
HHS	(U.S. Department of) Health and Human Services
HIV	Human immunodeficiency virus
IADL	Instrumental activities of daily living
IMR	Infant mortality rate
IOM	Institute of Medicine
IVF	*In vitro* fertilization
IUD	Intrauterine device
MDG	Millennium Development Goals
MMR	Maternal mortality rate
NCHS	(U.S.) National Center for Health Statistics
NIA	(U.S.) National Institute of Aging
NIH	(U.S.) National Institutes of Health
NGO	Non-governmental organization
PGD	Pre-implantation genetic diagnosis
QALY	Quality-adjusted life years
STD	Sexually transmitted disease
TB	Tuberculosis
TFR	Total fertility rate
UN	United Nations
UNAIDS	(Joint) United Nations Program on HIV/AIDS
UNHCR	(Office of the) UN High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNFPA	United Nations Fund for Population Activities
WHO	World Health Organization
YYL	Years of life lost to premature mortality
YLD	Years of life lost to disability

Index

J.S. Siegel, *The Demography and Epidemiology of Human Health and Aging*,
DOI 10.1007/978-94-007-1315-4, © Springer Science+Business Media B.V. 2012

D

E

N